A PRELIMINARY STUDY ON AUTOMATED FRESHWATER ALGAE RECOGNITION AND CLASSIFICATION SYSTEM

HAYAT MANSOOR ABDULLAH

DISSERTATION SUBMITTED IN FULLY FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE
OF MASTER OF BIOINFORMATICS

INSTITUTE OF BIOLOGICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR
2012

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **HAYAT MANSOOR ABDULLAH** (Passport No: 02494321)
Registration/Matric No: **SGR080080**

Name of Degree: **MASTER**

TITLE ("A PRELIMINARY STUDY ON AUTOMATED FRESH WATER ALGAE RECOGNITION AND CLASSIFICATION SYSTEM"):

Field of Study: **BIOINFORMATICS**

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any Copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:
Subscribed and solemnly declared before,

Witness's Signature Date:
Name:
Designation:

Witness's Signature Date:
Name:
Designation:

Declaration

I certify that this dissertation is based on my independent work, except where acknowledged in the text or by reference. No part of this work has been submitted for degree or diploma to this or any other university.

Affirmed by:

HAYAT MANSOOR ABDULLAH

Date:

Supervisor: **Dr. SORAYYA BIBI MALEK**

Signature:

Co-supervisor: **Prof. Datin Dr. AISHAH BINTI SALLEH**

Signature:

Institute of Biological Sciences, Faculty of Science
University of Malaya,
Kuala Lumpur,
Malaysia.

ACKNOWLEDGEMENTS

I wish to express my great thanks to my supervisor **Dr. Sorayya Bibi Malek** for supporting, guiding, reading and correcting and flow me step by step I appreciate her continuous help during working and writing, also great thanks to my Co-Supervisor **Professor Datin Dr. Aishah Binti Salleh** for her motivation and consistent support.

Many thanks go to **Professor Dr. Rosli Bin Hashim** Dean of the Institute of Biological Sciences, Faculty of Science for his kind advice, helping ,and supporting.

Special mention for my **mother** whose push me to complete my study. And all my love to my idol **father** thanks for your inseparable support and prayers.

Words fail to express my appreciation and thanks to my husband (**Mogeeb A.A Mosleh**) deepest gratitude understanding & endless love, without his ideas, effort and suggestions I could not complete all this work, big love to my soul my sweaty daughter **Nesma.** Last but not the least, to my sisters and brothers.

ABSTRACT

Freshwater algae can be used as indicators to monitor freshwater ecosystem condition because algae react quickly and predictably to a broad range of pollutants. Research reported that Algae can provide early signals of worsening environment. This study was carried out to develop a computer-based image processing technique with artificial neural network (ANN) approaches to automatically detect, recognize, and identify algae genera from the divisions of *Bacillariophyta*, *Chlorophyta* and *Cyanobacteria* in Putrajaya Lake. Based on literature review, automated identification of tropical freshwater algae is even non-existent yet, and this study is designed to fill this gap.

The development process of the automated freshwater algae detection system involves with many techniques and computer methods such as image preprocessing, segmentation, feature extraction, and classification process by using ANN. Several image preprocessing steps was designed to contrast the images, remove the noise, and improve image quality and overall appearance. Then, Image segmentation applied by using canny edge detection algorithm with specific morphological operation to isolate the image objects components. Image segmentation was divided each input images into sub images where each sub images includes one object only. Feature extraction process was applied to extract some shape and texture features of algae image such as shape index, area, perimeter, minor and major axes, entropy, and Fourier spectrum. Then principal component analysis (PCA) was applied to normalize the extracted features. Novel techniques of auto-alignments with shape index procedures was developed here, where auto-alignments function was used to aligned image objects with horizontal coordinates to extracted object features in similar position. Shape index techniques are also considered a novel techniques developed to assist system in classification of algae based on their biological metrics and taxonomy. Shape index function is an index

number of different shape of algae as one component feature of algae diversity. Finally, 41of geometrical, texture, and novel features were normalized to feed into artificial neural network (ANN) for classification and recognition purposes. The Feed-forward multilayer perceptron network with back propagation error algorithm (MLP) initialized, and trained with extracted database feature of selected algae image samples. Experiment for comparison between manual process identification by experts with automated recognition process performed by system. The Proposed system was automatically able to classify five kinds of freshwater algae successfully, and experimental results showed that our approach is workable, and had a great accuracy results with more than 93%. Results also indicated that our approach is faster in execution, efficient in recognition rate, and easier for using and implementation if compared with similar developed systems.

This study demonstrated application of automated algae recognition of five genera of freshwater algae, there are *Navicula* form Bacillariophyta, *Scenedesmus* and *Chroococcus* from the Chlorophyta division, *Microcystis* and *Oscillatoria from* the Cyanobacteria division. The results indicated that MLP is sufficient, and optimal enough to be used for classification of the selected freshwater algae. The accurate results was obtained due to the specific preparation for algae image, well segmentation approach, and the novel methods of auto-alignments and shape index techniques which extremely enhanced system classifier of algae. However, for further improvements, we recommended to be included more features with different ANN such as support vector machine (SVM) and radial basis function (RBF) for better recognition rate as the number of algae species studied increases.

ABSTRAK

Alga air tawar boleh digunakan sebagai petunjuk untuk memantau keadaan ekosistem air tawar kerana alga bertindak balas dengan cepat dan boleh diramal untuk pelbagai bahan pencemar. Penyelidikan dilaporkan bahawa Alga boleh memberikan isyarat awal persekitaran yang semakin teruk. Kajian ini telah dijalankan untuk membangunkan teknik pemprosesan imej berasaskan komputer dengan rangkaian neural tiruan (ANN) pendekatan secara automatik mengesan, mengiktiraf, dan mengenal pasti genus alga dari bahagian Bacillariophyta, Chlorophyta dan Cyanobacteria di Putrajaya Lake. Berdasarkan kajian literatur, pengenalan automatik alga air tawar tropika walaupun tidak wujud lagi, dan kajian ini direka untuk mengisi jurang ini.

Proses pembangunan sistem pengesanan alga air tawar automatik melibatkan banyak dengan teknik dan kaedah komputer seperti pra pemprosesan imej, segmentasi, penyarian sifat, dan proses pengelasan dengan menggunakan ANN. Langkah-langkah pra pemprosesan imej beberapa bentuk untuk bezakan imej, menghapuskan hingar tersebut, dan meningkatkan kualiti imej dan rupa keseluruhan. Kemudian, segmentasi Imej memohon dengan menggunakan algoritma pengesanan pinggir hati-hati dengan operasi morfologi khusus untuk mengasingkan komponen objek imej. Segmentasi imej telah terbahagi setiap imej input kepada imej kecil di mana setiap imej sub termasuk satu objek sahaja. Ciri-ciri proses pengekstrakan telah digunakan untuk mengekstrak beberapa ciri-ciri bentuk dan tekstur imej alga seperti indeks bentuk, kawasan, perimeter, paksi minor dan major, entropi, dan spektrum Fourier. Kemudian analisis komponen utama (PCA) telah digunakan untuk menormalkan ciri-ciri yang diekstrak. Teknik novel automatik penjajaran dengan prosedur indeks bentuk dibangunkan di sini, di mana auto-penjajaran berfungsi digunakan untuk objek imej yang sejajar dengan koordinat mendatar kepada ciri-ciri objek yang diekstrak dalam kedudukan yang serupa. Teknik indeks bentuk juga dianggap sebagai teknik novel yang dibangunkan untuk

membantu sistem dalam alga klasifikasi berdasarkan Metrik mereka taksonomi dan biologi. Fungsi indeks bentuk diberi nombor indeks tentang bentuk badan yang berbeza, alga sebagai salah satu komponen ciri kerana kepelbagaian alga. Akhirnya, 41of geometri, tekstur, dan ciri-ciri novel telah kembali biasa makan ke dalam rangkaian neural tiruan (ANN) untuk tujuan pengelasan dan pengiktirafan. Feed-hadapan perceptron rangkaian berbilang dengan kesilapan algoritma rambatan kembali (MLP) dimulakan, dan dilatih dengan ciri-ciri pangkalan data yang diekstrak sampel alga imej terpilih. Eksperimen bagi perbandingan antara pengenalan proses manual oleh pakar-pakar dengan proses pengiktirafan automatik dilakukan oleh sistem. Cadangan sistem adalah secara automatik dapat mengklasifikasikan lima jenis alga air tawar berjaya, dan keputusan eksperimen menunjukkan bahawa pendekatan kami adalah yang dapat dilaksanakan, dan mempunyai keputusan ketepatan besar dengan lebih daripada 93%.

Hasil kajian menunjukkan bahawa MLP adalah mencukupi dan optimum cukup untuk digunakan untuk pengelasan alga air tawar yang dipilih. Keputusan yang tepat telah diperolehi kerana penyediaan khusus untuk imej alga, pendekatan segmentasi baik, dan kaedah novel auto-penjajaran dan teknik indeks bentuk sistem yang sangat dipertingkatkan pengelas alga. Walau bagaimanapun, bagi penambahbaikan, kami mencadangkan untuk dimasukkan lebih banyak ciri-ciri dengan ANN berbeza seperti mesin sokongan vektor (SVM) dan fungsi asas jejarian (RBF) untuk kadar pengiktirafan yang lebih baik sebagai bilangan spesies alga mengkaji kenaikan.

TABLE OF CONTAINS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION AND OBJECTIVES

1.1 Introduction

Freshwater ecosystems are including many components such as rivers, lakes, ponds, wetlands, streams, and springs. Freshwater habitats classified based on different factors including temperature, light penetration, and vegetation. Algae can be used as indicators to offer relatively exclusive information in ecosystem conditions. Algae react quickly and predictably to a broad range of pollutants, thus providing potentially constructive early caution signals of a worsening environment and the possible causes. Algae also provide some of the few standards for establishing water quality conditions and an early caution signal of worsening ecological conditions.

Several species of algae are capable of producing potentially harmful toxins as well as unpleasant taste and odor. Blue green algae, which are widespread in eutrophic lakes, have become a critical problem worldwide because of their toxicity. Surveys carried out in different countries demonstrated that about 75% of lake water samples contain toxic cyanobacteria. Blue green algae are also considered as a parameter for water quality control, and are recommended as a factor of risk assessment plans and safety level in different organization standards.

In Malaysia, a study on the eutrophication status of 90 lakes showed that 56 lakes (62%) are eutrophic or in a poor situation and requires immediate rehabilitation and restoration. The other 34 (38%) lakes are classified as mesotrophic.

Previous labors works was spurred by threats of human health, research attempts to understand and monitor freshwater ecosystems. The early monitoring of ecosystems is focused on including three essential components, namely, chemical indicators, bacteria, and algae. A new type of monitoring involves different organisms, such as macro-invertebrates, macrophysics, and fish, as well as the associated stream conditions.

1.2 Problem Statement

Malaysia is considered as a tropical zone area containing several lakes that can be used as water resources. Unfortunately, most of these lakes are in a poor situation, and require extensive efforts to be treated well before they can be used as water resources.

Conventional methods for analyses and measurements of water quality depend on the manual collection, capturing, and identification of different types of microorganism in microscopic images. However, the manual process is subjected to human errors, and considered a tedious and time consuming process.

Recently, image processing and computer imaging has grown at a fast pace, and computer architecture and components have become sufficiently powerful enough to solve complex tasks in processing image data. Computer-based image processing approaches are widely involved in solving many problems in the biology field. Several studies attempted to automate water quality analysis. Image processing combined with some other approaches have been employed to automate the detection and identification of microorganisms; however, each individual type of alga has its own features and requires a special model to develop a recognition process. Several algae are found in Malaysian lakes; however, there is no system that automates the process of detecting and identifying certain types of algae.

1.3 Research Objectives

In this study, we employed image processing with artificial neural networks (ANN) methods to develop a prototype system for detecting, identifying, and classifying several types of freshwater algae automatically.

The proposed system was developed to be used as a tool for monitoring water quality and estimating the density of microorganisms found in collected water samples by counting only objects in microscopic images. The purpose of the system was to detect and identify selected algae found in microscopic images based on the taxonomy and feature extraction of the algae.

This research developed a computer software program that automates the process of detecting, identifying, and classifying algal image samples based on some techniques of image processing with ANN. Several models were developed to facilitate the automated manipulation of the process for the input images, and each model was employed to perform essential tasks for achieving system goals. An image processing algorithm was implemented to achieve several tasks including contrast, filter, isolate, and recognize algae objects from the microscopic images of the collected samples of freshwater algae. These algorithms were implemented into several modules to improve the accuracy results of the freshwater algae recognition process. An image preprocessing module was used to contrast images, to remove noise and improve image quality. A segmentation module was used to isolate the objects found in the input image. Some new techniques are developed to extract image features including geometric and texture features. A combination of feed forward ANN with a feature extraction module was used to train and recognize the selected freshwater algae images. Finally, the accuracy rate, system reliability, and performance of the developed system were evaluated.

1.4 Organization of Thesis

This dissertation is structured as follows: Chapter one provides the summary of research including problem statements, objectives, and aim of study. Chapter two provides a literature review of about freshwater algae and current developed systems. Chapter three identifies the materials and methods used in this study, and describe the development process for each module in proposed system. It also describes the functional requirements of the methods and algorithms used during system design. Chapter four describes the system results and discusses the experiment results process with the system performance criteria. Chapter five draws the main conclusions and provides some guidelines for future works.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Water is an essential element of life on earth, and human beings give special consideration to water resources. Water covers most of our planet surface (approximately 70%; however, less than 1% of the total amount of water can be used. Only freshwater is suitable for human use, and can be diverted from different resources such as rivers, lakes, ponds, streams, springs, and wetlands. Water is used not only for human consumption, but also for domestic, livestock, agricultural irrigation, and different industrial applications (Wurbs and James, 2002). There are several organisms that can affect the quality of water and render it useless. Thus, water quality monitoring processes were spurred for ecosystems due to threats on human health. Early monitoring processes were focused on chemical indicators, bacteria, fungi, protozoa, and algae. However, modern approaches for monitoring water quality involve different groups of organisms such as macro-invertebrates, macrophysics, and fish, as well as the stream condition associated with each type.

Recently, a research reported that blue green algae play an important role in both short and long term processes for the determination of water quality in freshwater lakes. Algae affect water properties such as water color, odor, taste, and chemical compounds, which may cause potential hazards for human and animal health. Traditional processes used to remove algae from water include the coagulation processes such as pre-oxidation by ozone, chloride dioxide, and chlorine or permanganate (Gilbert, 1996; Gao et al, 2009).

They classified phytoplankton ranging from unicellular to multicellular as a kind of algae that float freely in freshwater ecosystem. Phytoplankton are found in colonies or long-chain filaments and appear as scum on water surfaces. Scum is a layer of dirt resulting from a mixture of various algal species. Phytoplankton play a vital role in all aquatic ecosystems and are primary producers that form the base of the food chain. Abnormal or excessive growth of this type of algae (phytoplankton) interferes with human enjoyment of aquatic resources and can even be harmful. algal community changes is depends on some reflect of pollutants occurrence, or other environmental stressors especially nutrients that make algae increased dramatically, and lead to decreased oxygen level in water which harmful other organisms in the aquatic food chain (Johnstone et al., 2006; Camargo and Alonso, 2006).

Algae are responsible for wide range of chemical and toxin compounds in their environments thus make algae a very good indicator for ecological conditions, also because of their highly receptive in changing the environment (Anton, 1991). Abundance of algal species is commonly used to detect environmental changes, and to indicate the trophic status, oxygen level, and nutrient problems of a lake (Patrick, 1994). Using algae as a biological indicator has been suggested by several studies to supplement the traditional method of monitoring. Algae provide unique information on the ecosystem condition, which is potentially useful as an early warning sign of deteriorating condition and its possible causes (McCormick and Cairns, 1994; Knoben et al., 1995; Masseret et al., 1998; Hillebrand and Sommer, 2000; Pipan, 2000; Rauch et al., 2006). Algae are mostly used as bioindicators due their rapid response to environmental changes and reproduction within a short period (Hobson & Welch, 1992).

Algae from Bacillariophyta and Chlorophyta, especially desmids (e.g., *Scenesdesmus*), are used as bio-indicators for monitoring water quality because their highly sensitive to their changes in environmental parameters (Coesel, 1983; Coesel, 2001; Leclercq, 1988). However, several species of algae are capable of producing potentially harmful toxins as well as unpleasant taste and odor. Chlorophytes are often abundant in eutrophic lakes, and blooms of *Staurastrum sp* have created grassy odor problems. For example, *Navicula sp* is a member of the group of algae called Bacillariophyta which does not decompose even if cell dies because it has hard cell walls. Their remaining skeletons of the cells generate several problems when they obstruct the filters uses for water treatment. Algae measurements are often used as key components of water quality monitoring because of their importance in aquatic ecosystems and susceptibility to environmental changes (Addy and Green, 1996).

Cyanobacteria are responsible for producing nuisance blooms in eutrophic waters. Some species of cyanobacteria, such as *Microcystis* and *Anabaena*, contribute to toxin, taste, and odor problems in water. Cyanobacteria have become a critical problem worldwide because of their toxicity and wide distribution in eutrophic lakes. Studies performed in different countries confirmed that about 75% of lake water samples contain toxic cyanobacteria (Chorus et al., 2000; Azevedo, 2001). Thus, cyanobacteria are considered as a parameter for water quality control a factor for risk assessment plans and safety level in some organization and standardization, such as the World Health Organization and several national authorities worldwide (Falconer, 2001; Codd et al., 2005; Walsby and Avery, 1996).

2.2 Water resource in Malaysia

Malaysia is a tropical area whose main freshwater sources are rain, river, and lakes. There are several rivers in Malaysia; some of them are located in peninsular Malaysia, such as Sungai Perak (390 km), Sungai Selangor (80 km), Sungai Muar (190 km), Sungai Kelantan (250 km), and Sungai Pahang (500 km). Others are located in East Malaysia such as the longest river in Malaysia called Sungai Rajang (760 km), and Sungai Kinabatangan (560 km) (World Wide Fund for Nature). There are also several lakes located in Malaysia that utilized for many different uses.

Lakes and reservoir are important sources of water in Malaysia and can have multiple purposes. They form part of storage basins for municipal and industrial water supply, and also function in agriculture and hydropower plants. Water resources in Malaysia are used extensively for domestic needs, agriculture, aquaculture, industries, hydroelectric power, and recreation (Ho, 1995).

A lake is considered as amount of water localized in a basin which surrounded by land, it is apart from a river, stream, or other form of moving water. Lakes are individual inland and not considered as a part of sea or ocean. Most lakes are fed and drained by rivers and streams which have distinct meaning from lagoons, and ponds. There are many type of lakes based on specific terms, however our research area is putrajaya lakes which refers to artificial lakes which created by flooding land behind a dam called an impoundment or reservoir. Putrajaya Lake dimensions 400 hectares created by flooding the valleys of Sungai Chuau and Sungai Bisa. It constructed within two different phases, initial phase designed to form approximately 110 hectares involved the construction of a temporary dam across Sungai Chuau, and second phase by extended it to be 400 hectares later on.

Putrajaya wetland is considered as the first man-made wetland in Malaysia, and also as one of the largest fully constructed freshwater wetland in the tropics. It designed with modern technology and stringent environmental management's method with yield 197 hectare project resulted for transforming an oil palm site into wetland ecosystem.

2.3 Algae in Malaysia

Patrick (1936) performed the first study on freshwater algae in Malaysia. He worked on the taxonomic identification of existing diatoms found in tadpole intestines (a type of frog) collected from Perak (Anton 1991). The earliest algology studies recorded the existence of some algal species, such as desmids and Euglenophyta from the Cholorophyta division, as well as *dinoflagellate* from the Chrysophyta division (Prowse 1957; 1958; 1960). Some studies reported that phytoplankton exist in Malaysia; thus extensive studies were conducted to determine the relationship of the water quality with phytoplankton such as the periphyton community and diatoms (Khan, 1985; 1990; 1991). Other studies explored the structure and species composition of periphytic algae and their relationship with the water quality in the Sungai Pinang basin (Maznah & Mansor, 2002).

Several recent studies were performed to assess the eutrophication status of 90 lakes in Malaysia, and showed that 56 (62%) lakes were eutrophic or in a poor condition, and requires immediate rehabilitation and restoration. The other 34 (38%) lakes were classified as mesotrophic. The most common phytoplankton in Malaysian lakes is cyanobacteria, and *Anabaena* is responsible for most cyanobacterial blooms (Tisdale, 1931; Chen et al., 2005; Fatimah et al., 1984). Most studies on Malaysian lakes reported the existence of several types of freshwater algae, these most common types of freshwater algae illustrates in Table 2.1.

Table 2.1 Common Algae Existing in Malaysia

Algae	Region	References
Periphytic algae Diatom	Pinang River	Wan Maznah and Mansor (2002)
Euglenophyta Pyrrhophyta,		Chong(2002)
176 species recorded 65 unidentified	Teluk Bahang	Yasser(2007)
7 genera Cyanophyta 22 genera Chlorophyta 8 genera Bacillariophyceae 2 genera Chrysophyceae, 3 genera Euglenophyta 1 genus of Pyrrhophyta	Paya Bungor Lake	Fatimah et al., (1984)
Diatom	Malaysian State	Patrick (1936)
periphyton especially diatom	Sungai Linggi basin	Khan (1985, 1990, 1991)
periphytic algae on stony substrates	Maliau River systems	Anton et al(1998)
17 genera Diatom 5 genera Cyanobacteria 4 genera Chlorophyta	Gunung Stong, Jeli, Kelantan	Faradina Merican, Wan Asmadi W A,Wan Maznah W O and Mashhor M (2006)
3 genera Chlorophyta (*Cosmarium, Closterium, Eustrium*) 8 genera Bacillariophyta(*Navicula,Synedra, Diatoma,Nitzschia,Fragilaria,Gomphonema, Tabellaria and Cymbella*) 1 genera Cyanophyta (*Oscillatoria*)	Kinabalu Park Sabah	Maznah and Mashhor (1999)

2.3.1 Freshwater Phytoplankton in Putrajaya Lake

Our research were agreed with previous research about the main division found at Putrajaya Lake, there are Bacillariophyta (Diatoms), Chlorophyta (green algae), and Cyanophyta (blue green algae) Research reported that most freshwater algae found in Putrajaya Lake mostly contain divisions of Cyanobacteria (28%), Chlorophyta (26%), Pyrrophyta (18%), Chrysophyta (17%), and Bacillariophyta (11%) as reported by (Sorayya et al., 2011). In this study, we selected three of these common divisions of freshwater algae including Bacillariophyta, Chlorophyta, and Cyanobacteria which describe in more details below:

a. Bacillariophyta (diatom)

The diatoms are considered single celled microscopic algae which commonly distributed in diverse water ecosystems such as lakes, rivers, oceans, wetlands and even soils. They are rejoining quickly with environmental change because their ability of immigrating and replicating in rapid fashion. Actually, diatoms are used to gather information changes in pH and nutrient status in lake sediments, and also to detect climate change. They are also used widely to deduce water quality in current marine systems. Many types of diatoms were found in Putrajaya lakes but we selected one gens from this division which is *Navicula*.

b. Chlorophyta (green-algae)

Chlorophyta is responsible for the unpleasant taste and odour of drinking water. This division can clog filtration equipment, and it can decrease the oxygen supplier for other organisms by forming scums when it is population increased in water source. Green algae are more common in brightly lit aquariums than in gloomy ones, and perhaps considered as a sign of good environmental conditions, green algae is a food for many freshwater fish and invertebrates thus occasionally a planktonic green algae bloom turns the water green. Also we used one type of this division in our research which is *Scenedesmus*.

c. **Cyanophyta (blue green algae)**

Cyanobacteria are colonial and filamentous photosynthetic organisms spread in irritating organic waters, wetlands, and soils. Some large forms of cyanobacteria are development quickly in certain conditions such as high temperature session with rich organic waters; it may be appear in water blooms or red tides. Also, cyanobacteria produce many harmful substances for zooplankton, mollusks, fishes and other marine organisms including neurotoxic alkaloids and hepatotoxic peptides. Table 2.2 illustrates some of toxin produced by cyanobacteria. In this study we selected three spices of this division because it is highly effects in water quality.

Table 2.2 Toxinsand Acute Effect of Cyanobacteria

Toxin	**Acute Effect**
Saxitoxin, Neosaxitoxin	Neurotoxicity
Nodularin	Hepatotoxicit
Microcystin	Hepatotoxicit
Cylindrospermopsin	Hepatotoxicity, renal toxicity, chromosome breakage, aneuploid
Anatoxin-a	Neurotoxicity

2.4 Algae Recognition Process

Over the last decades, only traditional methods were used to recognize and identify each individual type of algae, including water sample collection, slide sample preparation under a microscope, and algal type identification under a microscope by a human expert. Unfortunately, these methods are tedious, time consuming, and subject to human error.

Recently, with the rapid evolution of technology, computers and workstations become powerful enough to analyze and process huge amounts of data. Computer vision and image processing can now perform most conventional processes that depend mainly on human experts. Image processing using standard scientific tools and image processing techniques are now applied in virtually all natural sciences and technical disciplines (Jähne, 2002). Computer-based image-processing approaches are widely applied in solving many problems in biology and other fields. Extensive studies have been conducted to develop a computer system that can mimic the conventional approach in detecting and identifying different algal types found in microscope images. In Table 2.3 comparisons between manual and automatic techniques.

Table 2.3 Comparison between Manual and Automated Recognition Processes

	Manual Recognition	**Automated Recognition**
Accuracy	Subjected to human error	Subjected to approach and techniques used
Speed	Subject to the experts knowledge	Subjected to the training set of objects, image resolution, and system complexity.
Reliability	Subjected to experts knowledge, and equipment use.	Subjected to the features selected, and training results
Cost	Subjected to prior knowledge, experts efforts.	Subjected to software cost only. And power consuming of devices.

Advantages of using computerized processes for automated recognitions:

- Automated identification, classification, and recognition processes are always faster than conventional methods.
- Computer calculations are often more accurate than human ones. The accuracy of the manual process is subject to expert knowledge and human errors.
- The cost of a recognition process using manual processes is higher that of using computer programs. The latter incurs costs for only one time, and the former incurs costs hourly.
- The cost of learning and training is always cheaper using a computer than employing humans.
- The automated recognition process is easier to use, more convenient, and more efficient than the manual process.
- Automated recognition processes support digital documentation, which eases the searching process and documentation.

- The manual recognition process is constrained in certain environments and cannot be used for online data monitoring, whereas automated methods can provide a real estimation of the monitoring process.

2.4.1 Computer vision and pattern recognition

Computer vision is a field that used for developing several methods in processing digital images including acquiring, processing, analyzing, and understanding images. It also generally involves analog data from the real world to produce numerical or symbolic information for digital representation. Computer vision technology covers many field of automated image analysis to provide a robotic guidance for industrial application. Computer vision is connected with some other field to link the theory of computation with the particular works such as artificial systems to develop specific application for processing image information. Image data come with different forms, such as video sequences, views from multiple cameras, or multi-dimensional data from a medical scanner (Shapiro & Stockman 2001; Morris, 2004).

The classical problem of image processing is to determine whether image included specific information or contains some specific objects, features, or activities. Identification task is solved automatically by using digital image processing field with less human effort. Many approached and methods have been designed to deal with these type of problems including (recognition of simple geometric objects, matching human faces, matching biometric objects such as iris and finger print, and OCR application for handwritten), other methods are focused more in enhancing images, defined area, background, and pose of image object.

Different methods applied to solve the recognition problem for most existing systems. Early methods were involved with object recognition, where matching techniques is used without any learning criteria, it recognizes by matching the template of object. However previous methods were not accurate enough in recognition process. Furthermore other methods involve in identification approach used to identify individual components of image object such as face, finger, and iris techniques. Finally, recent techniques which involves in identifying and detection approaches by using learning features algorithms for recognition purposes. Last types of methods are showed a great accuracy and performance in recognition and classification images.

Computer vision system is extremely application dependent on other area to accomplish recognition tasks. Specific implementation of a computer vision system depends mostly on specific functionality. In this research we choose the recent techniques of recognition and identification approaches which describes briefly in the following section.

- **Image acquisition**: Digital images are produced by several image sensors including range sensors, tomography devices, radar, ultrasonic cameras, etc. the process of transferring the analogue image into digital images performed by using one of more types of previous sensors. The process of image acquisition is required to transfer the analogue images into digital forms. The main unit of digital images is pixel which represent the values correspond to the light intensity in one or more spectral bands.
- **Pre-processing**: captured images is usually suffering from different problems such as varying in colors, darkness, and brightness because of many factors influence such as light source, lens, and method of capturing. The process of preprocessing performs on the image data to assure good appearance and clearer details. Some methods apply such as re-sampling image coordinate system to endure its

transforms correctly, Reducing sensor noise to introduce proper information, and contrast enhancement to make relevant information clearer and detectable.

- **Detection/segmentation**: Most images contain different objects; it is rarely to obtain microscope images with one object only. Some method of image processing is used to isolate image into different object components. A detection process for specific image points or regions performs to determine the object for further processing. This process of isolating and dividing the image into a sub-image, where each image contains at least one object is called segmentation.
- **Feature extraction**: there are a huge number of data can be derived and extracted from the digital images at various levels of complexity including lines, edges, ridges, corners, and blobs. There are several categorizes of feature such as geometrical, shape, texture, and color feature. Feature extraction is the process of selecting the suitable parameter and values of image to be used for identification process of image objects.

- **High-level processing**: This type of processing is typically performed on a small set of data, e.g., an image region that contains a specific object. Examples of this processing are verification data, estimation of object measurements, classifying and detecting objects into different categorizes, as well as comparing and combining two different views of the same object.
- **Decision making:** This process of making final decision about the input images is required for most recognition application such as automatic inspection results true or false, recognition application match or not, and for security and recognition application a flag of matching.

2.4.2 ANNs for the recognition process

ANN is a mathematical or computational model designed to simulate the structure of biological neural networks, and learning functions. A neural network is a collection of different artificial neurons groups connected together. ANNs are usually used to organize the complex relationships between inputs and outputs in mathematical models, and to find the essential patterns in data. Ability to learn is the most attractive feature of neural networks. Learning by ANN means using a set of observations to find the best function that solves tasks in some optimal sense. There are three major learning methods used in ANNs and each one corresponds to a particular abstract learning task. These methods include supervised learning, unsupervised learning, and reinforcement learning, and there are many types of artificial neural networks (ANN) developed to support these methods which designed specifically to mimic real human behavior of neurons and electrical messages. The input of neuron can be considered as eyes or even nerve, processed by brain to take the output decision. Some other types of ANNs called adaptive systems used to model things such as environments and population. Common types of neural network and recognition approach describes below in some details.

Feed forward neural network

It is considered as one of first ANN which designed with most simple type of artificial neural network. It can be constructed with different unites without loop or cycles. The information moves in one direction from input nodes through hidden layer to the output nodes. Sometimes it constructed with back propagation algorithms to support training and learning objectives. It designed with single or multi-layer based on layer architecture of application. Figure 2.1 illustrates basic diagram of this type of ANN (Schmidhuber, 1989).

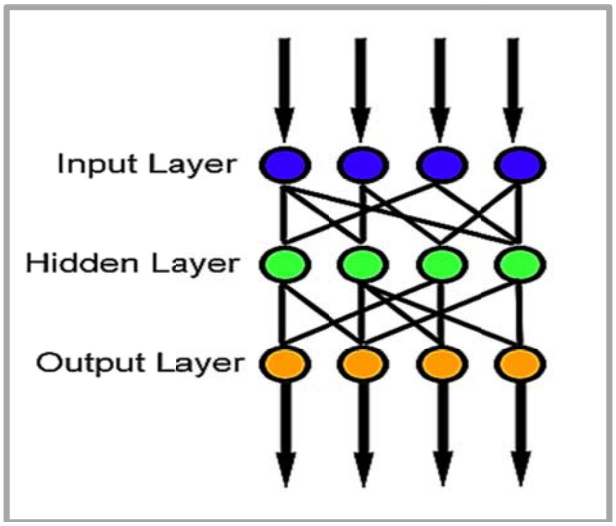


Figure 2.1. Basic Diagram of Feed forward neural network.

Recurrent neural network

The recurrent neural network (RNN) is type of neural network developed to create internal connection state between networks unite as a directed cycle. It is used mostly for unsegment tasks such as handwriting recognition which achieved best known results (Grayes et al., 2009). It designed with many type based on application requirements such as Bi-directional RNN, Continuous-time RNN, Hierarchical RNN, and Recurrent multilayer perceptron.

Hopfield neural network

A Hopfield network is developed by John Hopfield which considered as one form of recurrent artificial neural network. It works as content addressable memory with binary threshold unit to provide a model for human memory understanding. Figure 2.2 shows simple diagram for this type of ANN (Hopfield, 1982).

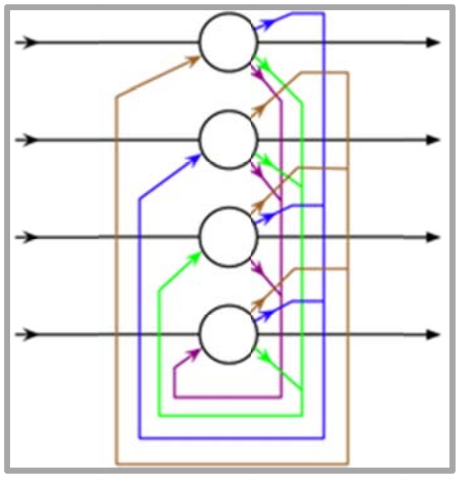


Figure 2.2. .Hopfield neural network diagram.

Radial basis functions

Radial basis functions are considered one of most powerful ANN techniques which used for interruption the information in multidimensional space. A RBF is designed based on a distance criterion between the centers with respect to branches. It used in neural network area to replace the sigmoidal hidden layer in multi-layer perceptrons. RBF networks achieved process in two phases, the first phase where input is mapped onto each RBF in hidden layer, and the second phase the mean predicted outputs calculated using linear combination of hidden layer values. RBF is not suffering from local minima as found in Multi-Layer Perceptrons because only parameters that are adjusted during learning process are mapping from the hidden layer to output layer. However it required good coverage of input space by using radial basis function (Yee et al., 2001). Figure 2.3 illustrates examples for this ANN architecture.

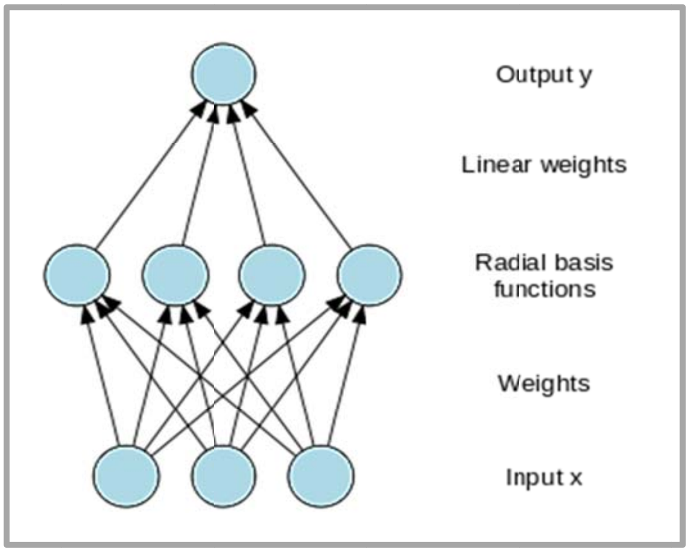


Figure 2.3. Simple example for RBF ANN.

Cascading neural networks

The Cascade neural network is a supervised learning algorithm designed with multilayer structure by Scott Fahlman for adjusting the weight of neuron and to train and add automatically the hidden layer. It has many advantages such as learning very quickly, automatic adjusted for size and topology, and retains the network structure if training set have been changed, however it required more complex feature with extra time for training and learning process (Fahlman & Lebiere, 1990).

Support vector machine (SVM)

Support vector machine (SVM) is a supervised learning methods proposed for statistics data analyze with computer science to recognize patterns based on classification and regression analysis. A set of inputs is process by standard SVM to predict the outputs by using probabilistic binary linear classifier. It has two main types linear SVM and nonlinear SVM, where each input data is classified in two or more categorizes based on the data redundancy. It is an efficient methods used widely for most of pattern recognition system which showed high accurate results in classification and recognition process (William et al., 2007).

2.5 Existing Studies for Automated recognition Algae

Based on our review of most existing systems developed for algal recognition, we found that there are two different types of algae systems. The first type which used to estimate of the algal density by counting and calculating the objects found in collected water samples. The second type of system involves with identification of the object itself in given water samples based on taxonomic characteristics. Actually, both approaches are important to indicate the water quality of freshwater lakes using computerize methods for the recognition process. Extensive studies have been conducted on computer-based approaches combined with image processing, ANNs, genetic algorithms, and fuzzy logic to develop computer software that can detect, count, identify, and classify types of algae. Some of these studies were efficient with 90% accuracy (Jefferies et al., 1984; Katsinis et al., 1984). Other studies were also used to determine the shape, size, volume, and other features of diverse microorganisms with great accuracy (Estep et al., 1986). At the end of the 1990s, image processing combined with some other techniques such as fuzzy logic, genetic algorithm, and ANN began to be used for better accuracy in the classification and recognition of microorganisms. Most developed tools are used for different purposes, such as online monitoring and density measurements of microorganisms in water. Other tools were used to assist in the recognition process, such as image enhancement, noise elimination, edge detection, image extraction, and segmentation (Kamath et al., 2005; Junna et al., 2009). Other techniques used image processing with genetic algorithms or ANN to improve the accuracy of recognition process (Schultze-Lam et al., 1992; Blackburn et al., 1998). The following sections explore most existing systems developed for algal recognition.

Previous studies reported that the conventional method of counting and identifying different algal types by microscopy is a time-consuming process that requires substantial specialist knowledge and inevitably subjected to human error (Simpson et al., 1993). Recently, considerable efforts were exerted in producing a computer system that enables the automated analysis and identification of algal samples. Basically, we categorized most existing systems developed for algal recognition into three different types based on system functionality and components. The first type, which was developed from the early 1980s to nearly the beginning of the 1990s, mostly used image processing techniques with some geometrical measurements for algal shapes. The second type of recognition system was developed during the mid-1990s and used image processing combined with ANN and some basic features of extraction algorithms. Finally, the third type of these developed systems can be considered the modern system because it uses all available technology for the automated recognition of algae, such as supervisor machine learning, fuzzy logic, expert system, and genetic algorithms approaches. This type of system was developed within the last 10 years, and showed great accuracy in the classification and identification of many types of algae.

2.5.1 Early recognition system

Early research on this area was started in the beginning of the 1980s by Jeffries et al. (1980; 1984), who used an Eclipse SA40 with six satellites and a Colorado Video frame grabber to identify some types of zooplankton by analyzing some object parameters, such as length, width, perimeter, and area. They reported that their procedures can identify eight taxonomic groups with 89% accuracy at a speed of about 35 organisms per minute. Dietrich & Uhlig

(1984) interfaced the Quantimet system to a Digital PDP 11/23 computer to measure the area, length, width, and ratio of length to width of *Artemiasalina* for biomass termination, and then used those parameters to classify and count different stages of mass-cultured. Estep et al. (1986) interfaced a Macintosh computer to image analysis computer to define the shape, volume, size, abundance, and surface area of a variety of organisms ranging from bacteria to fish. At the end of the 1980s, Gorsky et al. (1989) developed an image analyzer system to automate the process of identification of three types of algae. The developed system was evaluated by comparing its results of counting and measurements with those obtained by visual analysis. They reported no significant difference between both sets of results. They also reported that if the image resolution can be improved, their system can identify 26 types of algae instead of only 3. Actually, they used the size and shape in the measurement process, which limited the identification process.

Most developed systems before the 1990s can be considered as simple image processing systems with many constrains and limitation. These limitations existed because most of them depended on some basic calculations of geometrical parameters for algal images. The image processing field cannot be employed to develop full automated systems to process the identification of algal groups without combining them with other computer science areas, such as ANNs, fuzzy logics, expert systems, and genetic algorithms. The process of detecting and recognizing objects in a given image requires certain intelligent processes analogous with human brain activities. Artificial intelligence technologies evolved widely in the early 1990s, and are involved in solving most critical scientific problems.

2.5.2 Moderate generation of recognition system

In the early 1990s, image processing combined with artificial intelligence was employed to develop useful methods for detecting, recognizing, and classifying image objects. These new techniques improved the image processing system performance, such as accuracy, speed, and reliability. The reusability of the object-oriented system was used to enhance the developed system with less time and effort. Over all, new technologies have offered new tools and techniques in designing an appropriate system for solving existing problems.

Simpson et al. (1991) began to develop an automated classification system for detecting some species. They were the first to use image processing with ANN techniques in the classification process of certain types of algae. They worked to improve the detection of biological images using some pattern recognition methods with ANNs for identifying biological objects found on digital images. They proposed an image processing system with a neural network model to analyze some plankton data derived from previous counting techniques. The backward-error propagation method with three layers was used in the training mode of learning, and they showed that a neural network with two layers of weights was also able to learn a large data set by the significant results achieved in separating novel images of two co-occurring species of the *Ceratium* domain (Simpson et al, 1992). The proposed method was used to classify some plankton types, such as *Ceratium* class Dinophyceae for evaluation purposes (Simpson et al, 1993; 1994). One of Simpson's colleagues, Culverhouse, attempted to improve their proposed methods by evaluating the proposed system by including another kind of plankton. He extended the ANN classification model to improve system performance by increasing the extracted parameters of texture images. The developed system was able to classify the majority of three toxic and noxious phytoplankton blooms (Culverhouse et al, 1996). Later on, Culverhouse et al., (2003; 2006) performed several studies to improve the accuracy of their system. A software system named DiCANN developed to demonstrate the feasibility of applying ANN pattern categorization methods to the laboratory identification of toxic and noxious dinoflagellates. DiCANN is considered as modern application for laboratory pattern recognition system which developed to categorize various marine HAB Dinoflagellate specimens automatically. DiCANN system is able to classify of 23 species of Dinoflagellate from microscope images successfully with responsible accuracy. Calibration techniques are then developed as standards for this new class of marine observation method. DiCANN included many functions such as internet distributed database, advanced image analysis techniques with ANN to perform object identification and categorization.

Boddy et al. (1994) performed an extensive study on the identification of 40 marine phytoplankton species from different taxonomic divisions. Image processing methods with two different approaches were adopted for the recognition process based on flow cytometric data of species such as integral fluorescence, horizontal and vertical forward light scatter, and time of flight. A back propagation neural networks with single hidden layer was trained to distinguish species by using patterns recognizing based on their flow cytometric signatures with different data testing sets. The first approached employed a single layer of ANN to identify the major taxonomic group based on cell fitted. The second approach used a large ANN architecture to discriminate some of the major taxonomic groups. They reported that cryptophytic species were identified successfully and half of other groups were identified reliably in using a single-layer network, whereas all other species were identified almost well. They concluded that the application of neural computing techniques to identify large number of species must be represented well, and preliminary studies should be considered and integrated for further development.

Early phase of ADIAC project led by du Buf & Bayer (2002), was started on May 1998 and taken for three years later. They perform experimental study based on the application of image processing and pattern recognition tools to automate the identification process of diatoms using computer processing. The ADIAC was considered as innovative system that designed to identify diatoms by using feature and information of image including taxonomists, shape, ecologists, and ornamentation of consortium. Image database were captured and processed by experts. Then feature extraction for identification is performed to achieve 90% of recognition rate which considered good results in identification of some species divisions if compared with other study results.

2.5.3 Recent Recognition System

Modern recognition systems were developed by applying image processing techniques with several artificial intelligent approaches to improve the identification methods for objects in given images. The Matlab software is an essential programming tool for most scientific studies, and has been used to solve complex problems efficiently, especially in the image processing field. Matlab used mostly in the system development of an image recognition process because it has an integrated technical computing environment suitable for algorithm design and development. It is also a high-level programming language that includes hundreds of built-in functions that can be reused to support the development process for such applications (Gonzalez et al., 2004).

Furthermore, Embleton et al., (2003) developed a computer application by using image processing with pattern recognition methods to identify, count, and measure selected groups of phytoplankton automatically lake LoughNeagh in Northern Ireland. Some image processing techniques are used to isolate, and measure features of phytoplankton images. A combination of ANN with a simple rule-based procedure was used for measurements specific object features to identify and classify the selected samples of phytoplankton. The obtained parameter of measurements included 74 parameters for all four phytoplankton groups, which were stored in a database for later use. The developed system was trained with 75 image samples for each individual type, and then tested over the total volume of image samples. A comparative analytical method was performed on both manual and automated identification processes to obtain system accuracy. Their experimental results showed that automatic system was within 10% of manual detection process over the total estimated cell volume. They reported that results of their system were close to fit with the manual process. Some variations between both manual and automated processes are found however the automated process was reported as faster and accurate in counting the total cell volume. Finally, they reported that developing a computer system for the automated identification process was visible, and the accuracy and speed of the automated process was efficient compared with conventional methods.

Tang et al., (2006) proposed a prototype system that included several descriptors for extracting shapes and features, and used a normalization multilevel dominant eigenvector to extract the best feature set for the binary images of selected plankton. They combined the new shape features with common shape features used to produce a compact feature vector for the classification process. For feature extraction, they used several existing methods such as Fourier descriptor to normalize the features around the centroid of the object. Moment invariants were used to compute the invariants of rigid objects. A granulometry was used to extract the size distributions in binary images. Circular projection was used to reproduce the longest linear structure of the object and the smoothness of the kernel boundary. They also calculated the object width and density. Finally, they used principle component analysis (PCA) for the feature combination and normalization process. PCA was used due it is ability to reduce the feature-selecting process, and its capability to compact useful information into dominant features. They used 3147 binary image samples of seven plankton classes for their experiment. They used each individual feature vector for first-stage classification; however, the accuracy obtained was below 65% because of the large differences that occurred within each class of plankton image. They developed a new algorithm called NMDEE to combine all the long and short feature vectors of plankton images. After applying their algorithm, they found that the proposed system accuracy improved to 91% in the classification process.

Sosik & Ropert (2007) developed an analysis and classification approach to increase the ecological insight that can be obtained from rapid automated micro-plankton imaging systems. Machine learning algorithms for the classification process of several phytoplanktons were developed by using a combination of image processing with neural networks to categorize 22 types of phytoplankton. The developed system depends mainly in preprocessing techniques, and feature extractions. Image preprocessing used in developing this system such as edge detection, morphological operation, boundary representation. In addition, extract procedure for several types of features was designed including size, shape, symmetry, texture characterization, invariant moments, diffraction pattern sampling. Extracted features combined the selected features in a scalar feature vector for training purposes and machine learning. Co-occurrence matrix statistics results showed that system classifier was able to categorize all selected phytoplankton with 88% recognition rate. This approach is used to provide taxonomically resolved estimates of phytoplankton abundance with fine temporal resolution.

Verikas et al., (2010) performed another study for automated detection and recognition process of phytoplankton species. They concerned mainly on the development of algorithms for the detection of objects in phytoplankton images. They selected *Prorocentrum minimum* objects which representing one of invasive species and considered as one of known harmful blooms in estuarine and coastal environments. They developed novel techniques to combine many modules in one system including image segmentation, congruency-based detection of circular objects, and stochastic optimization. Their experimental results showed that automated object recognition of phytoplankton is possible by using images processing techniques. Their system recognition rate was 93.25% of objects representing one phytoplankton only where *P.minimum* cells were detected accurately with this technique.

Luo et al., (2011) developed an image-processing system with pattern recognition using MATLAB 7.0 for the automated identification of seven types of circular diatoms based on texture features. Many steps were performed to achieve their goal, including the application of a canny edge detector, image segmentation used to find the location of diatoms, eigenvector obtained by applying Fourier spectrum features, and BP neural network used to classify the circular diatoms effectively. During their system development, they focused more on extracting the varying features of diatoms, which can be used to improve the training set of neural networks. They reported that their system obtained a promising result with 94.44% accuracy from 12 species of circular diatoms. They reported that circular diatom identification using microscopic images approach is potentially applicable in the future automated identification of microalgae in the field of phycology.

More recent, Dimitrovski et al., (2012) proposed a new approach to classify specific diatoms by considering the hierarchical structure of diatom taxonomy. A combination of contour based and texture based features for automated classification process was used in this study. They found that random forest approaches has better predictive performance and more efficient than SVM approaches in classification process.(Dimitrovski et al., 2012) In addition, Wu, et al., (2012) performed study to measure specific morphological features of Spirulina microalgae filaments by using some image processing approaches such as length, diameter, width, and degree of spiralisation to improve the algae production. Experimental results showed that their algorithms can be considered as optimal if compared with manual approach where the means error between manual and automated approach were 4.8% for length, 5.6% for diameter, 6.2% for width, and 4.7 for spiralisation degree. However, they found their system is faster than manual approached in measurements process; it takes about 30 seconds while a manual approach takes about 5 minutes approximately (Di Wu, et al., 2012).

2.6 Problems of Current systems

Current methods for the automated identification of algae, including absorption spectroscopy, fluorescence spectroscopy, liquid chromatography, flow cytometry, and molecular genetic techniques, are not only tedious but also solely depend on the physiological state of algae under low resolution.

Unfortunately, most developed tools are designed for a specific division of algae, such as plankton, due to the difficulties in implementing a system that can detect all algae divisions. These difficulties can be attributed to the variations that can be found in the algal shapes for each division (Wilkins et al., 1999; Yao et al., 2007). Studies on the automated identification of tropical freshwater algae to identify specific groups or either certain species are also limited. There are many constraints and limitations in developing an application that can identify and classify all algal divisions in a certain area. However, the development of a process that is standardized and can integrate all tools for building successful applications that can perform automated recognition for some algal division should continue.

One of the most common problems in developing tools is the issue of accuracy, which is dependent on the specific parameters used in the design of a system, such as the number of species included in the recognition process, selected features extracted for training process, number of objects found in the sample image, and image processing techniques used with other techniques. Most developed systems have variations in accuracy ranging from 50% to 80%.

Finally, the identification process time can be considered as an essential problem for these applications due to the length of time required during training and during the detection process. System performance is constrained by time, which has a close relation with the number of algae included in the recognition process. Most developed systems require time to train a set of features, and still more time to identify the objects inside images. Both training and recognition times vary from 5 min to 3 h based on several parameters, including image resolution, extraction method, training set of image data, computer performance, and type of ANN used and its components. However, the processing time consumed by automated recognition methods can be considered as an advantage compared with manual recognition methods.

2.7 Chapter summary

In this chapter, many concepts related to our thesis have been discussed, including water quality and importance of water for all organisms, which are described in the main parts of this chapter. Several issues on algae and their effects on water quality are also discussed. A list of extensive studies on algae division is also presented. The next section describes the water resources and the common freshwater algae found in several lakes in Malaysia. Comparison between advantages and disadvantages for both manually and automatically approaches of Algal recognition. Computer vision with ANN and other technology approaches are also involved in the recognition of algae division. Finally, an extensive review of most existing systems is presented in this chapter with the problems encountered in these systems are also discussed.

CHAPTER 3

RESEARCH METHODS

3.1 Research Materials

Equipment and devices used to accomplish the initial tasks of algae image preparations are described in more details in the following sections.

3.1.1 Plankton nets

Plankton nets with handles are mostly used for collecting plankton. The net is made of precise, fine polyamide material mounted onto a round metallic frame with a handle. The mesh size of the netting can vary and has a standard depth of 70 cm. It transforms into a U shape if the upper and lower portions have the same width or a V shape if the net is narrowed toward the lower portion shows in Figure 3.1(a). The following steps were performed to obtain the samples.

- The net was lowered vertically into the water until the bottle was filled.
- The plankton net was slowly raised from the water.
- The filtered water was used to wash all plankton retained on the inner surface of the net into the bottle.
- The water from the plankton net bottle was divided into different small bottles

(a) Plankton Net **(b) Plastic Bottles**

Figure 3.1(a) Plankton Net & (b) Plastic Bottle

3.1.2 Slides and cover slips

The slides and cover slips were prepared as follows.

- The slides and cover slips Figure 3.2 (a) were thoroughly cleaned, dried, and ensured of being free from dust, debris, and grime because it touches the object being observed and has greater potential to contaminate the specimen if careful handling is not undertaken.
- The flat slide was placed on a clean, dry surface.
- A few drops of the sample were obtained using plastic pipettes Figure 3.2 (b) (sample taken from a clear surface). A small amount is collected from the green area (sample taken from the bottom) with a pair of tweezers Figure 3.2 (c) and placed on the center of the slide.
- One drop of liquid sample was squeezed out onto the direct center of the flat slide.
- The cover slip was gently lowered onto the flat slide. One edge of the cover slip was placed down first before lowering the rest. The cover slip must not be pressed down once it is in place. The slide and cover slip combination was picked up and gently placed on the viewing tray of the microscope.

(a) Slide with Cover **(b) Plastic Pipettes** **(c) Tweezers**

Figure 3.2 Materials used during preparation process of algae slides

3.1.3 Microscope device

The microscope is a device used to enlarge the views of small objects not visible to the naked human eye. The microscope is a device which has two essential elements, a primary magnifying lens and a secondary lens system. In our study, we used an electronic microscope model MTC#BI-220ASA, as shows in Figure 3.3. Prepared slides were placed under microscope lenses with magnification powers of 10×, 20×, and 40×.

Figure 3.3 Picture for Electronic Microscope used in This Study

3.1.4 Microscope Camera

The AM423X Dino Eye digital shown in Figure 3.4 is used to acquire the algal image from the microscope into computer storage. This camera was selected because of its unique design, allowing users to fit it into most microscope eyepiece slots. The capture images were obtained with a resolution of 1280 × 1024.

Figure 3.4 Picture for the Dino-Eye Camera that used in this study

3.2 Study Area

Putrajaya Lake is a man-made freshwater lake that covers an area around of 650 ha, and is located at the new capital city of Malaysia known as Putrajaya. The lake was constructed to provide a landscape feature and varied recreational activities for the city population, as well as to create wildlife habitats (Shutes, 2001). Putrajaya Lake is warm polymictic, oligotrophic to mesotrophic, and is located at the south of the densely inhabited Klang Valley in Malaysia. Major inflows from upstream outside surrounding areas contain certain level of pollutants. Nutrient loading at the lake come mainly from non-point sources. These include the use of agrochemicals, fertilizer, land clearing, and soil leveling at the surrounding areas. The Putrajaya Lake Catchment Figure 3.5 is a small river catchment with an area of about 52.4 km^2 located in the middle of the Sungai Langat River Basin.

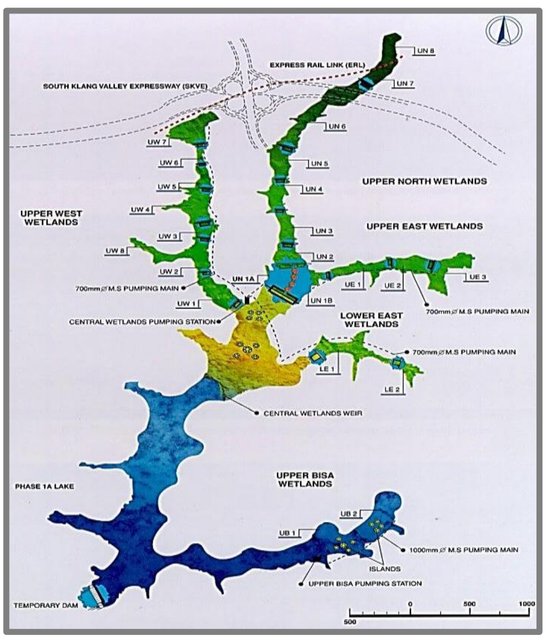


Figure 3.5. Putrajaya Lake Catchment

The wetlands are the largest freshwater source in the tropics (Perbadanan Putrajaya and Putrajaya Holdings SdnBhd, 1999). Putrajaya Lake includes wetland cells adopted with a multi-cell design strategy as shows in Figure 3.6; it comprises six defined wetland arms and lakes such as Upper North, Upper West, Upper East, Lower East, Upper Bisa, Central Wetland, and Putrajaya Lake (1998). Putrajaya Lakes and Wetlands water were classified to indicate that the water is relatively clean based on oligotrophic to mesotrophic.

Figure 3.6 Multi-cell layout Map for Putrajaya Lake.

The images of freshwater algae that used in this work were captured from water samples collected from different locations at Putrajaya Lake, Malaysia. In this study, we selected five common species from the three main divisions found in Putrajaya Lake as preliminary study for classification process of freshwater algae. The genera of selected algae in this research are *Navicula* from Bacillariophyta division, *Scenedesmus* from the Chlorophyta division, *Chroococcus*, *Microcystis* and *Oscillatoria* from the Cyanobacteria division. These types of algae used in this study are described below briefly.

(a) Scenedesmus

Colonies are formed by the lateral joining usually with 4 or 8, or rarely with16 cells. It has spines in the terminal cells as shows in Figure (3.7-I).

(b) Chroococcus

Microscopic colonies usually found in colonies of two, four, or eight cells with a transparent protective covering sheath containing photosynthetic pigments shows in Figure (3.7-II) (Davidson,2003).

(c) Oscillatoria

Filamentous unbranched, the single cell with cylindrical shape, occurring singly or in colonies shown in Figure (3.7-III) (Tiffany & Britton 1971).

(d) ***Navicula***

Single cell with boat-shaped the central area distinctly expanded with acute, rounded or capitate ends the cell content lines arranged parallels to the apical axis as shows in Figure (3.7-IV) (Tiffany & Britton 1971).

(e) Microcystis

The single cell with round-shape or oval usually formed in large colonies that form irregularly. The cells color appears brown, black or purple as shows in Figure (3.7-V) (Prescott 1984).

I.

Phylum Cyanobacteria
Class Chlorophyceae
Genus ***Scenedesmus sp***

II.

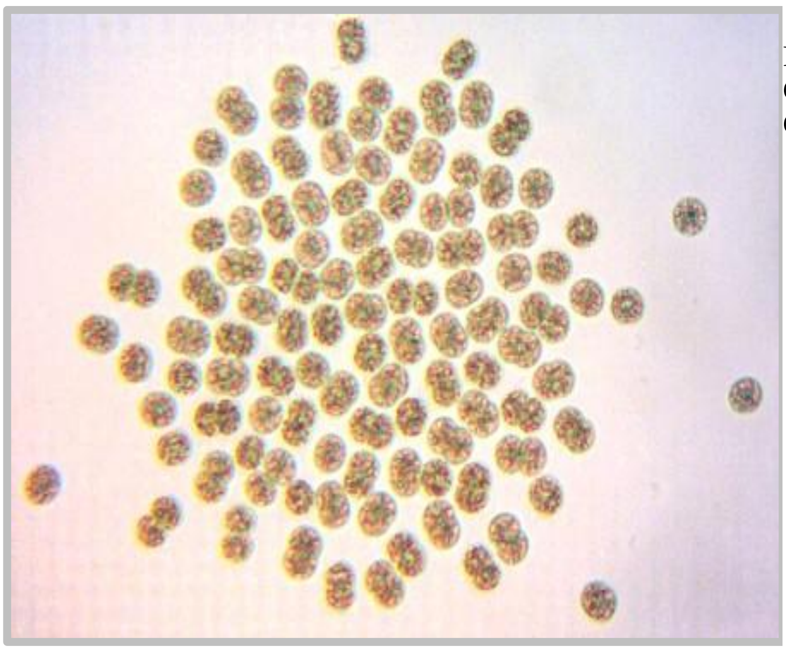
Phylum Cyanobacteria
Class Cyanophyceae
Genus ***Chroococcus sp***

III.

Phylum Cyanobacteria
Class Oscillatoriaceae
Genus ***Oscillatoria sp***

Phylum Bacillariophyta
Class Bacillariophyceae
Genus ***Navicula sp***

IV.



Phylum Cyanobacteria
Class Cyanophyceae
Genus ***Microcystis sp***.

V.

Figure 3.7 (**I** *Scenedesmus sp* , **II** *Chroococcus sp*, **III** *Oscillatoria sp,*
IV *Navicula sp,***V** *Microcystis sp*

Water samples were collected from different sampling sites at Putrajaya Lake. The water samples were analyzed and examined using an electronic microscope model no. is (MTC#B1-220ASA). A microscope eye-piece camera model (AM432X) was attached to the microscope lens and connected to a PC via a USB port for image acquisition. It was used to capture, load, and store the images directly into computer. Our data set contained four genera of cyanobacteria, and 100 image samples were collected for each genus. All sample images were divided into two groups; one used for ANN training purposes and the other used for system testing to avoid biasness in results.

3.3 Methods of System Developments

The objective of this study was to develop an automatic recognition system that can identify and classify selected algal samples. Several modules were built to develop our proposed system with a graphical user interface that helps users run system functions. MATLAB ver. 7 was used in the system development process because of its ability to integrate a technical computing environment suitable for algorithm design and development. MATLAB is considered as a high-level programming language that includes a number of integrated functions. Based on our analysis of system requirements, we found that for any image recognition system, the essential module must include image preprocessing, image segmentation, feature extraction, and classifier techniques. Our proposed system component module with flow chart diagram is illustrates in Figure 3.8.

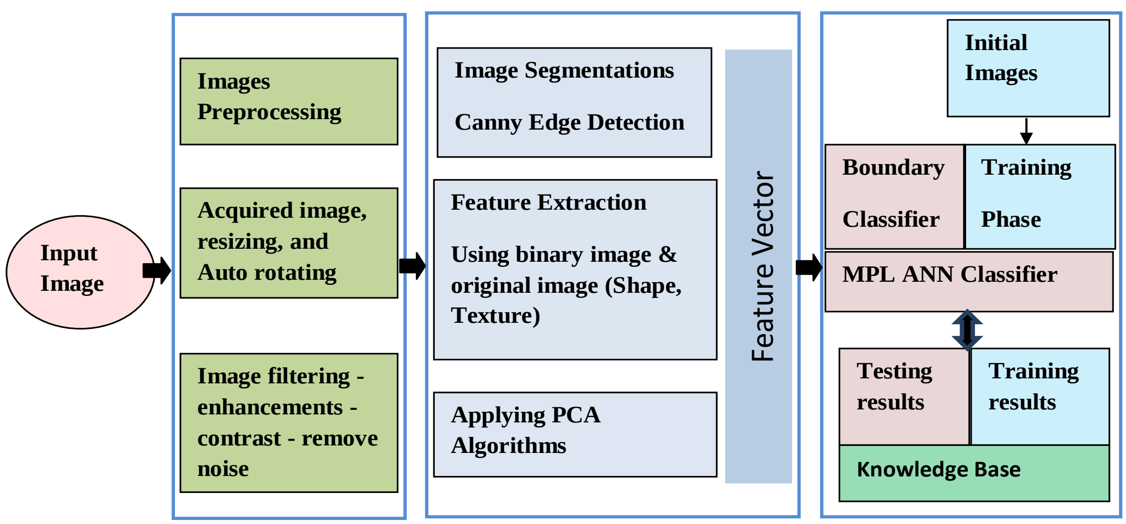


Figure 3.8. System Architecture and Flow Chart Diagram

The developed system was designed with two separate interface models; the first one was used in the preparation system database from algal images by supporting user with necessary functions in extracting features and storing it in the database file. The other interface was used in initialization network parameters to perform the training phase on the system database, enable users to upload algal images for the testing mode, and to obtain classification results automatically. Interface system built with simple, easy, and friendly appearance to make interaction process with user efficient.

Based on analysis of functional requirements of recognition system we developed our system with two individual modules. First module is used for preparation process of algae database by using a set of algae training images. This module is including all the required functions to transfer the set of training images into a vector of values, and then store it in single database file such as uploading images, image enhancements, segmentation, and feature extraction as illustrates in Figure 3.9. Second module is used for initialization the ANN parameters such as number of nodes in input, hidden, and output layers, learning rate, mean square errors, and maximum number of epoch. It also used to trigger the learning process of MPL networks by training the ANN with the stored database, and finally this module is used to display the results of classification process as shows in Figure 3.10. System has been designed with simple and easy to use interface to satisfy the biologist users. System functions have been developed using graphical user interface appearance with a simple buttons for each function to facilitate the interaction process with system. A guide about how to use our system is written and user can invoked it by selecting help in system main menu. Extra images and explanation about the developed system are described in more details in appendix (II)

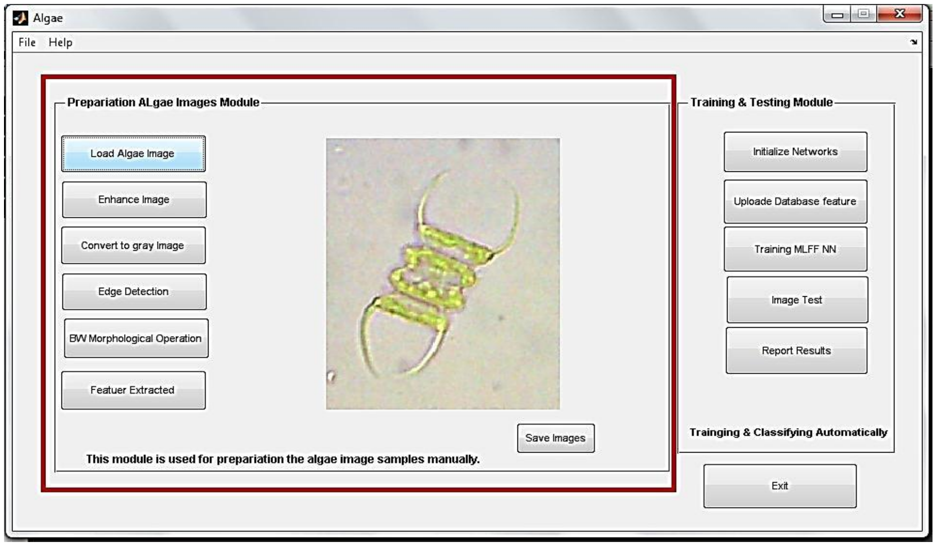


Figure 3.9 System screen snapshot with highlighting the prepration module.

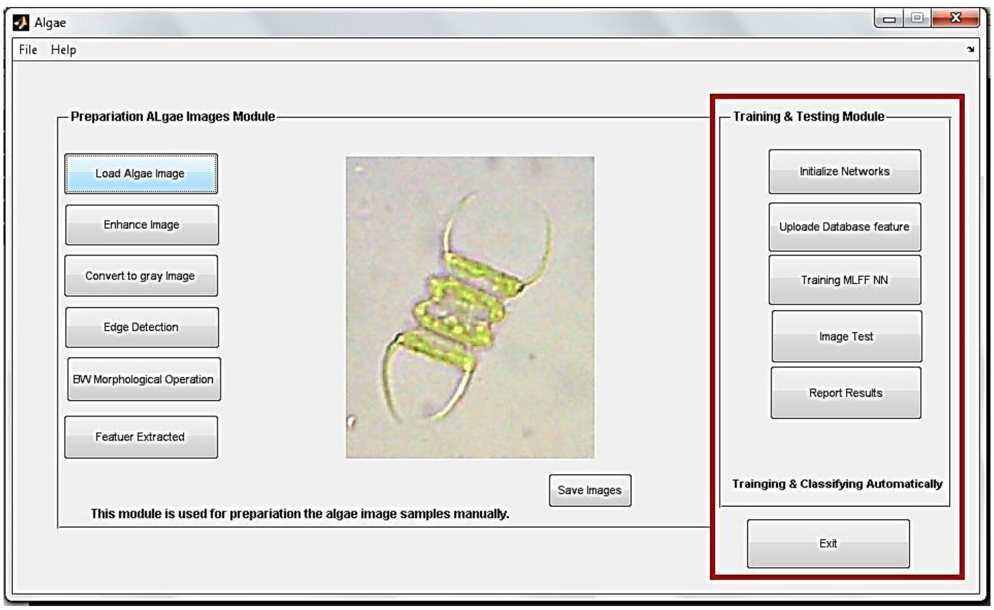


Figure 3.10. System snapshot highlighting the training and testing module.

In the system development process, each main module was divided into small units of function and procedure to perform several tasks involved in the system implementation process. The developed system has been developed with several functions where each function associated with index number as illustrated on Figure 3.11, for example uploading images is given number (1), preprocessing function is given number (2), segmentation task is number (3), extract feature is number (4), initialization and training tasks is number (5), uploading testing images is number (6), and finally result report is number (7).

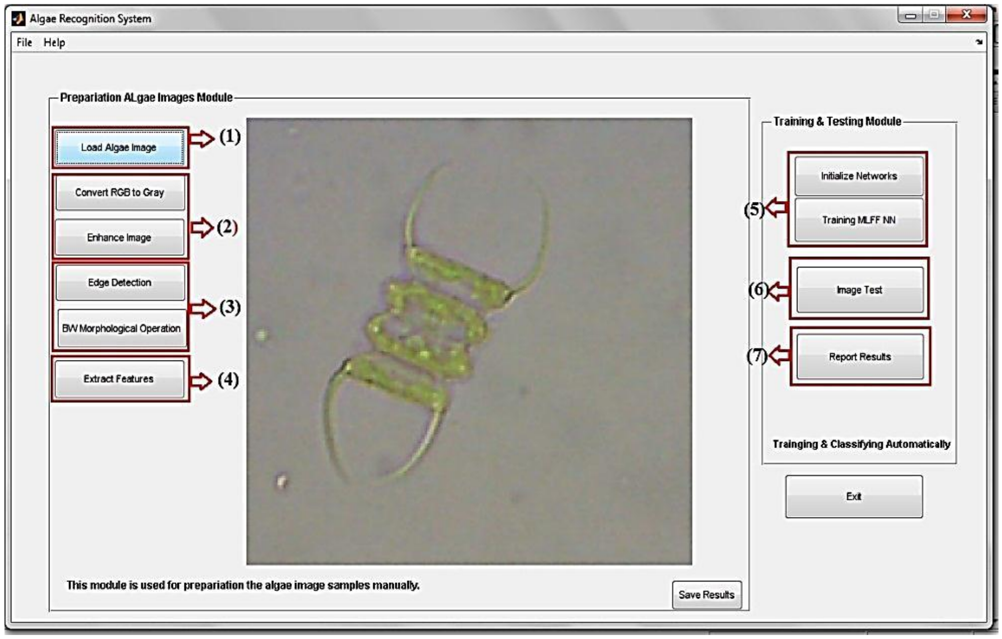


Figure 3.11. Proposed System Tasks assigning each member function.

In contrast, there are many module have been designed to accomplish individual tasks, the main module are image preprocessing, image segmentation, feature extractions, and ANN design and training parts. For example, system preprocessing is including many functions such as image acquisitions, image enhancement, image filtering, improve image contrast, and image conversion into gray scale. In the following section, we explore in more detail each step involved in our system design and implementation.

3.3.1 Image Preprocessing Module

Microscope images commonly suffer from noise and low contrast quality. Noise exists mostly due to random variations in brightness or color information produced by captured devices such as scanners and cameras. Microscope images also include unavoidable scum existing beside target cells, and some holes or small objects that strongly affect image quality. Images may contain unwanted areas and appear blurry, as shown in Figure 3.13.

For these reasons, we need to perform special procedures to reduce image noise and improve image quality. The preprocessing of captured images is a preparation and treatment process used to enhance the features of images to produce clearer details, remove noise, remove intelligibility of images, and improve the overall appearance of images. There is no specific technique for enhancement image; however, there are some common techniques used in image enhancement, including image conversion to grayscale, filtering, histogram conversion, and color composition. Image enhancement is defined as the conversion of an image quality to create a clear image and ensure the accuracy of the process of feature extraction.

In this study, we carefully selected image enhancement methods based on tried and tested methods until the best results for the dataset of images were obtained. In the following section a list of the basic steps were performed for the automatic image preprocessing in this study.

1. Captured images were uploaded into the system using the graphical user interface (GUI) of the system. This function was implemented to ease the process of selecting images in storage devices.

2. Contrast enhancement was performed to enhance uploaded images, remove dark areas, increase image brightness, and make images clearer. Histogram equalization was applied to enhance the contrast of the color image intensity before the image was converted into a gray scale image. The frequency occurrence of pixel intensities was given by the histogram and mapped to a uniform distribution, and then image intensity was adjusted to increase image contrast. Histogram equalization is one type of gray scale conversion; it used to convert the histogram of the original image into an equalized histogram. An accumulated histogram was calculated from original image and then divided into a number of equal regions. The corresponding gray scale in each region was assigned to a converted gray scale. The effect of histogram equalization was the enhancement of image parts that have more frequency variation, whereas parts of an image with less frequency were neglected. This step was performed to improve the appearance of the images in terms of the image contrast. Figure 3.12 shows a comparison between the original and converted images after histogram equalization was applied, and shows a comparison between the histogram of the original image with the histogram of a produced image.

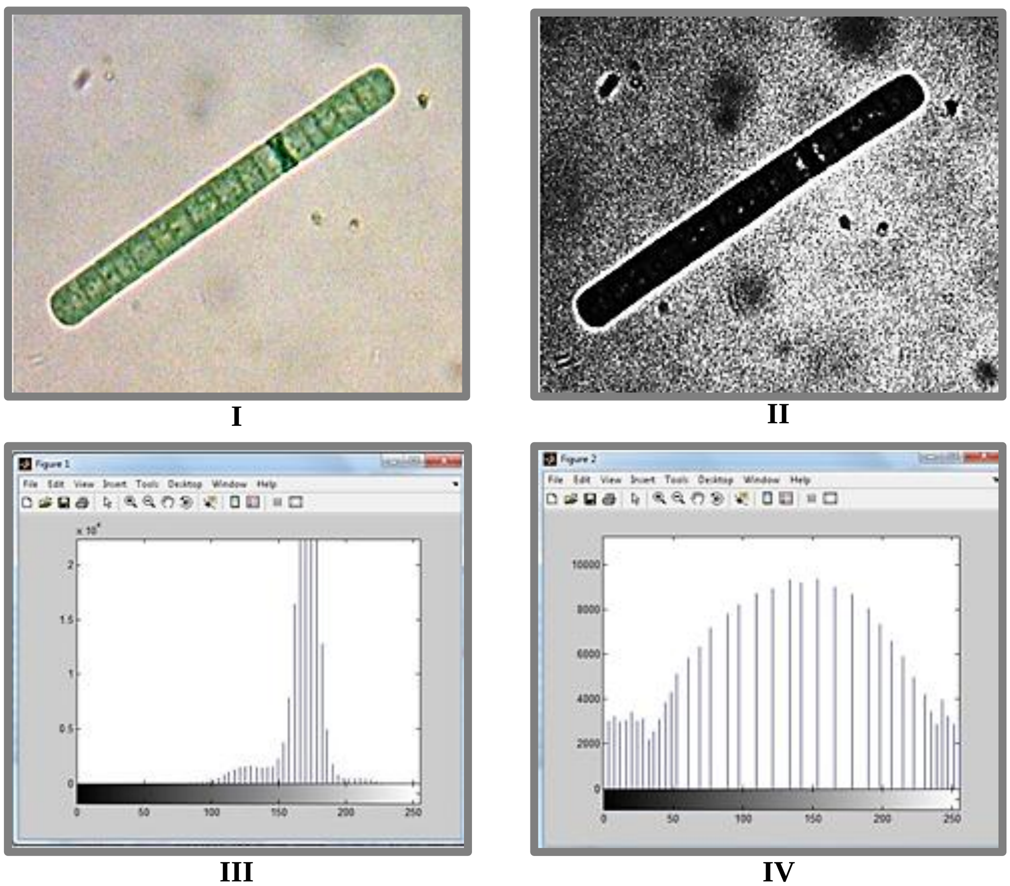
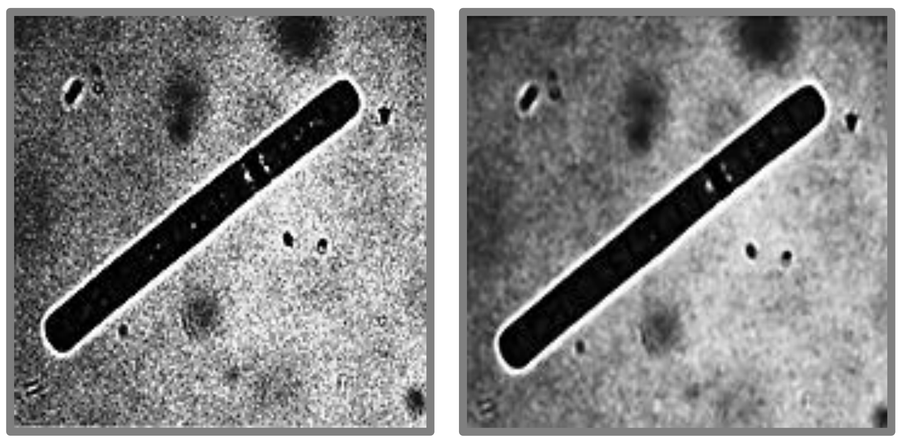


Figure 3.12 Examples for Applying Histogram Equalization Process as
I- Original Image of *Oscillatoria sp* ,**II-** Result Image After Equalization.,
III- Accumulated histogram of original image, **IV**- Accumulated histogram for

3. Median filter with (3 X 3) pixel size was used to reduce image noise, and to preserve edges. Some unwanted area and small objects is removed when median filter is applied. Median filtering is a nonlinear operation used mostly to reduce image noise in better way than other methods such as convolution.

Some unwanted areas and small objects were removed when the median filter was applied. Median filtering was a nonlinear operation used mostly to reduce image noise in better way than other methods such as convolution approach (Lim, 1990). Figure 3.13 shows some examples for image results after applied median filter.



I *Oscillatoria sp* image in gray scale Before the process

II- *Oscillatoria sp* after Median Filter was applied.

Figure 3.13 Examples for applying Median Filter

3.3.2 Image segmentation Module

The image segmentation process was used to isolate individual objects in captured images. Images of selected algae genus rarely exist alone and mostly contain several objects such as microorganisms and other algae. Image segmentation is used to identify the number of detected object in binary image and divided the original image into several sub-images based on the number of objects detected. Image segmentation uses preprocessed images to segment it into sub-images. In this study, we used a canny edge detector algorithm to detect the objects and perform image segmentation. A canny edge detector is considered as the most powerful edge detector for image segmentation (Canny, 1986). It is used to identify discontinuities in an image intensity value or the edge of the image. The canny edge detector was implemented using the steps described below.

a) The image was converted from gray scale to binary, and the image size was reduced to increase the system performance. The image samples used in this research were acquired with a resolution of 1024 × 1280 pixels. This higher resolution had a strong effect on the overall system performance including the processing time required to process each individual image. If the resolution of the images varies, incorrect results can arise during feature extraction or classification. A small routine was written to reduce the image size to 300 × 300 pixels.

b) A Gaussian filter was applied to sharp the binary image. This filter was used to reduce the noise of binary images by applying specified standard deviation sigma (σ) with default value for sigma was 1.

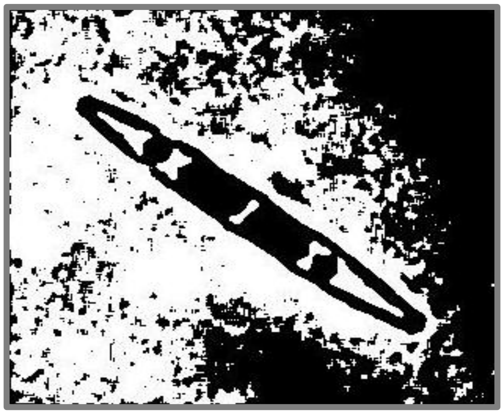
c) The local gradient and edge direction were computed for each point using equations 1 and 2. The Gx and Gy for each point were calculated using the first derivative of pixel intensity. Local maximum for gradient direction is used to identify the edge points.

$$g(x, y) = [G_x^2 + G_y^2]^{1/2} \qquad (1)$$

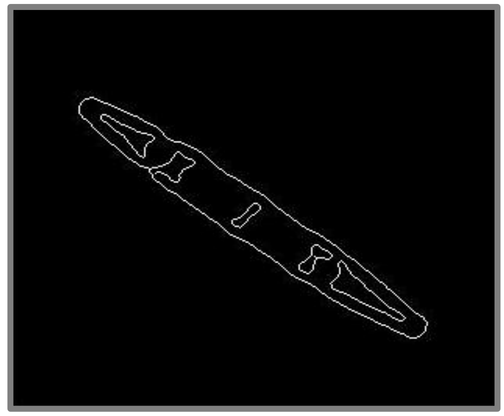
$$\alpha(x, y) = \tan^{-1}(G_x^2 + G_y^2) \qquad (2)$$

d) Non-maximal suppression in the gradient magnitude image was achieved to give a thin line where each ridge of the edge points was calculated in (2). Then, the threshold with 0.8 was applied on the ridge pixels to ignore the edges that were weaker than the threshold.

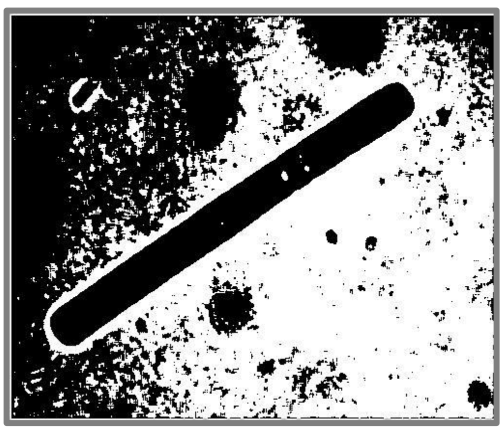
e) Finally, the algorithm performed edge linking by incorporating the weak pixels connected to the strong pixels. Figure 3.14 illustrates two examples of result images after applying canny edge detection.



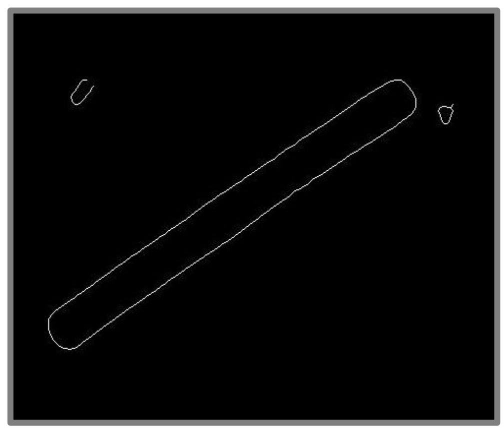
I- WB Image for ***Navicula sp***



II- Canny edge detection for ***Navicula sp***



III- WB Image for ***Oscillatoria***



IV- Canny edge detection for ***Oscillatoria***

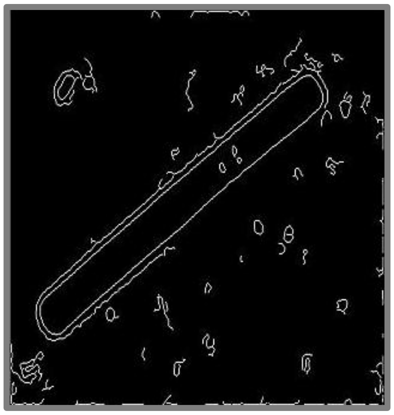
Figure 3.14 Examples of Results after Canny Edge was applied

After using the canny edge approach to detect the objects on the binary image, essential morphology steps were performed on the resulting image, including removing image borders and small objects, as well as filling the boundary area. Morphology operation is a set of image processing operations that process images based on shapes. Morphological operations apply a structuring element on input image to create an output image of the same size. In our system, we used dilation and erosion, which are considered the most basic morphological operations. In morphological operations, the value of each pixel in the output image was compared with the corresponding pixel in the input image of its neighbors. Morphological operation was performed by selecting the size and shape of the neighborhood pixel for each cell.

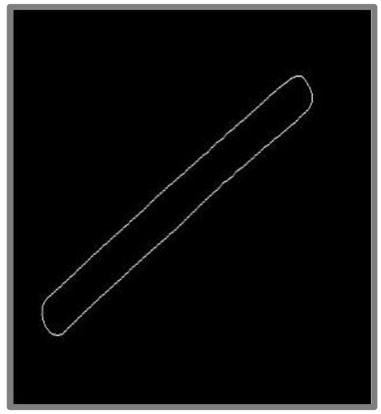
Below is the morphological steps performed in our system to improve the process of image segmentation.

- **a)** An open binary image operation was used to remove small objects on the binary image. It first determined the object-connected components, the area of each component was calculated and any small region less than 80 pixels were removed.
- **b)** Dilation operation was applied to enhance the object boundary and close the open region of objects. Dilation was used to expand the structuring of the element objects. We then performed a flood-fill operation on background pixels for resulting images is then performed to fill in the object boundary.
- **c)** Erosion operations were applied on the binary fill image by the erode function, which determines the center element of the object neighborhood and erodes the binary image.
- **d)** The binary image was segmented into sub-images based on the region number found in the binary images, and the exterior boundary objects were calculated.
- **e)** Finally, the segmented area was used to outline the original image to extract the algal objects for feature extraction purposes.

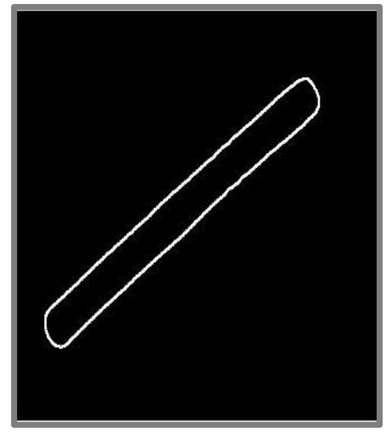
Figure 3.15 shows the steps in applying the morphological operation on the binary images, also appendix (II) showed the process of preprocessing with segmentation for selected algae.



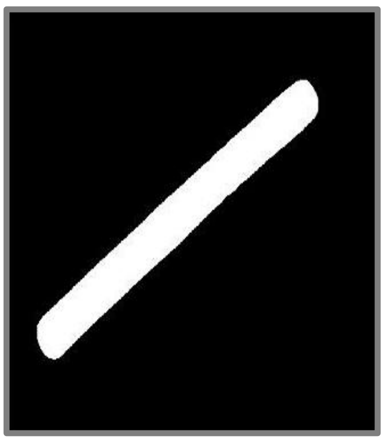
I-BW Image for *Osillatoria* after Canny Edge detection.



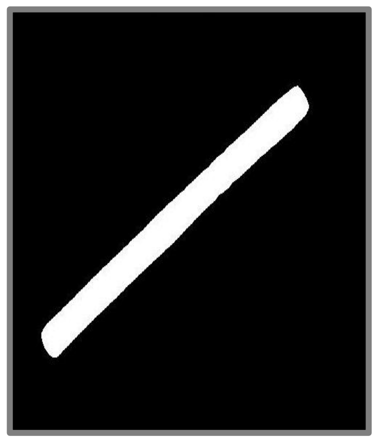
II-Open Binary process to remove small objects



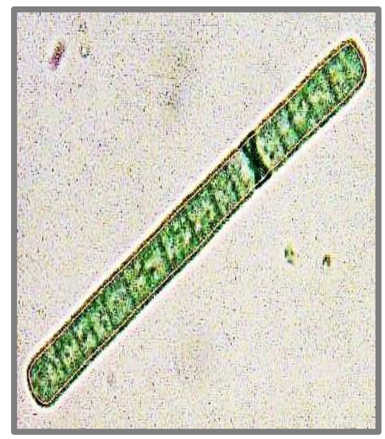
III-Dilation process to enhance Detected Edge



IV-A Flood Fill Operation to fill object area



V-Erosion Operation to focus more on detected object



VI-Outline same area of object on original image

Figure 3.15 Image Samples for Morphological Operation on *Oscillatoria sp*.

The images of a selected algal genera are rarely exist in isolated forms; thus, image segmentation approach is used in this study to divided algae input images into different images based on the number of detected objects, where each image must contains only one object. In the segmentation process, each object is isolated in separate sub images and counted as one item to overcome the problem of overlapping objects in an image. To separate individual objects, we implemented a simple routine that copies each regions enclosed by a rectangle with a maximum length of <80 pixels and width of <50 pixel into a new binary image. Overlapping filaments are represented as one region, which isolated as single object to solve algae overlapping problems.

Image segmentation is used to separate the objects inside the original images into sub-images that are processed individually. The region of binary image is detected using canny edge approach, and each region is represented on the sub-image. Each sub-image was used as a mask to obtain the same region of original image (color image); both regions of color and binary images are associated with the corresponding index number for storing purposes in the database. Our segmentation method was designed based on the theory of successful segmentation (Shapiro & Stockman, 2001). The boundaries of each segment was noted to be jagged or ragged, also whether the segment was smooth and spatially accurate. Small holes were removed from the interior region, whereas small objects were removed from the exterior region. Adjacent regions of a segment were distinguished significantly compared with their characteristic descriptions. Overlapping objects were segmented separately to process or neglect them during the classifying process.

3.3.3 Novel Technique of Image objects Alignments

Before shape extraction is performed, we developed a technique that performs automatic rotation for objects to be aligned with the horizontal axis. This step increased the accuracy of our system accuracy as we will later on in the discussion. A small routine is developed to obtain the angle of inclination for object automatically. This routine is then used for rotating image objects to be aligned horizontally as shows in Figure 3.16.

The angle of inclination is calculated automatically by obtaining the longest path between each two points on the object boundary. Identified point P1 (X1,Y1), P2(X2,Y2), and Origin point P(0,0) are used to determine the angle of inclination using the following equation:

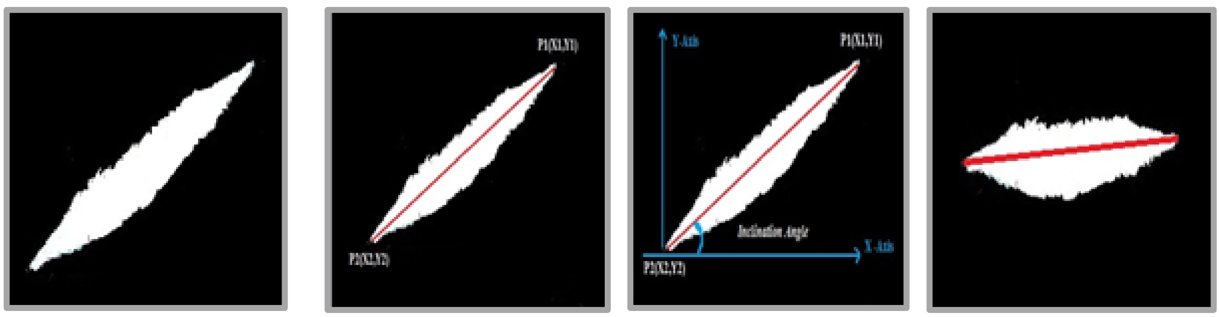
$$\theta = tan^{-1} (m1\text{-}m2)/(1+ m1*m2) \,, \qquad (3)$$

where m1 and m2 are the slopes of lines that form the angle that is obtained using the following equations:

$$m1 = (Y2\text{-}Y1) / (X2\text{-}X1) \qquad (4)$$

$$m2 = (Y1\text{-}Y0)/(X1\text{-}X0). \qquad (5)$$

This routine is designed to align the rotated shape into horizontal lines which will ease the feature extraction process, and improve the accuracy and performance of the recognition process by ensuring that all the extracted features are calculated with similar positioning of the object coordinates.



I- BW object in normal position

II- Find two points of longest distance

III-Obtain angle of inclination

IV-Alignments object Horizontally

Figure 3.16 Sample Steps of Auto-Orientation Novel Method.

3.3.4 Feature Extraction Module

Feature extraction is a process of determining various attributes and properties associated within a region or objects. It is an image processing approach that is used for object characteristic measurements using an algorithm to generate a set of descriptors and characteristic attributes from a binary, gray, or color images to present meaningful features. It is used to transform an image into a set of recognized parameters that can be used later for the classifying process (Chandra & Majumder, 2000). Image feature extraction process is involved in three different aspects including spectral features such as color, tone, ratio, and spectral; geometric features such as edges and lineaments; and textural features, such as pattern, homogeneity, and spatial frequency.

In this study, a specific feature set was selected to obtain the essential characteristics for the selected algae. Feature extraction process is implemented to extract some parameters from both binary and color images of algae including shape index, area, perimeter, minor and major axes, centroid, equivalent perimeters, bounded box, and Fourier spectrum with principal combination analysis (PCA). Other features were derived from the main feature to obtain the relation between shape parameters such as that between area and parameters. In the following section, we describe in detail the steps involve in the extracted features performed in this study.

i) Area

The area represents the actual number of white pixels in the selected region. It is calculated by using the number of white or “1” pixels inside the image object as shows in Figure 3.17. Area is used as a parameter in classifying operations because it gives an indication of the size of objects.

Figure 3.17 Example for extract area

ii) Perimeter

The perimeter of the object is the summation of the distance between each adjoining pair of pixels around the object border. It shows in red pixel in Figure 3.18.

Figure 3.18 Example for obtain Perimeter, red pixels is the Perimeter.

iii) Major and Minor Axes

Major and minor axes are extracted where two points are identified automatically by calculating the maximum distance between given points in the objects vector. The major axis represents the line segment connecting between the base points in the X axis, whereas the minor axis represents the maximum width which is perpendicular to the major axis. The major axis can be used as object length and minor axis as object width as represented on Figure 3.19.

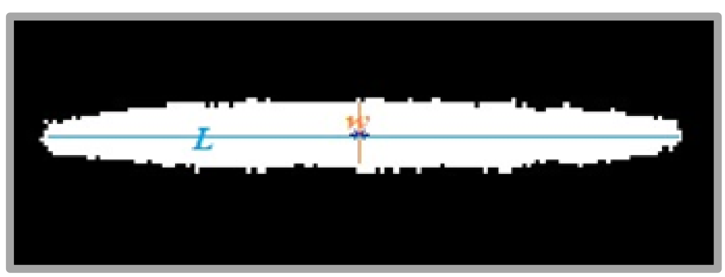
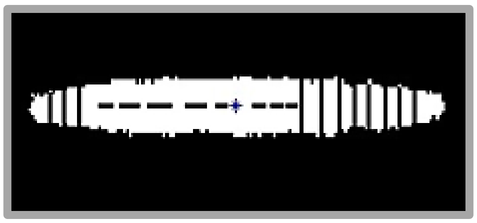


Figure 3.19. Sample for Extract Major and Minor Axes.

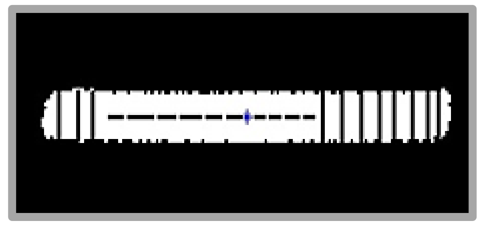
iv) Object width factor

A new feature was selected in our study, called the object width factor. It is calculated by slicing across the major axis and parallel to the minor axis, the feature points are normalized into a number of vertical strips, and for each strip, the ratio of strip length to the object width is calculated as the following equation:

$R_c = W_c/L$, where R_c is the ratio at column c, W_c is the width of object at column c, and L is the object length as shows in Figure 3.20. This feature is used to differentiate between some algae that identical in width and length, and have some essential variations in width at different position such as *Navicula* which has similar value of width with *Oscillatoria* at middle position; however it has varies values of width in different position as illustrates in Figure 3.20.



I- Sample for Slicing *Navicula*



II-Sample of Slicing *Oscillatoria*

Figure 3.20 Sample images for Slicing process on *Navicula* and *Oscillatoria*

v) Equivalent Diameter

Equivalent diameter is the feature that determines the diameter of a circle with the same area of a segmented image. It is computed as:

$$EquDi = Sqrt(4*Area/pi).$$

vi) Euler Number

We used this feature to identify the number of holes inside the image object. We applied this function into the segmented image before it was filled by holes. The Euler number is equal to the number of objects in the region minus the number of holes in those objects. Figure 3.21shows one example of using Euler number to improve our classification method.

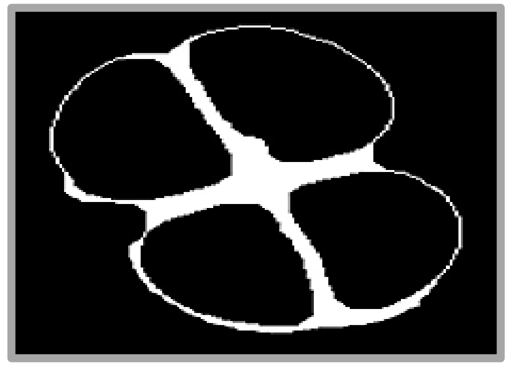


Figure 3.21 Sample of Extract Euler number for *Chroococcus sp*.

vii) Bounding Box

Bounding box is identified as the smallest rectangle that contains the region. The box has four feature parameters, including the left corner point of the rectangle (X1, Y1), length, and width of bounding rectangle. Figure 3.22 shows one example for object bounding box.

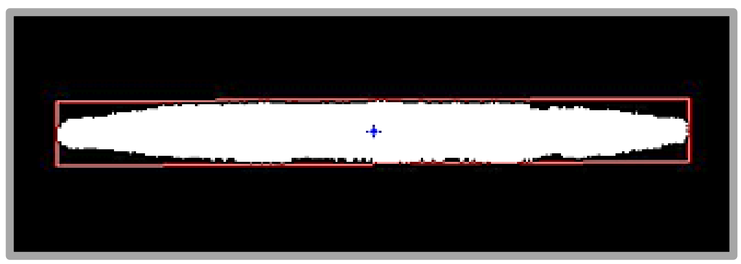


Figure 3.22 Sample for extract bounding box parameters for *Navicula*.

We also used the rectangle bounding the object to compute the extent parameter which represents the proportion of pixels in the bounding box within the region. The results are obtained by dividing the object area with the area of bounding box.

viii) Centroid

The centroid is the center of mass for the segmented region, and helps us determine the object center coordinates. The elements of centroid are two points, the x- and y-coordinates, of the center of mass. Figure 3.23 represents the centroid of object as small blue star in center of object.

The centroid as parameter does not provide a distinguishing feature for selected algae but we can use it to drive a new feature to enhance the classifications method. We also used the centroid coordinate to extract the length and slope of line that connects between the centroid coordinates points and the left corner point extracted from the bounding box, the line is shown also in Figure 3.23 by blue color.

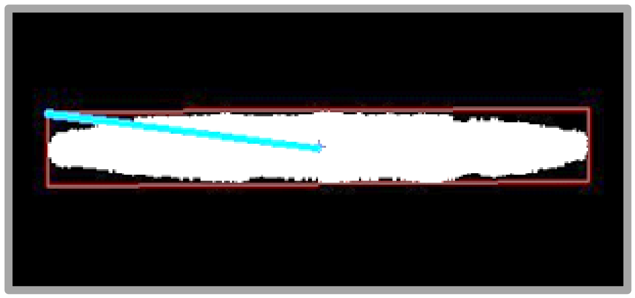


Figure 3.23. Illustrates Centroid and Slope Lines.

ix) Derived features

We used some other relations to compute the selected feature ratio for improving the classifying process such as the ratio of the area and perimeter (perimeter/area), ratio of perimeter and object length (perimeter/ length), ratio of perimeter and object width (perimeter/width), and ratio of minor to major axes (Minor/Major).

x) Shape Index

Extracted shape index is one of the novel techniques developed in this study. We used our biology background of algal taxonomy based on algal shapes to develop a new function that can categorize algal shape into three different types which are circular, spiral, and irregular shapes. This function is designed to identify algal shape type and provided us with a new feature named shape index, which is equal to "1" if the shape is circular, "0.5" if the shape is cylindrical, "0" if the shape is spiral, and "-1" if the shape is irregular. This classifier is used to improve accuracy rate and optimize the time of recognition process (Hayat et al., 2011). The results obtained from shape extraction will be included as one of the input parameters for algal classification using MLP. The shape index feature can be used as early stage classifier for the algal taxonomy area. The measurement of the shape index is obtained by evaluating the roundness or cylinder of the objects as described in the following steps:

- Comparing the longest and shortest diameters for the object.
- Comparing the area with the formula $\pi \times r^2$.
- Comparing the perimeter with the formula $2\pi \times r$.
- Eccentricity is also used to improve the results of shape index measurements. Eccentricity refers to the ratio of the distance between the foci of the ellipse bounded the object and its major axis length.

xi) Entropy of gray image

Entropy is defined as the statistical measurement of randomness that can be used to characterize the texture of the input image. Entropy is used to extract texture features on the gray scale image. During segmentation processes, the detected object is used to obtain the outline of both color and gray images. Entropy is implemented in our system to obtain one output feature from the outlined gray image. Entropy is given in the following equation:

xii) Fourier Spectrum with PCA

Fourier spectrum was applied to extract some of the texture features for algal images. Using texture feature increases the accuracy of system detection. Fourier spectrum is ideally suitable for describing the directionality of periodic or almost periodic two-dimensional patterns.

Power spectrum is obtained for each image by applying second Fourier transforms techniques. The power spectrum is extracted as vector for each individual image.

The Fourier transforms is used to represent input images as points in space spanned by asset of spectra. Vector of linear models is extracted to represent overall image features. Then, the PCA approach is used to normalize the basis of power spectra in small vectors with 15 elements dimension.

PCA is used almost to normalize the extracted values by reducing the redundancy values exists between the pointes of features. For easier implementation and best performance we restricted to use only the first basis vectors of power spectra which given by 15 largest eigenvalues of PCA. Thus, also used because the orientation of image objects have been solved previously.

Features extracted from the same algal object may contain some inevitable and redundant features. The PCA approach is considered to be useful statistical technique which used widely in many image processing applications. It has the ability of serves to de-correlate redundant features, and also the function of energy packing property which compressed useful information into a few dominant features (Tang et al., 1998).

In this study, the PCA algorithm is used to reduce and summarize the extracted features using the Fourier Spectrum method for removing redundancies. Additionally, we applied the principal component analysis to the power spectra of 300×300 pixel size of original image outlined. The result of power spectra is used as input data for PCA only. For more performances, we calculated the strongest eigenvectors and eigenvalues of the covariance matrix without explicitly computing the covariance matrix itself. The resulting images were normalized for visualization using linear transformation of the minimal and maximal values of the 15 components to the minimal and maximal values of the grayscale range (0–255). The covariance matrix was extracted and mean values calculated for each input images to yield 15 features used in our classification methods as shows in Figure 3.24.

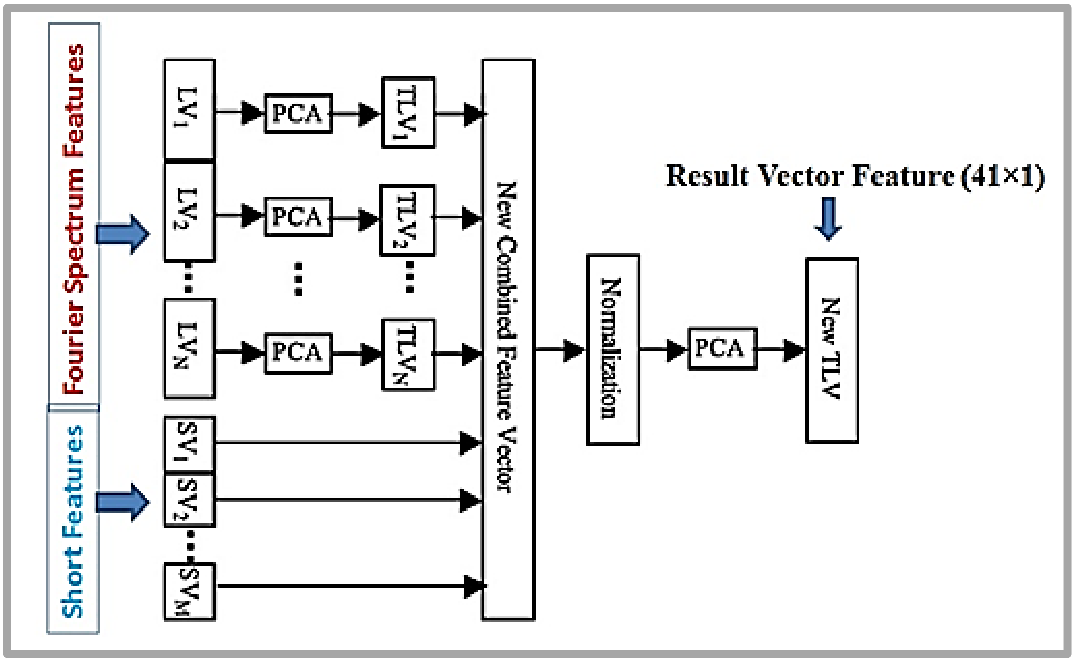


Figure 3.24. Using PCA for Features Normalization Process.

At this stage, we completed the feature extraction process, which is a combination of shape and texture features used to extract different parameters as shown in Table 3.1.

Table 3.1. Extracted Feature as Vector used in this study.

Feature No.	Feature Descriptions
F1	Area
F2	Perimeter
F3,F4	Major Axis, Minor Axis
F5-F8	Object Width Factor Strips
F9	Equivalent Diameter
F10	Euler Number
F11 – F 14, F15	Bounding Box Parameters (smallest rectangle bounded the object), and Extent area
F16, F17	Centroid coordinates
F18, F19	Length and Slope for the line that connected between centroid and bounding box corner.
F20, F21, F22, F23	perimeter/area, perimeter/ length, perimeter/width, Minor/Major
F24, F25	Shape index (circular = 1, cylindrical = 0.5, spiral = 0, and irregular = -1), and Eccentricity
F26, F27-F41	Entropy, Fourier Spectrums Normalized by PCA

3.3.5 Classification and Identification module

Our system objective is to develop an application for recognition and classification of freshwater algae. All the previous steps were performed to obtain freshwater algae that can be used to identify each type of selected algae. Previous studies have reported that selected features for any objects affect the classification process. Artificial neural networks are considered the heart of the process for any recognition or classifying approaches. There are many types of neural network models that are commonly used for identification, recognition, and classification of image objects. These types of ANN are widely used for pattern recognition and classification. In this study, one hidden layer feed forward neural network was chosen primarily because it has been proven that such a topology can approximate any continuous function as listed previously in literature review.

i) Neural Network Architecture design

In our research, a multilayer feed-forward with back propagation error (MPL) is used for training and recognition process of selected algae. The extracted feature is combined to construct a vector feature which is represented as matrix of (41×1). The MPL ANN designed with seven source nodes that constitute the input layer, one hidden layer contains of eight nodes, and one output layer of computation nodes with five output nodes as shows in Figure 3.25. The input vector propagates through the network in a forward direction on layer by layer basis. Learning for the MPL ANN consists of two passes through different layers of the network. Forward pass and backward pass is used for learning process of network. The input pattern was applied to the input nodes on forward pass of the network. The forward pass is followed by a backward pass that propagates the difference between actual outputs of networks and target outputs. The synaptic weights are adjusted to cause the actual response of the network to be closer to the desired response. The standard root mean squared error function (RMSE) was used to access network performance, and a momentum value of 0.05 was set based on trial and error. With the above parameters fixed, optimal step sizes taken in weight space were made a function of the learning rate of 0.85 width and epoch size of 400.

The final architecture of MPL ANN consists of one vector with forty-two individual elements, one input layer with seven input nodes, One hidden layer with eight hidden nodes, and one output layer with five output nodes as shows in Figure 3.25.

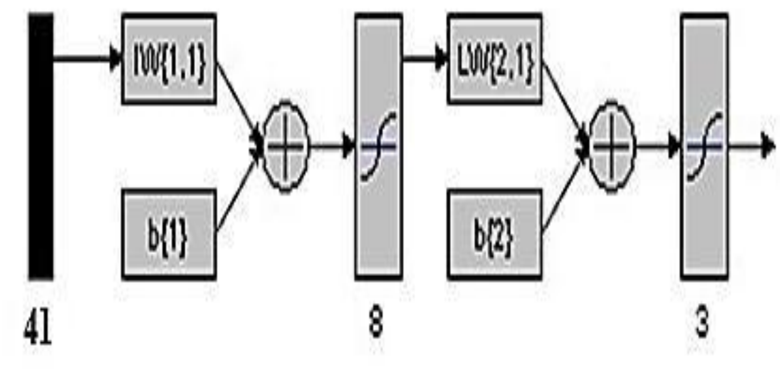


Figure 3.25 MPL ANN Architecture design

In our multilayer network, *a sigmoid* transfer function is used in the hidden layers and function *logsig* is used in the output layer to generate output values between the range of 1 and 0. Multiple layers of neurons with nonlinear transfer functions allow the network to learn nonlinear relationships between input and output vectors. The initialization parameter for our MPL NN was implemented on our system function to run simply in graphical user interface.

ii) Image preparation for training

The input to MPL NN is a vector feature with 41 × 1 dimension features extracted from the input image. The element of vector features has been shown previously in Table 3.1. The training set of images was used to extract the feature vector for each image and each selected alga has 50 images for training mode. All training images were used to extract the feature of selected algae and the results stored in the database under five different classes. Each class represents one of algal division type. Some of extracted features, including area, length, and width, have large values, whereas some features, such as shape index and width/length have small values. The normalization process is applied on features that have large values for both input and target vectors in the data set. It is also used to increase the classification accuracy, optimize the time of network training, and make the network output fall into the normalized range consistently. During the training phase, input data and desired responses were presented repeatedly to the network. The network uses momentum learning algorithm to determine the weights in the network. Weights values are adjusted to minimize exists error for each step between the desired and actual output. We tried to minimize the exist error between the desired response and networks towards zeros by increasing the number of epoch as progressed training. As mentioned previously that MLP with single hidden layers can be used to virtually approximate any input values with output map. We found that is possible over-training a network by developing a network to classify the training data perfectly however we found it is unable to generalize and classify new 'unseen' data.

The training process is implemented to upload the database file, which includes the normalized feature extracted from all the training image samples, and the training process set to stop when either the learning rate is achieved or when the epoch is becomes larger than 400 steps. During the training phase, the MPL accesses to the input and desired data repeatedly to adjust the neuron weights to minimize the error between the desired and actual outputs. The extracted database was split into training data and cross validation to improve the generalization of the neural network. Figure 3.26 shows the training function interface on the developed system. The next chapter discusses in more details the system testing and evaluation.

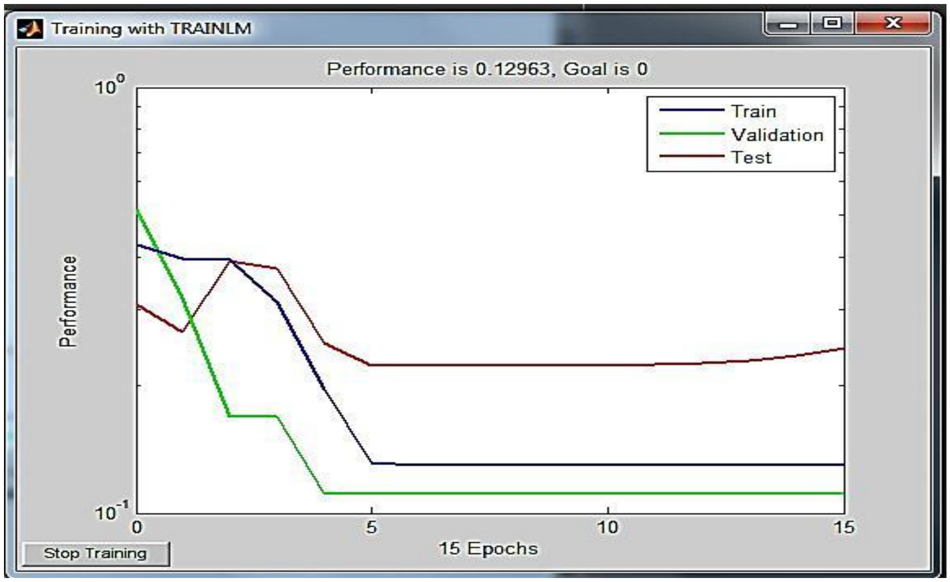


Figure 3.26 Step of training process in MATLAB

3.3.6 System Evaluation Approaches

After developed the system we performed three different type of evaluation to ensure that our system meet user requirements and other critical criteria such as, reliability, accuracy, and performance. These three methods are testing system functionality, comparison between manual and automated classification, and using confusion matrix.

During system development process we performed functional tests over developed functions to verify each function is run correctly, and achieved their target tasks. Then, two types of evaluation were performed to measure overall recognition accuracy rate for developed system. First method, we initiate experiment to classify the test images manually by experts, and then use developed system to identify the same images automatically. Then, experimental results of comparison between both manual and automated classification were used to obtain recognition rate. The second evaluation method depends on confusion matrix which considered most common method used to assess image accuracy classification (Kohavi & Provost, 1998). Confusion matrix as evaluation methods are used in this research to compare between the individual algae for obtaining system accuracy. Both previous methods of system evaluation are commonly used by most researchers to determine the recognition rate of their developed systems such as (Tang et al., 2006; Sosik & Ropert, 2007; Verikas et al., 2010; Dimitrovski et al., 2012)

In other hands, we also evaluated system performance based on the speed of classification process between both manual and automatic approaches. In experimental results we calculate the time spent by experts to classify the selected algae, and then compared it with the system time exhaust to perform classification process. The results of system evaluation in terms of performance and accuracy are described in more details on next chapter.

3.4 Thesis Contribution

Image preprocessing and image segmentation was designed in appropriate manner to detect microscopic algal images. Image segmentation methods showed accurate results in terms of isolating each object inside input images. It also solved some common critical problems exists mostly in algal images such as small area, and overlapping objects.

In our study, we used new techniques to ensure the most accurate results for recognition and classifying processes. For example, we extracted different types of algal features, such as shape, texture, and domain frequency features. A novel routine was also implemented to align the image object with horizontal coordinates, and extracted function to obtain the algal shape index that provides proof of its suitability to support the classification process and improve system recognition accuracy.

CHAPTER 4

RESULTS AND DISCUSSIONS

Developed system is an application for complete automated analysis process of selected freshwater algae, which includes all steps from the image acquisition to identify, recognize, and count the algae objects. In According to previous chapters, we selected five types of freshwater algae, where 100 image samples collected and stored in jpg format for each individual alga. Then, image samples are classified into two data set for training and testing samples, where 50 image samples used for training purpose and the other 50 image samples used for testing system performance, and evaluating system accuracy. In this chapter we will describes with more details about the results of system testing and evaluation phase process, and then we will discusses our finding based on these results..

4.1 RESULTS:

The success of this research is evaluated according to whether useful data were obtained in terms of temporal coverage and whether results obtained in terms of speed and accuracy of identification.

The testing data set included three genera of Cyanobacteria, one genera of Chlorophyta, and one genus of Bacillariophyta. 50 image samples used for testing purpose for each selected algae genus. For system testing and evaluation purposes we used two different methods of test as mentioned in previous chapter. In first method we used the comparison between the manual and automated recognition results, and in other test method we used confusion matrix method to obtain system recognition rate. Figure 4.1 is illustrates snapshot screen example of result for system recognition process. In following section we will describes both methods that used for testing and evaluation proposed system.

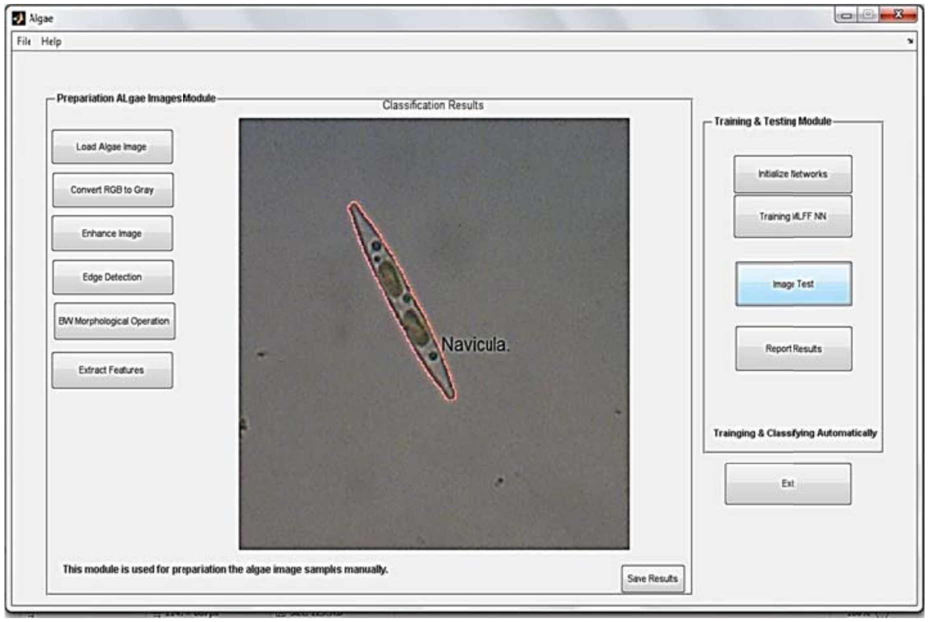


Figure 4.1 Snapshot screen example for System Classification Results.

4.1.1 Results comparison of manual and automated process:

A complete automated algae recognition system has been developed to provide a test environment for our thesis statements. The developed system contains all essential function requirements to perform the automatic recognition and classifying for selected freshwater algae.

An experiment performed to compare between the identification results of manual and automated approaches for measuring proposed system accuracy. There are 5 freshwater algae involved in our study, and each type has 50 images as testing dataset, a total of 250 images are involved on this testing methods. Two experts took a part in the experiments to identify manually the algae object found on the testing samples images, and the same dataset of testing images is uploaded to our system. Results are recorded in datasheet file for both manual and automatic identification methods over all the tested image samples. Table 4.1 represents the comparison results between manual identification that perform by experts with the automated recognition process achieved by proposed system.

Table 4.1 Comparison Results of Manual and Automated classification process.

Detected Object	Manual	Automatic	Accuracy for Auto./Man.%
Unidentified	**324**	**285**	**87.9 %**
Chroococcus	**123**	**108**	**87.8%**
Microcystis	**109**	**98**	**89.9%**
Navicula	**147**	**141**	**95.9%**
Oscillatoria	**132**	**124**	**93.9%**
Scenedesmus	**113**	**109**	**96.5%**

First column in Table 4.1 above represents the names of selected freshwater algae used in this study with additional row at named unidentified used for object that misidentify during manual and automated recognition process. Second and third column contains the results for the total objects found in image dataset in manual and automated process, only correct object identification is counted in this methods. The last column represents the comparison results obtained by comparing the automated and manual identification process. Figure 4.2 shows the graph chart of comparison results between manual and automatic identification process which appear approximately close to fit with small variation, where the higher differences values exist more in detecting unidentified objects which can be ignored.

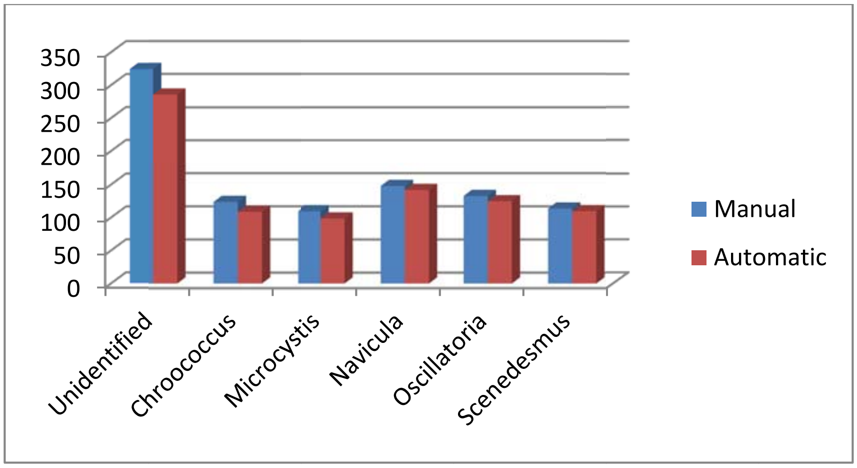


Figure 4.2. Chart of Comparison Results for Manual and Automatic Process

Number of extracted region for each alga was examined to ensure if the object was an alga or other object. The process for separating the objects which found in the image was resulted some short irregularly shaped image regions, which mostly misclassify these small segments, and the other small object was excluded during the segmentation process. Any other objects found on the algae image were classified by the MPL ANN as unidentified. Comparison of results of algae identification by automated system shows an accuracy rate of 90% of the manual identification. The results of comparison showed that developed system is able to identify the algae objects in given images within the approximate accuracy of manual procedure.

4.1.2 Results of Confusion matrix for image testing dataset:

Confusion matrix is used for testing our system to validate the data results between the selected freshwater algae. After system training process, we performed system testing for each type of selected algae images. Each selected algae involved in this study has 50 image samples reserved for testing purposes. System recognition results by using confusion matrix illustrates in Table 4.2. First column contains name of selected freshwater algae used in this study, while column 2 includes the total numbers of testing image samples for each selected algae. Columns 3 to 6 are the system results for recognition process of each individual algae. Finally, column 7 shows the system accuracy results for recognition rate which calculated by using the validation methods for individual classification. According to the results of confusion matrix as shown in Table 4.2, some errors of misclassification occurred between some of selected algae which they have some similarity in specific shape or features such as *Navicula* and *Oscillatoria*.

Table 4.2. Confusion matrix for testing dataset

Name	No. of Test samples	System recognition results						System Accuracy
		Chr.	Mic.	Nav.	Osc.	Sce.	Unknown	
Chroococcus	50	46	2	0	0	0	2	92%
Microcystis	50	2	43	0	0	0	5	86%
Navicula	50	0	0	49	1	0	0	89%
Oscillatoria	50	0	0	2	47	0	1	94%
Scenedesmus	50	0	0	0	0	49	1	98%

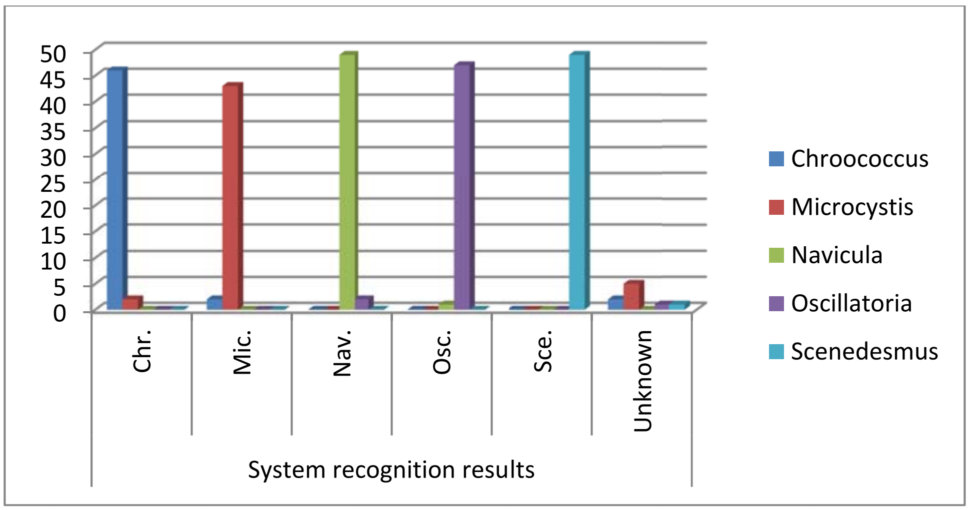


Figure 4.3. Confusion matrix results graphic charts.

4.1.3 System Accuracy:

The developed system is evaluated to measure the accuracy of classifying process on image data set. Evaluation system accuracy performed by using the previous results of comparison, in first method by measuring the accuracy results between the manual and automated identification process as shown in Table 4.1, and illustrated as chart in Figure 4.4.

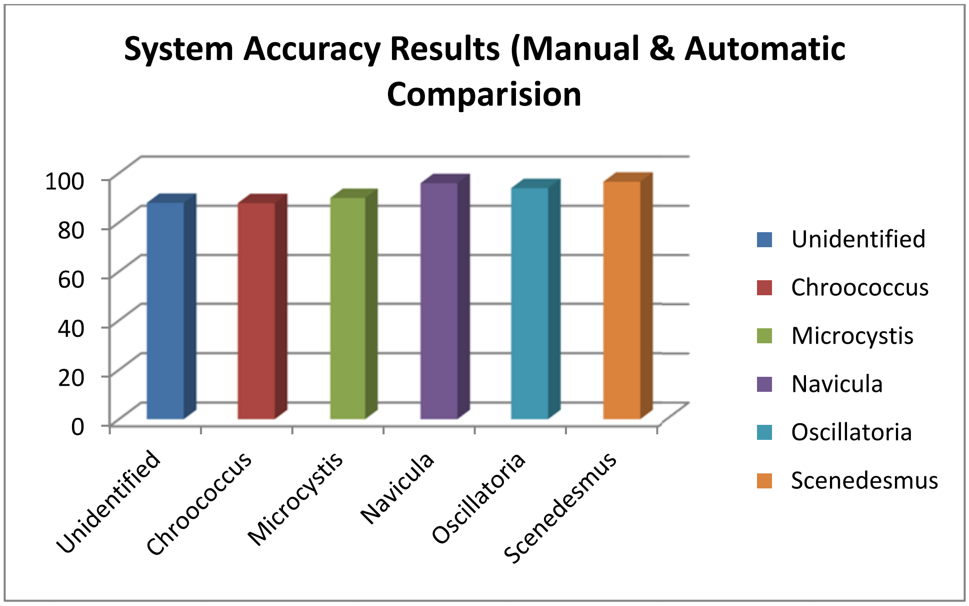


Figure 4.4. Chart of System Accuracy for Manual and Automatic Comparison

Results of comparison between manual and automated process was used to evaluate our system accuracy. Both results showed a great accuracy between the manual and automated approaches. Recognition rate for object detection and classification has some variation values between 87% to 96%, and the average system accuracy calculated based on this method is approximately 91.9 % which consider a great accuracy achievements. The system recognition accuracy for both used methods is calculated as follows:

System recognition accuracy = (automatic recognition/ Manual identification) * 100/100

The results of comparison showed that developed system is able to identify the selected algae images within approximate accuracy of manual procedure.

Confusion matrix method is used to obtain recognition accuracy rate as shown in Figure 4.5 and described in Table 4.2 previously. The testing results in this method are showed a recognition accuracy rate varies between 86 and 98.

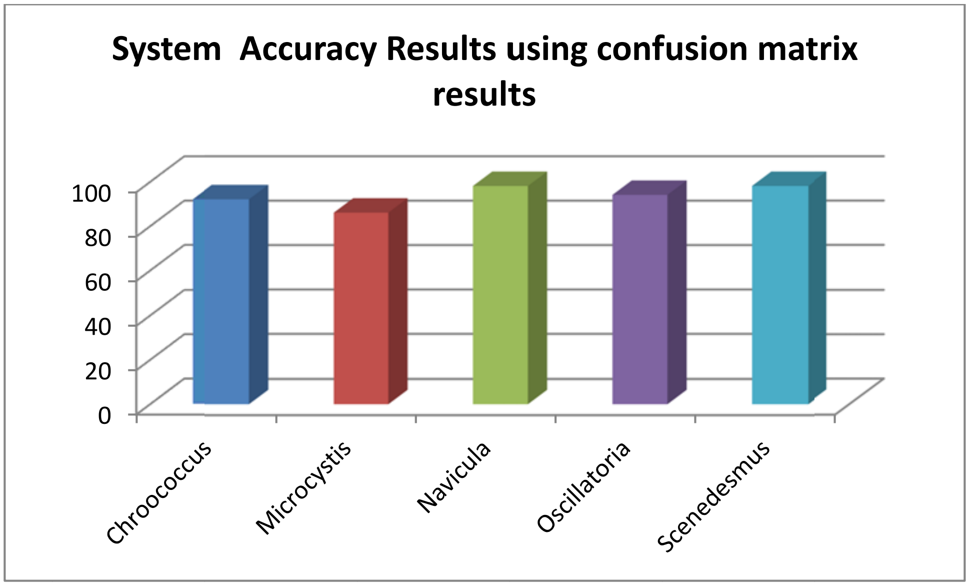


Figure 4.5. Chart for System Accuracy Results of Confusion Matrix

Actual classification accuracy examined by confusion matrix is shown as chart graph in Figure 4.5. In this method classification results given by the automated analysis are compared with the desired classification (as defined by a human expert). The results demonstrated that proposed system identified 93% of input images successfully from total of 100 input images, where results indicated that the average recognition accuracy rate for proposed system is approximately 93.6% when included all extracted features into the MPL ANN classifier. Overall, the total accuracy for system recognition rate is calculated as the average accuracy between both evaluated methods which had a great system recognition accuracy of 92.75%.

4.1.4 System performance

System performance is mainly concerned about many factors such as friendly interface, system reliability, and system processing time. For developed system we tried to make the interface simple as possible which build by using GUI interface, that facilitate the interaction between the user and system interface. Some system users had reported that developed system is easy to use, understandable, and has friendly interface.

System performance is mostly concern more about the time that system needs to solve the problems, or the time required to perform tasks and obtaining results. Actually, for this type of system evaluation, there are mostly two critical time needs to be measured, training time, and classification time. There are training time, and classifying time which calculated to evaluate system performance individually.

To calculate the training time a small routine was developed to obtain the system time before entering the training process, and then routine is access system time again at the end of training process. Then the training time is calculated based on the differences between the two time obtained at beginning, and ending process of training. The training process was run several times to capture training time to obtain reliable values for training time during training process. Then we found that the average time of the training process is varying between 3 to 5 minutes.

In other hand, the same routine is used to compute the classification time that system required to identify each upload images. Several images also is uploaded to obtain reliable results of recognition time, we found that the average time that system need to classify the objects inside the uploaded algae images is varying between 0.5 minute to 1.8 minute. This variation in time occurs due the varies number of objects found on each individual images, where images that includes more algae objects takes more time than the images that contains a few objects only.

Overall, the automatic procedure for training process takes approximately 5 minutes; and the time required for identifying and classifying process for the input images takes about 1.5 minute in worst cases. This results shows that developed system is achieved training and classification process in responsible time if compare with some existing system that developed for the same purposes as mention in review.

4.2 DISCUSSIONS

Our goal was to develop an image processing system with artificial neural networks to identify and classify specific freshwater algae to substantially increase ecological insight that can be achieved from rapid automated imaging systems. This research can be considered as a preliminary study towards developing an application that can be used to detect and classify selected of freshwater algae.

System design and experimental results showed that our approach meets the challenges of many category classifications. Furthermore, the extent of training and testing data ensures that our approach is robust and reliable enough to classify algae objects a cross several of conditions such as change in taxonomic composition, variation in image quality, variation in algae position and orientations which describes in more details as follows.

The training and testing data sets were collected from Putrajaya lakes with different sampling dates over more than six months, and then practical work performed in laboratory to acquisition the algae images which will be used in our study. The acquisition process was a tedious process which spans 5 weeks of continuous sampling and algae image acquisition.

In according to previous section, the system recognition results are showed a great accuracy and performance for the developed system. The accuracy and performance of our automated classifier exceeds than the expected consistency between manual microscopists with automated application which reported in review chapter by (Culverhouse et al. 2003, Grosjean et al., 2004; Blaschko et al., 2005; Hu and Davis, 2005; Luo et al., 2005).

Our developed system provides unbiased quantities abundance estimates with taxonomic resolution similar to manual analysis of algae samples. The highest accuracy rate of system recognition was achieved for identification of *Scenedesmus* as this alga has the most distinct feature compared to the other algae genus used in this study. Meanwhile *Chroococcus* has the lowest classification rate because of the process for separation resulted in the production of some short, irregularly shaped image region representing the algae. *Microcystis* which is circular in shape is difficult to distinguish because these algae exist in colonies and the images captured are prone to overlapping which cause the MLP to misclassify the algae to unidentified. The accuracy rate for *Navicula* and *Oscillatoria* can be misclassified with each other by few samples because they have spiral shape seems similar for the classifier and extracted feature for both of them are matching in most extracted parameters. The overall system accuracy of developed system depends essentially on the ability of system to detect object within input image and the ability of the classification system to identify the detected object based on the extracted feature. Accuracy rate achieved in this study is acceptable and consider higher rate if compared with other similar studies. The higher accuracy for system classification obtains due the specific approaches used in our researcher which describe in more details below:

4.2.1 Image pre-processors manipulation:

In this step, we employed some image processing techniques to get the appropriate results for the microscopes images. For examples, most of microscope image covered with shadow and dark area so we increase the image contrast by using the histogram operator. In this process, we tried to obtain sufficient contrast for image analysis to separate cells from background accurately. Microscopes image mostly contained noise and small unwanted area so we applied the median filter to remove the news and small area exists.

4.2.2 Automatic Segmentation of algae objects

Before segmentation process a well-prepared samples in an adequate amount of water is used, where the probe is sampled without disturbance of the sediment to avoid interference by detritus. Segmentation procedure that used in developed system is designed with specific condition to detect the algae objects found in input images, where the scums and other objects are removed automatically. Segmentation process is also designed with additional morphological operation to solve the overlapping problems and image border which exists in most of algae images.

A large amount of the misclassified results was occurred when cells touching each other, or overlapping with other objects, and also the small objects and or scums existed in microscopic images. In our segmentation method we used the canny edge detector with some image processing methods such as erosion and dilation of image regions to separate the touching objects, overlapping between objects, and to remove the unwanted area and small objects as showed in previous chapters.

Segmentation method is playing important rules in detecting objects, and can give more accurate results if the object shape identify well. In addition, the algae images which consisting of several optically weakly connected components are detected as separated objects, such as some of *Navicula, Osillatoria*, and *Scenedesmus* algae. This problem also solved by closing the region of objects to detect the object correctly. No segmentation performed for detected objects with area smaller rounded by a square of 50 × 50 pixels, and also no dissection for big overlapping or weakly connected components. That's lead to decrease the erroneous of detected objects in our segmentation approach.

4.2.3 Position and orientation of detected algae:

Algae often have more than one representation, position of algae in input images, and also orientation are varying diversity for most of algae images. It is difficult to find an alga objects inside two image samples within the same position coordinates or even within the same orientation. In these cases it appears clearly that the classifier should be designed specifically for certain algae structures in position, representation, and orientations. However, in our developed system we designed specific routines to solve this problem that move the detected objects into fixed coordinates, and also it performs the operation of alignments the detected objects with horizontal coordinates. By using this solution there is no need for multiple representations, and also there is no need to limit our classifier with some position of detected algae.

4.2.4 Differentiation of algae by using measurable features

In our developed system we used morphological features which measure the detected object properties such as area, perimeter, length, width, etc. however, to increase the classifier accuracy we included some texture features such as entropy and Fourier spectrum. In addition, we developed some new features to be extracted which can support our classifier to distinguish accurately between the selected algae such as slice width factor which essentially used to differentiate between *navicula* and *oscillatori*a, and Euler number which give different value for each selected algae. Anther novel feature is used as simple classifier for detected objects based on the shape index which describes the algae shape in numeric value rather than in words such as circuitry, spiral, ellipse, and irregular. This simple classifier is used as one of extracted features and can be consider as one step classifier that improves the accuracy rate of our classifier significantly.

In contrast, texture features as well as shape features are necessary to discriminate between selected algae, for examples in initial of system testing where morphologic features used only, the classifier accuracy obtained was below 60%, however when the texture feature used the classifier accuracy obtained was increased to more than 90%. Application of texture information also substantially improved separation of algae from none algae objects. Furthermore, color parameters are not considered in our feature due the color similarity for most of freshwater algae.

Generally, the identification quality as presented in Table 4.1 was very good when compared with the range described with manual counting. Also the accuracy results obtained in confusion

4.2.5 ANN Classifier:

The neural network approach employed in our research study had been used in a wide range of scientific classification problem successfully, and it showed greater results of identification and recognition of some algae studies (Boddy et al., 1994; Chtioui et al., 1996). This MLP neural network is required much less computation to arrive at their output desired. It is also simple in design and implementation than other neural networks. This leads to optimize the required time for training and testing, and also assist in simpler implementation.

The combined of MPL neural network and the rule-based classification used in this study is showed probably results near optimal based on recognition accuracy achievements showed in Table 1 and Table 2 in previous sections.

MPL was used in this study instead of SVM or RBF because the data utilized in this study are limited to few numbers of algae. The limited number feature has been utilized in this study because of the selected features are sufficient to detect and classify selected algae with considerably high accuracy rate. The improvements to increase the accuracy results must come from image processing and analysis. This improvement can be obtained in the methods used to drive the information from images in their present form or by imaging cell by using additional techniques, and also by extracted more features and combined with the data used during training mode.

Some of most misclassification occurs between the similar algae in shape and structures, which reduced by using some unique features that is not similar in these types of algae. Also, some misidentified regions could be results due the overlapped objects and connected objects. Furthermore, further improvements can be made by using different neural network such as SVM and RBF when data volume of algae and/or the extracted features is increased.

CHAPTER 5

CONCLUSIONS & FUTURE WORKS

5.1 Conclusions

In this study a system has been developed to identify some freshwater algae on microscopic digital images automatically. In this research, an image processing techniques with ANN approach presented to identify and classify five genes of freshwater algae from three different divisions varies in color, size, and shapes. This study illustrated that computational recognition approach is important for freshwater algae, and proved that the classifying process are feasible for automated identification of the selected freshwater algae. The developed system was automatically able to classify 5 kinds of freshwater algae successfully, and experimental results showed that our approach is workable, and had a great accuracy results with more than 93%. Results also indicated that our approach faster in execution, efficient in recognition rate, and easier for using and implementation if compared with similar existing methods.

The better accuracy resulted with higher recognition rate obtained was because we employed some specific technique on system developments, some of these approaches are listed below:

- Appropriate preprocessing techniques applied during images preparation to remove noise, and enhance image contrast assisted in present image feature with clearer details.
- Segmentation process used in this study showed appropriate results if compared with other methods, combination of canny edge detection with morphological operations applied were resulted to isolate each image objects appropriately. This approach facilities the process of image segmentation and object isolation in to sub images, we found that segmentation approach used in this study is showed optimal results in detecting image objects.

- The new developed function of auto-alignments image objects with horizontal coordinate showed a great effect on extracted features, consistency on feature values, reduce redundancy feature values of similar objects which improve overall recognition process significantly.
- We also found that numbers, and types of extracted features significantly affect recognition rate which are considered as important factors that affect the detection process. Increased the number of extracted features also increase the system accuracy, and good selected features this leads to better accuracy results including extracting shape feature, and texture features as used in this research. The process of selecting and using only the most appropriate features as inputs to the neural networks was a major step to get higher accuracy of recognition rate and successful classification process.
- Shape index function that develops in this study to be included as one features showed better improvements on both system accuracy and performance. We recommended the use of this feature for such similar application which required some classification based on shape. Also, we advise biologist to use this feature during their classification based computer approach.
- Training data used for system training process also can be effected system accuracy, for better accuracy achievements. We selected the training data set to represent full range of input data with wide varying diversion of selected images for each object of selected algae.

Overall, we concluded that developed system showed greater variability in cell counting than manual counts, however the conventional manual has a higher degree of accuracy if compared with automated system, but we should consider that the expert performing manual identification had over 10 years experiences in analysis algae samples. Manual identification process is considered time consuming, where developed system reduced time of identification process, even the operator time required to upload image into system is similar to the time required to put the slide under microscope.

Using additional rule based checks greatly improved the ability of the classifiers to reject foreign objects, detritus and aggregates that may have been misclassified by a neural network. Most of these misclassifications were objects mistaken for small objects. Misclassifications were also the result of objects wholly or partly overlapped with other objects in same image area. The main limitation of our system its inability to work well with images that include a huge number of objects.

This thesis addressed very important task of detection, and proved that water monitoring process is feasible to perform automatically, if robust system is developed well for automated detection, recognition, and derivation of quantitative concentration estimates of different freshwater algae using ordinary inverted microscope images. Density and types of freshwater algae detected by developed system as numerical measurements unit for water quality can be give a good indication about the situation of water inside lakes.

5.2 Future Work

The combined neural network and rule-based classification system used in this study is near optimal, so further improvements to identification accuracy must come from image pre-processing, segmentation, and extracted feature. These improvements could either be in the methods used for image processing or by solving the misidentified region, or overlapping objects.

Essential improvements further more should be given to include all freshwater algae division found in Malaysian lakes and improve the database training by extra features to accurate the identification and classification process. Also, future work can be expanded to develop an a repository for all freshwater algae within a database system to be available online via internet that assist researcher and students on their research area, and a new techniques can be developed for semantic algorithms based on ontologies to facilitate the searching, indexing, and tagging the available repository images for fast access and retrieves.

REFERENCES & BIBLIOGRAPHY

REFERENCES

Addy, K., and Green, L. 1996. Algae in Aquatic Ecosystems. *Natural Resources Facts*, 96(4).

Anton, A. 1991. Algae in the conservation and management of freshwaters. *Malayan Nature Society, International Development and Research Centre of Canada.*

Anton, A., J. Alexander, and Chan A. 1998. The Algae of Maliau Rivers. In *Maliau BasinScientific Expedition 1996*, edited by M. Mohamed, W. Sinun, A. Anton, M. N. Dalimin and A. Abdul-Hamid. Malaysia: Universiti Malaysia Sabah.

Azevedo, S. 2001. New Brazilian regulation for cyanobacteria and cyanotoxins in drinking water. In *Fifth International Conference on Toxic Cyanobacteria*. Noosa, Australia.

Blackburn, N., A. Hagström, J. Wikner, R. Cuadros-Hansson, and P. Bjørnsen. 1998. Rapid determination of bacterial abundance, biovolume, morphology, and growth by neural network-based image analysis. *Appl Env Microbiol* 64:3246–3255.

Boddy, L., C. W. Morris, M. F. Wilkins, G. A. Tarran, and P. H. Burkill. 1994. Neural network analysis of flow cytometric data for 40 marine phytoplankton species. *Cytometry,* 15(4):283-93.

Camargo, Julio A., and Álvaro Alonso. 2006. Ecological and toxicological effects of inorganic nitrogen pollution in aquatic ecosystems: A global assessment. *International Journal of Environment,* 32 (6): 831-849.

Canny, J. 1986. Computational Approach to Edge Detection. . *IEEE Transactions on Pattern Analysis and Machine Intell*, 8:619-698.

Chandra, B., and D. Dutta Majumder. 2000. *Digital Image Processing and Analysis*. New Delhi: Prentice Hall of India.

Chen, Jun, and Ping Xie. 2005. Tissue distributions and seasonal dynamics of the hepatotoxic microcystins-LR and -RR in two freshwater shrimps, Palaemon modestus and Macrobrachium nipponensis, from a large shallow, eutrophic lake of the subtropical China. *Jornal of Toxicon* 45(5):615-625.

Chong, A.M.Y. , Y.S. Wong, and N.F.Y. Tam.2000. Performance of Different Microalgal Species in Removing Nickel and Zinc from Industrial Waste Water.*J. of Chemosphere*, 41: 251 - 57.

Chorus, I., I.R. Falconer, H.J. Salas, and J. Bartram. 2000. Health risk caused by freshwater cyanobacteria in recreational waters. *J. Toxicol. Environ. Health* 3:323-347.

Codd, G.A., L.F. Morrison, and J.S Metcalf. 2005. Cyanobacterial toxins: risk management for health protection. *Toxicol. Appl. Pharmacol.* 203:264-272.

Coesel, P.F.M. 1983. The significance of desmids as indicators of the trophic status of freshwaters. *Schweiz Z Hydrol* 45: 388-393.

Coesel, P.F.M. 2001. A method for quantifying conservation value in lentic freshwater habitats using desmids as indicator organisms. *Biodivers Conserv* 10: 177-178.

Culverhouse, P. F., R. Williams, B. Reguera, and V Herry. 2003. Experts make mistakes? A comparison of human and machine identification of dinoflagellates. *Marine Ecology Progress Series,* 247:17–25.

Culverhouse, PF, RG Simpson, R Ellis, JA Lindley, R Williams, T Parisini, B Reguera, I Bravo, R Zoppoli, G Earnshaw, H McCall, and G Smith. 1996. Automatic classification of field-collected dinoflagellates by artificial neural network. *Marine Ecology Progress Series,* 139:281-287.

Culverhouse, Phil F., Robert Williams, Mark Benfield, Per R. Flood, Anne F. Sell, Maria Grazia Mazzocchi, Isabella Buttino, and Mike Sieracki. 2006. Automatic image analysis of plankton: future perspectives. *Marine Ecology Progress Series,* 312:297-309.

Davidson, M.W. 2003. Molekul Ekspresi: microscopist Feature: Wim van Egmond.

Detrich, A., and G. Uhlig. 1984. Stage specific classification of copepods with automatic image analysis *Crustaceana,* 3:159-165.

Dimitrovski, I., Kocev, D., Loskovska, S., and Džeroski, S. 2012. Hierarchical classification of diatom images using ensembles of predictive clustering trees. *Ecological Informatics,* 7(1):19-29, 19-29. doi: 10.1016/j.ecoinf.2011.09.001.

Du Buf, J. and Bayer, M. 2002. ADIAC achievements and future work. In Automatic Diatom Identification. Series in Machine Perception and Artificial Intelligence, 51:289–298.

Embleton, K. V., C. E. Gibson, and S. I. Heaney. 2003. Automated counting of phytoplankton by pattern recognition: a comparison with a manual counting method. *Journal of Plankton Research* 25(6):669-681.

Estep, K.W., F. MacIntyre, E. Hjörleifsson, and J.McN. Sieburth. 1986. MacImage: a user-friendly image-analysis system for the accurate mensuration of marine organisms. *Journal of Mar. Ecol. Prog. Ser.* 33:243-253.

Fahlman, S. E., and C. Lebiere. 1990. The Cascade-Correlation Learning Architecture. *Paper read at Advances in Neural Information Processing Systems* 2, at Los Altos CA.

Falconer, I.R. 2001. Toxic cyanobacterial bloom problems in Australian waters: risk and impacts on human health. *Phycologia* 40:228-233.

Fatimah, M.Y., A.K.M. Mohsin, and Mustafa Kamal A.S. 1984. Phytoplankton Composition and Productivity of a Shallow Lake. *PERTANIKA,* 7(3):101-113.

Gao, S., Jixian Y., Jiayu T., Fang M., Gang T., and Maoan D. 2010. Electro-coagulation–flotation process for algae removal. *Journal of Hazardous Materials,* 177 (1–3): 336-343.

Gilbert, J. (1996). Effect of food availability on the response of planktonic rotifers to a toxic strain of the cyanobacterium Anabaena flos-aquae. *Limnology and Oceanography,* 41(7):1565-1572.

Gonzales, R. C., R. E. Woods, and S. L Eddins. (2004). Digital Image Processing Using MATLAB: Prentice Hall.

Gorsky, G , P Guilbert, and E Valenta. 1989. The Autonomous Image Analyzer enumeration, measurement and identification of marine phytoplankton. *Marine ecology progress series,* 58(1-2):133-142.

Graves, A., Liwicki, M., Fernandez, S., Bertolami, R., Bunke, H., and Schmidhuber, J. 2009. A novel connectionist system for unconstrained handwriting recognition. *IEEE Trans Pattern Anal Mach Intell,* 31(5):855-868.

Hayat, M., Sorayya, M., Aishah, S., and Mogeeb, A. A. Mosleh. 2011. Automatic Recognition System for some cyanobacteria using image processing Techniques and ANN approach. *Paper presented at the IPCBEE, Int. Conf. on Envir. and Comp. Science Singapore*, 19:73-78.

Hillebrand, Helmut, and Ulrich Sommer. 2000. Diversity of benthic microalgae in response to colonization time and eutrophication. *Journal of Aquatic Botany,* 67(3): 221-236.

Ho, S.C. 1995. Statuse of Limnological Research and Training in Malaysia. In *Limnology in Developing Countries* edited by B. Gopal and R. G. Wetzel. New Delhi India: International Association for Limnology.

Hobson, K.A., and H.E. Welch. 1992. Determination of trophic relationships within a high Arctic marine food web using stable. *Marine Ecology Progress Series* 84: 9–18.

Hopfield, J. J. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the USA* 79, (8):2554–2558.

Jähne, B. 2002. Digital Image Processing: Prentice Hall.

Jeffries, H. P., K. Sherman, and R. Maurer. 1980. Computer processing of zooplankton samples. In *Estuarine Perspectives*, edited by V. Kennedy. New York: Academic Press.

Jeffries, HP, MS Berman, AD Poularikas, C Katsinis, I Melas, Sherman, and Bivins K, L. 1984. Automated sizing, counting, and identification of zooplankton by pattern Recognition. Journal of *Mar Biol,* 78:329–334.

Johnstone, C., J.G. Day, H. Staines, and E.E. Benson. 2006. Development of 2,2′-Azinobis-(3-Ethylbenzothiazoline- 6-Sulfonic Acid)Radical Cation Decolourisation Assay for Evaluating Total Antioxidant Status in an Alga Used to Monitor Environmental Impacts in Urban Aquatic Habitats. *Ecol. Indicators*, 6: 280-89.

Junna, C., Guangrong, J., Chen, F., and Haiyong, Z. 2009. Application of Connected Morphological Operators to Image Smoothing and Edge Detection of Algae. *Paper presented at the ITCS, International Conference on Information Technology and Computer Science*, Kiev, Ukraine. 2:73-76.

Kamath, S. B., Chidambar, S., Brinda, B. R., Kumar, M. A., Sarada, R., and Ravishankar, G. A. (2005). Digital image processing—an alternate tool for monitoring of pigment levels in cultured cells with special reference to green alga Haematococcus pluvialis. *Biosensors and Bioelectronics,* 21(5):768-773. doi: 10.1016/j.bios.2005.01.022.

Katsinis, C, A. D Poularikas, and H. P Jeffries. 1984. Image processing and pattern recognition with applications to marine biological images. *Paper read at SPIE 28th Annual International Technical Symposium on Optics and Electro-optics*, Aug, at San Diego.

Khan, I. S. A. Nather. 1985. Studies on the water quality and periphyton community inLinggi River Basin, Malaysia, Zoology Department, Universiti Malaya, Kuala Lumpur.

Khan, I. S. A. Nather. 1990. Assessment of Water Pollution using Diatom Community Structure and Species Distribution — A Case Study in a Tropical River Basin. *Internationale Revue der gesamten Hydrobiologie und Hydrographie* 75(3):317-338.

Khan, I. S. A. Nather. 1991. Effect of urban and industrial wastes on species diversity of the diatom community in a tropical river, Malaysia. *Hydrobiologia,* 224(3):175-184.

Knoben, R.A.E., C. Roos, and M.C.M van Oirschot. 1995. Biological Assessment Methods for Watercourses. UN/ECE Task Force on Monitoring & Assessment.

Kohavi, and F. Provost. 1998. Glossary of terms. *Journal of Machine Learning,* 30(2-3):271–274.

Leclercq. 1988. Utilization de trios indices, chimique, diatomique et biocénotique, pour l'évaluation de la qualité de l'eau de la Joncquiere, riviére calcaire polluée par le village de Doische (Belgique, Prov. Namur). *Mém. Soc. Roy. Bot. Belg.* 10: 26-34.

Lim, Jae S. (1990). Two-Dimensional Signal and Image Processing. Englewood Cliffs, NJ: Prentice Hall.

Luo, Q., Yahui G., Jinfei L., Changping C., Junrong L., and Chenhui Y. 2011. Automatic Identification of Diatoms with Circular Shape using Texture Analysis. *Journal of Software,* 6(3: 428-435).

Masseret, E., Christian A., and Gilles B. 1998. Changes in the structure and metabolic activities of periphytic communities in a stream receiving treated sewage from a waste stabilization pond. *Water Research,* 32(8): 2299-2314.

Maznah, Wan, and M Mansor. 2002. Aquatic pollution assessment based on attached diatom communities in the Pinang River Basin, Malaysia. *Hydrobiologia,* 487:229–241.

Maznah, Wan, and Mashhor Mansor. 1999. Benthic diatoms in the Pinang River (Malaysia) and its tributaries with emphasis on species diversity and water quality. 1(4):103-118.

McCormick, P.V., and J. Cairns, Jr. 1994. Algae as indicators of environmental change. *Journal of Applied Phycology,* 6: 509-526.

Morris, Tim. (2004). *Computer Vision and Image Processing.* : Palgrave Macmillan.
Patrick, R. 1936. A taxonomic and distributional study of some diatoms from Siam and the Federated Malay States. *Proceeding of Academy of Natural Sciences,* 88: 367-470.

Patrick, R. 1994. What Are the Requirements for an Effective Biomonitor. In Biological Monitoring of Aquatic Systems Lewis Publishers, 23 – 29.

Pipan, T. 2000. Biological assessment of stream water quality - the example of the Reka river. *Journal of Acta carsologica*, 29(1): 201-222.

Prowse, G A. 1957. An introduction to the desmids of Malaya. *Malayan Nat. J.* 11:42-58.

Prowse, G. A. 1958. The Eugleninae of Malaya. *Gards. Bull. Str. Settl.* 16:136-145.

Prowse, G. A. 1960. New and unusual flagellatta in Malaya. Paper read at Centenary and Bicentenary Congress of Biology, at Singapore.

Rauch, A., Fesl, C., and Schagerl, M. 2006. Influence of environmental variables on algal associations from a floating vegetation mat (Schwingmoor Lake Lunzer Obersee, Austria). *Journal of Aquatic Botany*, 84(2): 129-136. doi: 10.1016/j.aquabot.2005.08.004.

Schmidhuber, J. 1989. A local learning algorithm for dynamic feedforward and recurrent networks. *Connection Science*, 1(4):403–412.

Schultze-Lam, S., G. Harauz, and T. Beveridge. 1992. Participation of a cyanobacterial S layer in fine-grain mineral formation. *J Bacteriol* 174:7971–7981.

Shapiro, Linda G., and Stockman, George C. (2001). Computer Vision: Prentice Hall.
Shutes, R.B.E. (2001). Artificial wetlands and water quality improvement. *Environment International*, 26:441-447.

Simpson, R, P.F Culverhouse, R Ellis, and R Williams. 1991. Classification of euceratium gran. in neural networks. *Paper read at Neural Networks for Ocean Engineering, IEEE Conference 15-17 Aug.* , at Washington, USA.

Simpson, R, PF Culverhouse, R Ellis, and R Williams. 1994. Classificatlon of Ceratium and Dinophyceae plankton species. In *Proc Int Conf on Artificial Neural Networks, ICANN '94*, edited by M. M and M. PG. Sorrento, Italy: Springer-Verlag, Berlin.

Simpson, R, R Williams, R Ellis, and P.F Culverhouse. 1992. Biological pattern recognition by neural networks *MARINE ECOLOGY PROGRESS SERIES* 79:303-308.

Simpson, R., P. F. Culverhouse, R. Williams, and R. Ellis. 1993. Classification of Dinophyceae by artificial neural networks. In *Toxic Phytoplankton Blooms in the Sea*, edited by T. J. Smayda and Y. Shimizu. Amsterdam: Elsevier.

Simpson, R., P. F. Culverhouse, R. Williams, and R. Ellis. 1993. Classification of Dinophyceae by artificial neural networks. In *Toxic Phytoplankton Blooms in the Sea*, edited by T. J. Smayda and Y. Shimizu. Amsterdam: Elsevier.

Sorayya, M., Sharifah M. S., Sarinder K. K., Pozi M., and Aishah Salleh. 2011. Assessment of Predictive Models for Chlorophyll-a Concentration of a Tropical Lake. *Paper presented at the 10th International Conference on Bioinformatics*,12: 1-11.

Sosik, Heidi M., and Robert J. Olson. 2007. Automated taxonomic classification of phytoplankton sampledwith imaging-in-flow cytometry. *Limnol. Oceanogr. Methods,* 204–216.

Tang, X., Lin, F., Samson, S. and Remsen, A. 2006. Binary Plankton Image Classification. *IEEE Journal of Oceanic Engineering*, 31(3):728 - 735.

Tang, X., Stewart, W. K., Huang, H., Gallager, S. M., Davis, C. S., Vincent, L., and Marra, M. 1998. Automatic Plankton Image Recognition. *Artificial Intelligence Review,* 12(1):177-199. doi: 10.1023/A:1006517211724.

Tiffany, L. H., and M. E. Britton. 1971. The algae of Illinois. New York: Hafner Publishing Company.

Tisdale, E. S. 1931. Epidemic of intestinal disorders in Charleston, W. Va., occurring simultaneously with unprecedented water supply conditions. *American Journal of Public Health* 21:198–200.

Verikas, A., Zivile K., Marija B., and Adas G. 2010. Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey. *Soft Computing - A Fusion of Foundations, Methodologies and Applications,* 14(9):995-1010.

Walsby, Anthony E., and Avril Avery. 1996. Measurement of filamentous cyanobacteria by image analysis. *Journal of Microbiological Methods,* 26 (1–2):11-20.

Wilkins, M. F., Boddy, L., Morris, C. W., and Jonker, R. R. 1999. Identification of phytoplankton from flow cytometry data by using radial basis function neural networks. *Appl Environ Microbiol,* 65(10): 4404-4410.

William, H., Teukolsky, Saul A., Vetterling, William T., and Flannery, B. P. (2007). Section 16.5. Support Vector Machines, Numerical Recipes: The Art of Scientific Computing (3rd Ed.). New York, Cambridge University Press.

Wu, Di, Wang, S., Liu, K., Yu, X., He, Y., and Wang, Z. 2012. Rapid measurement of morphological features of Spirulina microalgae filaments using microscopy and image processing algorithms. *Biosystems Engineering*, 112(1): 35-41, 35-41. doi: 10.1016/j.biosystemseng.2012.02.002.

Wurbs, R. A., and James, W. P. (2002). Water Resources Engineering: Prentice Hall.

Yao, Zhihong, Minrui Fei, Kang Li, Hainan Kong, and Bo Zhao. 2007. Recognition of blue-green algae in lakes using distributive genetic algorithm-based neural networks. *Neurocomputing,* 70(4–6):641-647.

Yasser, A. Al-gahwari. 2007.Physico-Chemical Parameters and Microorganisms as Water Quality Indicators of Teluk Bahang Reservoir and Batu Ferringhi Treatment Plant. UNIVERSITI SAINTS MALAYSIA, PhD Thesis, 1-24.

Yee, Paul V., and Simon Haykin. 2001. Regularized Radial Basis Function Networks: Theory and Applications.

BIBLIOGRAPHY

Aik, Lee Kwong. 2003. *studies on the phytoplankton in the lower bisa area of Putrajaya lake,Putrajaya*. Dissertation, Technology(Enviromental Management), University of Malaya, Kualalumpur - Malaysia.

Alcocer, J., L. A. Oseguera, E. Escobar, L. Peralta, and A. Lugo. 2004. Phytoplankton Biomass and Water Chemistry in Two High-Mountain Tropical Lakes in Central Mexico. *Arctic, Antarctic, and Alpine Research*. 36 (3):342-346.

Álvarez-Borrego, Josué, and Selene Solorza. 2010. Comparative analysis of several digital methods to recognize diatoms. *Hidrobiológica* . 20 (2):158-170.

Anton, A., P.L. Teoh, S.R. Mohd-Shaleh, and N. Mohammad-Noor. 2008. First occurrence of Cochlodinium blooms in Sabah, Malaysia. *Harmful Algae* . 7:331–336.

Bachiller, Eneko, and Jose Antonio Fernandes. 2011. Zooplankton Image Analysis Manual: automated identification by means of scanner and digital camera as imaging devices. Vol. 18, *Marine Research Division*. Herrera Kaia, Portualdea: Unidad de Investigación Marina de Tecnalia.

Barsanti, Laura, and Paolo Gualteri. 2006. Algae: Anatomy, Biochemistry, And Biotechnology: CRC PressINC.

Botes, L. 2011. Phytoplankton Identification Catalogue, Global Ballast Water Management Programme. GloBallast Monograph.

Bouvy, Marc, Diogo Falcão, Mauro Marinho, Marc Pagano, and Ariadne Moura4. 2000. "Occurrence of Cylindrospermopsis (Cyanobacteria) in 39 Brazilian tropical reservoirs during the 1998 drought." *Aquatic Microbial Ecology* . 23 (1):13–27. doi: 10.3354/ame023013.

Ellis, R., R. Simpson, P.F. Culverhouse, and T. Parisini. 1997. Committees, Collectives and Individuals: Expert Visual Classification by Neural Network. *Neural Computing & Application* . 5 (2):99-105.

Ernst, B., S. Neser, E. O'Brien, S. J. Hoeger, and D. R. Dietrich. 2006. Determination of the filamentous cyanobacteria Planktothrix rubescens in environmental water samples using an image processing system. *Harmful Algae*. 5:281–289.

Fernandez-Canque, H L, B. Beggs, E. Smith, T. Boutaleb, H. V. Smith, and S. Hintea. 2006. Micro-Organisms Detection in Drinking Water Using Image Processing. Paper presented at 6th International IEEE EMBS Special Topic Conference on Information Technology Applicationsin Biomedicine (ITAB) Conference, at Ioannina–Epirus, Greece.

Fodorpataki, Laszlo, Csaba Bartha, and Zsolt Gyula Keresztes. 2009. Stress-physiological reactions of the green alga Scenedesmus opoliensis to water pollution with herbicides. *Analele Universitatii din Oradea, Fascicula Biologie*. 16 (1):51-56.

Grima, Carlos, Fernando Tadeo, and M Teresa Álvarez. 2003. Diatoms Classification Using Frequency Domain Techniques. *Empresa Depuradora de Aguas Residuales,Ávila* 1 - 5.

Grimm, M., and K. Kroschel. 2007. Robust Speech Recognition and Understanding: I-Tech Education and Publ.

He, Qin. 1999. Neural Network and Its Application in IR: a Graduate School of Library and Information Science. University of Illinois at Urbana-Champaign.

Hötzel, Gertraud, and Roger Croome. 1999. A Phytoplankton Methods Manual for Australian Freshwaters, Land and Water Resources Research and Development Corporation. *Canberra ACT Land and Water Resources Research and Development Corporation.*

Hudnell, H. Kenneth. 2008. *Cyanobacterial Harmful Algal Blooms*. Vol. 619. New York: Springer.

Jabla, Andrei C., Michael H.F, Wilkinson, and Jos B.T.M. Roerdink. 2004. Automatic Segmentation of Diatom Images for Classification. *Microscopy Research and Technique*. 65 ((1-2)):72-85.

Jalba, Andrei C., Michael H.F. Wilkinson, Jos B.T.M. Roerdink, Micha M, Bayer, and Stephen Juggins. 2005. Automatic Diatom Identification using Contour Analysis by Morphological Curvature Scale Spaces. *Machine Vision and Applications*. 16 (4):217-228.

Ji, Guangrong, Rui Nian, Shiming Yang, Lijian Zhou, and Chen Feng. 2006. *Cellular Recognition for Species of Phytoplankton Via Statistical Spatial Analysis*. Vol. 345: Springer Berlin Heidelberg.

Kilham, Peter, and Robert E. Hecky. 1988. Comparative ecology of marine and freshwater phytoplankton. Vol. 33: *American Society of Limnology and Oceanography*, Inc.

Lai, Mak Yoke. 2006. Classification of Lichen Species using Artificial Neural Networks, Computer Science and Information Technology, University Malaya, Malaysia.

Malek, S., A. Salleh, and M. S. Baba. 2009. Prediction of Population Dynamics of Bacillariophyta in the Tropical Putrajaya Lake and Wetlands (Malaysia) by a Recurrent Artificial Neural Networks. *Paper read at Environmental and Computer Science, 2009. ICECS '09. Second International Conference on*, 28-30 Dec. 2009.

Mansor, Mashhor, and Ganesan Muthaiya. 1998. Notes on Montane Stream Algae from Sayap-Kinabalu Park, Sabah. *ASEAN Review of Biodiversity and Environmental Conservation (ARBEC)*:123-128.

McCall, Helen, Isabel Bravo, J.Alistair Lindley, and Beatriz Reguera. 1996. Phytoplankton recognition using parametric discriminants. *Journal of Plankton Research*. 18 (3):393-410.

Merican, Faradina, Wan Asmadi W.A, Wan Maznah W.O, and Mashhor M. 2006. A Note On The Freshwater Algae Of Gunung Stong, Kelantan, Malaysia. *Bioscience* no. 17 (1):65–76.

Motamarria, Srinivas, and Dominic L. Boccellia. 2012. Development of a Neural-Based Forecasting Tool to Classify Recreational Water Quality Using Fecal Indicator Organisms. *Water Research* no. 46 (14):4508-4520.

Ognjanova-Rumenova, Nadja, and Kalina Manoylov. 2006. *Advances in Phycological Research.* Bulgaria.

Pacheco, Pech, G. Cristóbal Pérez J. L., J. Alvarez Borrego, and L. Cohen. 2001. Automatic system for phytoplanktonic algae identification. *Limnetica* no. 20 (1):143-158.

Pereira, Helder Cunha, Norman Allott, Catherine Coxon, Owen Naughton, Paul Johnston, and Laurence Gill4. 2011. Phytoplankton of turloughs (seasonal karstic Irish lakes). *Journal of Plankton Research*. 33 (3):385–403.

Potamianos, G., C. Neti, G. Gravier, A. Garg, and A. W. Senior. 2003. Recent advances in the automatic recognition of audiovisual speech. *Proceedings of the IEEE*. 91 (9):1306-1326. doi: 10.1109/JPROC.2003.817150.

Poulíčková, Aloisie, Martin Duchoslav, and Martin Dokulil. 2004. Littoral diatom assemblages as bioindicators of lake trophic status: A case study from perialpine lakes in Austria. *Eur J Phycol*. 39 (2):143-152. doi: 10.1080/0967026042000201876.

Reynolds, C.S. 2006. *The Ecology of Phytoplankton*: Cambridge University Press.

Rodenacker, K., P. Gais, U. Jutting, and B. A. Hense. 2001. (Semi-) automatic recognition of microorganisms in water. *Paper Proceedings read at International Conference on Image Processing*.

Rollanda, Anne, Frederic Bertrandb, Myriam Maumyb, and Ste´phan Jacqueta. 2009. Assessing phytoplankton structure and spatio-temporal dynamics in a freshwater ecosystem using a powerful multiway statistical analysis. *Water Research*. 43(13):3155–3168.

Ronald, Semyalo, Thomas Rohrlack, Kayiira Duncan, Kizito Yusuf S., Byarujali Stephen, Nyakairu George, and Larsson Petter1. 2010. On the diet of Nile tilapia in two eutrophic tropical lakes containing toxin producing cyanobacteria. *Limnologica*. 41 (1):30-36. doi: 10.1016.

Rouf, A. J. M. Abdur, Siew-Moi Phang, and M. Azmi Ambak. 2010. Depth distribution and ecological preferences of periphytic algae in Kenyir Lake, the largest tropical reservoir of Malaysia*. *Chinese Journal of Oceanology and Limnology*. 28 (4):856-867.

Sara, J. Ross. 2008. Molecular Phylogeography and Species Discrimination of Freshwater Cladophora (Cladophorales, Chlorophyta) in North America. Waterloo, Ontario, Canada: University of Waterloo in fulfilment.

Seckbach, Joseph. 2007. Algae and cyanobacteria in extreme environments. Edited by Joseph Seckbach. Vol. 11: Dordrecht : Springer.

Sidik, Madihah Jaffar, Md. Rashed-Un-Nabi, and Md. Azharul Hoque. 2008. Distribution of phytoplankton community in relation to environmental parameters in cage culture area of Sepanggar. *Estuarine, Coastal and Shelf Science*. 80 (2):251-260.

Solé, A., E. Diestra, and I. Esteve. 2009. Confocal Laser Scanning Microscopy Image Analysis for Cyanobacterial Biomass Determined at Microscale Level in Different Microbial Mats. *Microbial Ecology* . 57 (4):649–656.

Sole, A., J. Mas, and I. Esteve. 2007. A new method based on image analysis for determining cyanobacterial biomass by CLSM in stratified benthic sediments. *Ultramicroscopy*. 107:669–673.

Stich, T.J. 1997. The Application of Artificial Neural Networks to Monitoring and Control of an Induction Hardening Process. Vol. 16: Southern Illinois University at Carbondale.

Tabout, H., A. Souissi, and A. Sbihi. 2007. Region-Based Automated Relevance Feedback in Algae Image Retrieval. *Computer Science and Network Security*. 7 (12):124-131.

Takatoo, M., C. Onuma, and Y. Kobayashi. 1996. Detection of Objects Including Persons Using Image Processing. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*: IEEE Computer Society.

Thiel, Stefan, and Ron J. Wiltshire. 1995. The automated detection of cyanobacteria using ddigital image processing techniques. *Environmental international*. 21 (2):233-236.

Thiel, Stefan U., Ron J. Wiltshire, and Lance J. Davies. 1995. Automated object recognition of blue-green algae for measuring water quality—A preliminary study. *Water Research*. 29 (10):2398-2404.

Thiel, S. U., and J. A. Ware. 1995. Determination of water quality in fresh water lakes. Paper read at. *Fifth International Conference on Image Processing and its Applications*, 4-6 Jul 1995.

Thompson, J B, S Schultze-Lam, T J Beveridge, and D J Des Marais. 1997. Whiting events: biogenic origin due to the photosynthetic activity. C*yanobacterial picoplankton*. Vol. 42.

U, Gires, Leaw C.P, and Asmat A. 2002. Increasing importance of harmful algal blooms in Malaysia. Paper read at Proceedings of Regional Symposium on Environment and Natural Resources, at Kuala Lumpur, Malaysia.

Verikas, A., A. Gelzinis, M. Bacauskiene, I. Olenina, S. Olenin, and E. Vaiciukynas. 2012. Automated image analysis- and soft computing-based detection of the invasive dinoflagellate Prorocentrum minimum (Pavillard) Schiller. *Expert Systems with Applications*. 39 (5):6069-6077. doi: 10.1016/j.eswa.2011.12.006.

Verikas, A., A. Gelzinis, M. Bacauskiene, I. Olenina, S. Olenin, and E. Vaiciukynas. 2012. Phase congruency-based detection of circular objects applied to analysis of phytoplankton images. *Pattern Recogn.* 45 (4):1659-1670. doi: 10.1016/j.patcog.2011.10.019.

Wehr, J.D., and R.G. Sheath. 2003. *Freshwater Algae of North America: Ecology and Classification*: Academic Press.

Weller, Andrew F., Anthony J. Harris, and J. Andrew Ware. 2006. Artificial neural networks as potential classification tools for dinoflagellate cyst images: A case using the self-organizing map clustering algorithm. *Review of Palaeobotany and Palynology*. 141 (3-4):287–302.

Whitton, B.A. 2000. The Ecology of Cyanobacteria: Their Diversity in Time and Space. Springer.

Wilkins, M. F., Lynne Boddy, C. W. Morris, and R. R. Jonker. 1999. Identification of Phytoplankton from Flow Cytometry Data by Using Radial Basis Function Neural Networks. *Applied and Environmental Microbiology (AEM)*. 65 (10):4404–4410.

Wilkinson, M. H. F., J. B. T. M. Roerdink, S. Droop, and M. Bayer. 2000. Diatom contour analysis using morphological curvature scale spaces. *Paper Proceedings on 15th International Conference on Pattern Recognition*.

Young, D., C.A. Glasbey, A.J. Gray, and N.J. Martin. 1995. Cell identification in Differential Interference Contrast microscope images using template matching. In *Image Analysis*.

Young, D., and A. J. Gray. 1997. Semi-automatic boundary detection for identification of cells in DIC microscope images. *Paper presented at Sixth International Conference on Image Processing and Its Applications.* 14-17 Jul 1997.

APPENDICES

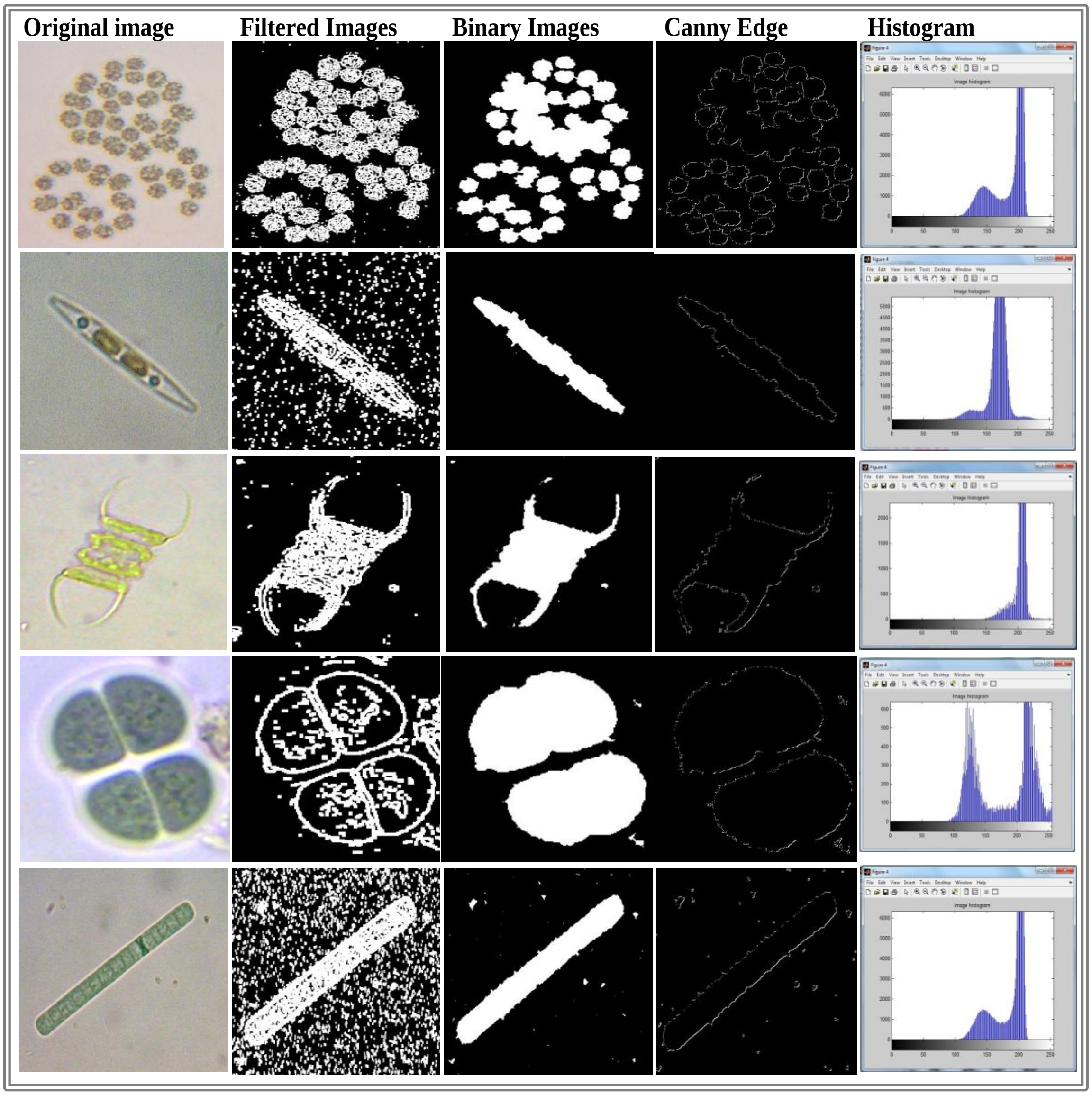
Appendix I -A

Table 2 Most common division and genera in Putrajaya Lake and Wetlands as reported by Sorayya, M., Aishah, S., Sapiyan, B on (2009).

Division	Genera
Bacillariophyta	*Cymbella* C.Agardh, *Cyclotella* Brébisson, *Diatoma* Bory de St-Vincent, *Fragilaria* Lyngbye, *Gomphonema* Ehrenberg, *Navicula* Bory de Saint-Vincent, *Pinnularia* Ehrenberg, *Stauroneis* Ehrenberg, *Surirella* Turpin, *Synedra* Ehrenberg, *Tabellaria* Ehrenberg ex F.T.Kützing, *Nitzschia* Hassall
Chlorophyta	*Closterium* Nitzsch ex Ralfs, *Cosmarium* Corda ex Ralfs, *Pachycladon* G.M.Smith, *Pediastrum* Meyen, *Scenedesmus* Meyen, *Staurastrum* Meyen ex Ralfs, *Ankistrodesmus* Corda, *Chlorella* M.Beijerinck, *Crucigenia* Morren, *Closteriopsis* Lemmermann, *Tetraedron* Kützing, *Closterium* Nitzsch ex Ralfs,
Chrysophyta	*Dinobryon* Ehrenberg
Cyanobacteria	*Chroococcus* Nägeli, *Oscillatoria* Vaucher ex Gomont, *Microcystis* Lemmermann, *Lyngbya* C.Agardh ex Gomont
Pyrrophyta	*Ceratium* F.Schrank, *Peridinium* Ehrenberg

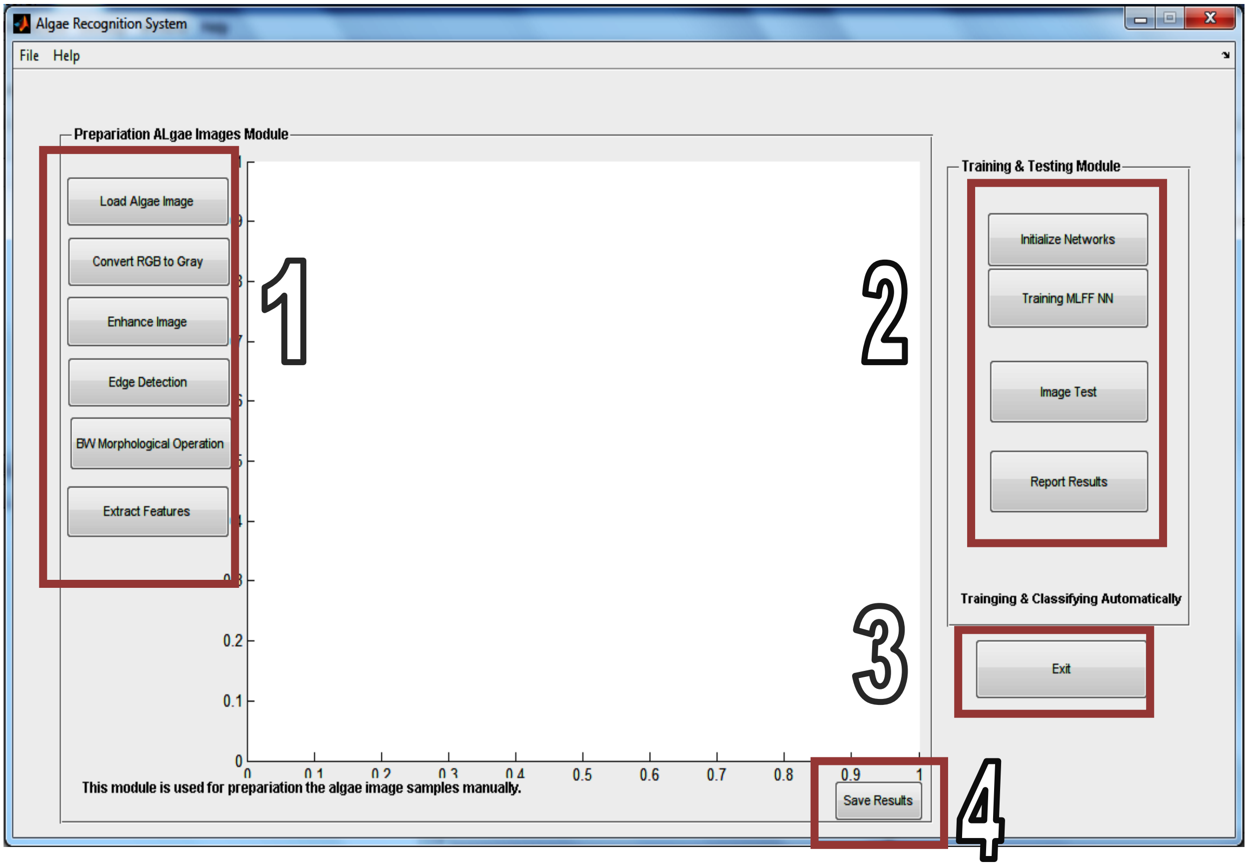
Appendix I -B

Image pre-processing and segmentation steps for all freshwater algae in this study.



Appendix II

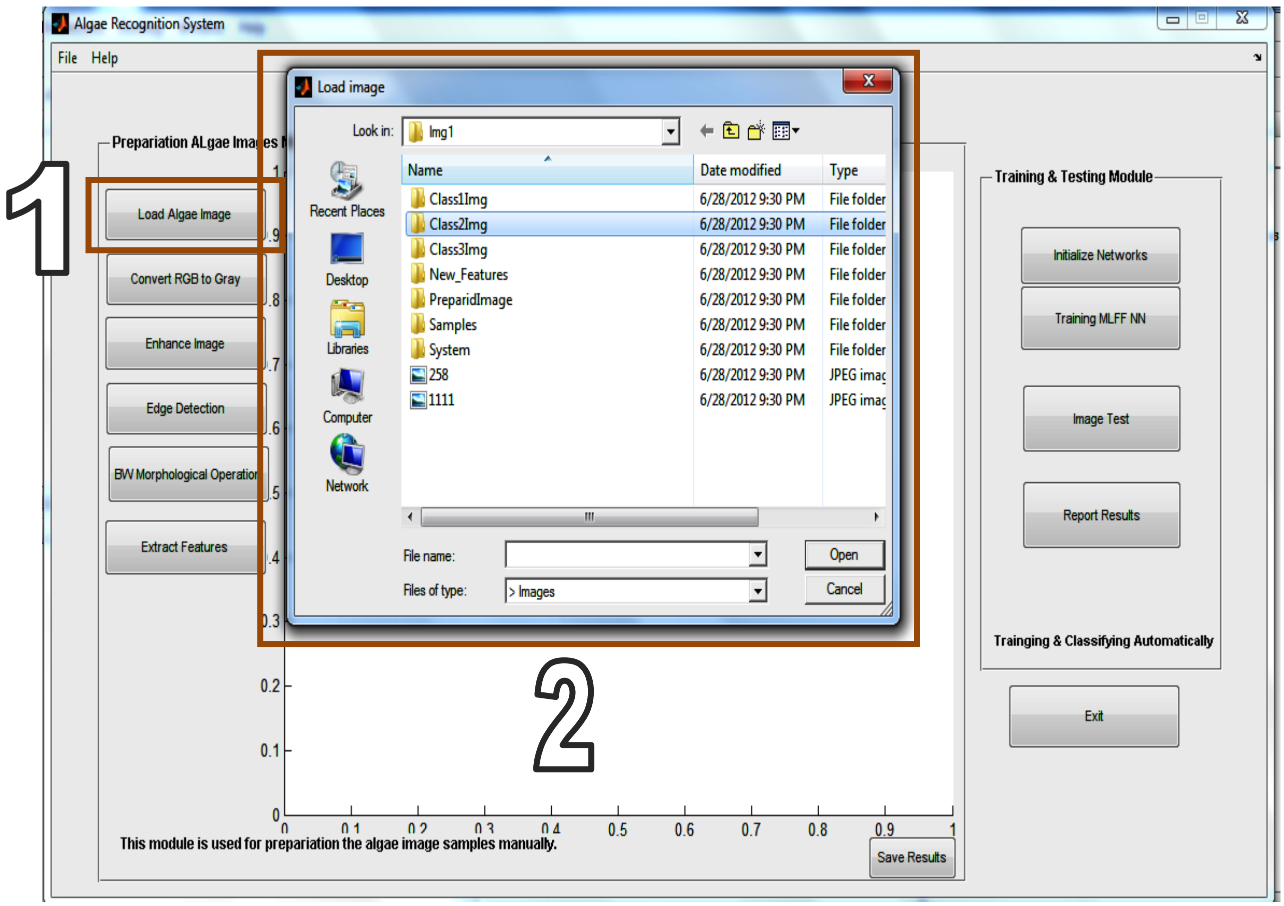
Procedure for preparation, training and testing of algae picture.



1 Preparation Algae picture
2 Training & Testing Algae picture
3 Exit the program
4 Save the result after testing the program

Appendix III

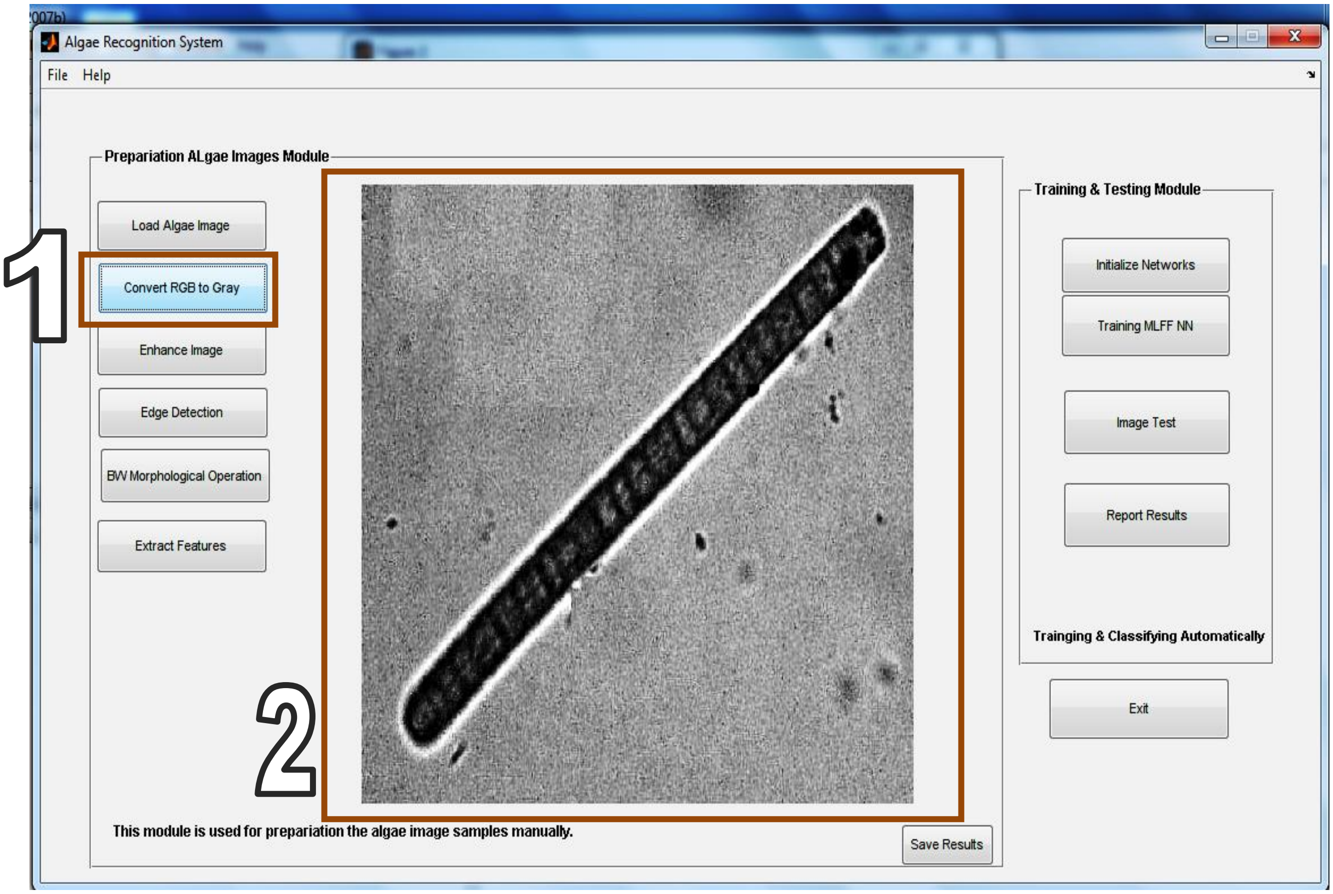
Procedure for uploading algae images.



1 First button to load image file
2 Select algae image

Appendix IV

Procedure for pre-processing of image sample of algae

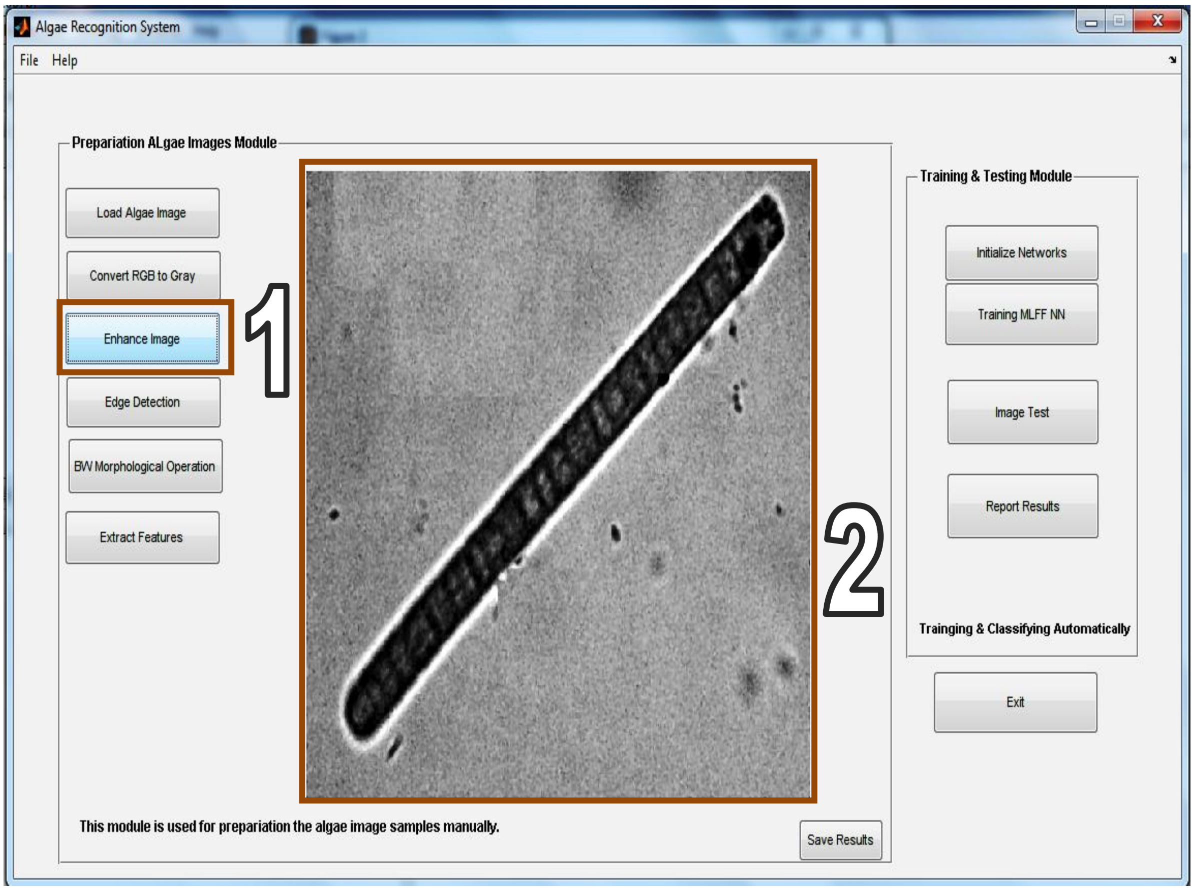


1 Step 2 in the above converts the color image to gray scale

2 Image appear in the work space in gray color

Appendix V

Procedure for image enhancement

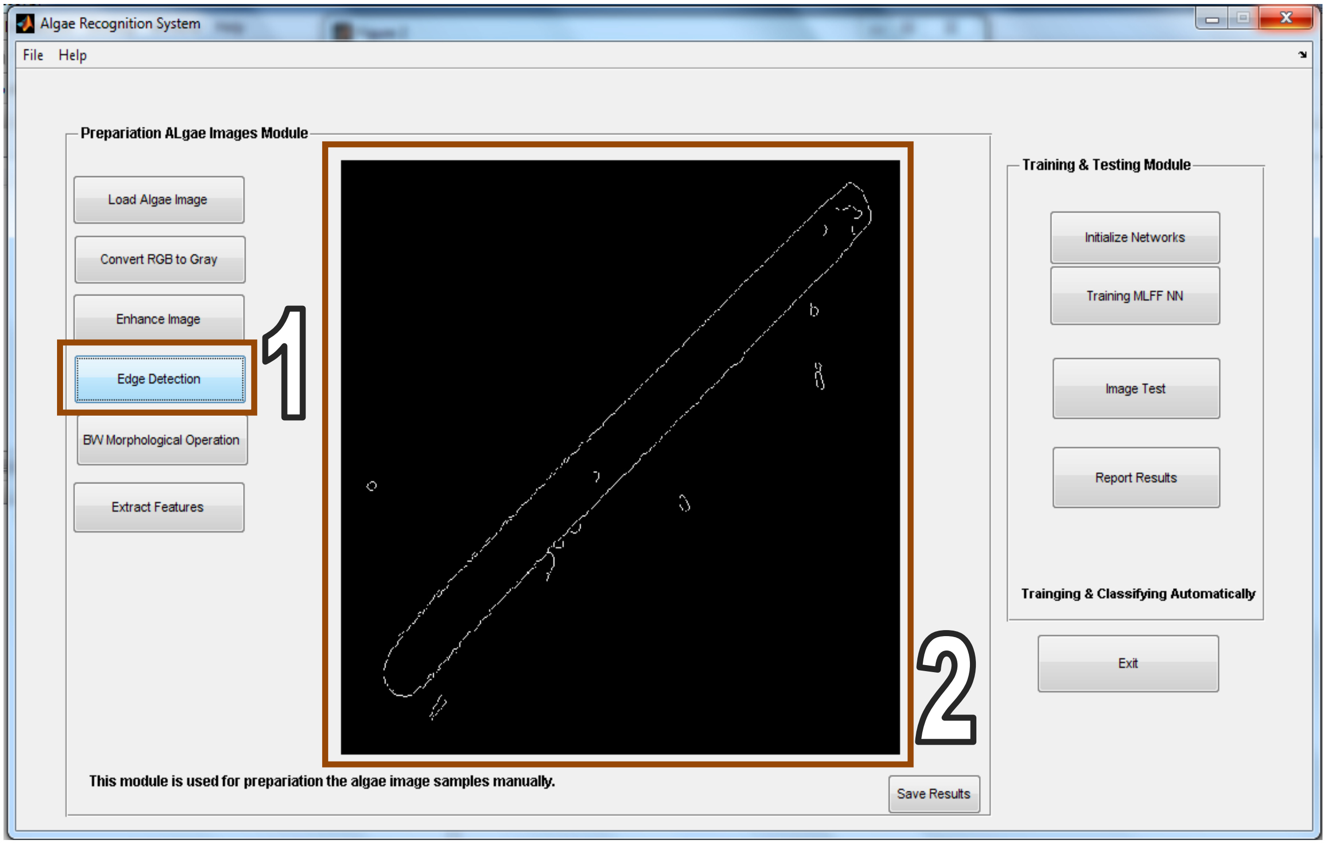


1 Enhance algae image

2 Algae in work space appear clear with all details

Appendix VI

Procedure for edge detection

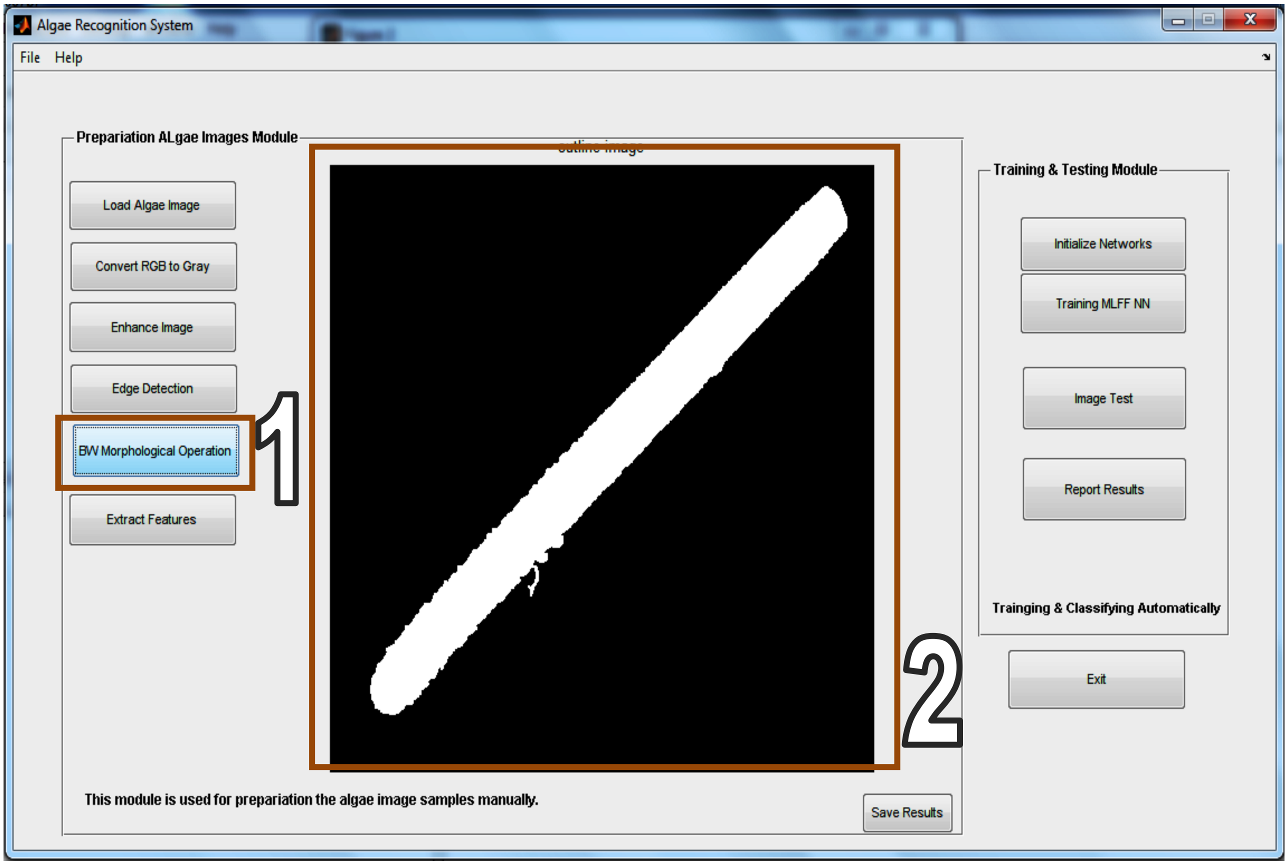


1 Edge detection operation

2 The image converts from gray scale to binary and remove the small object

Appendix VII

Procedure for morphological operation

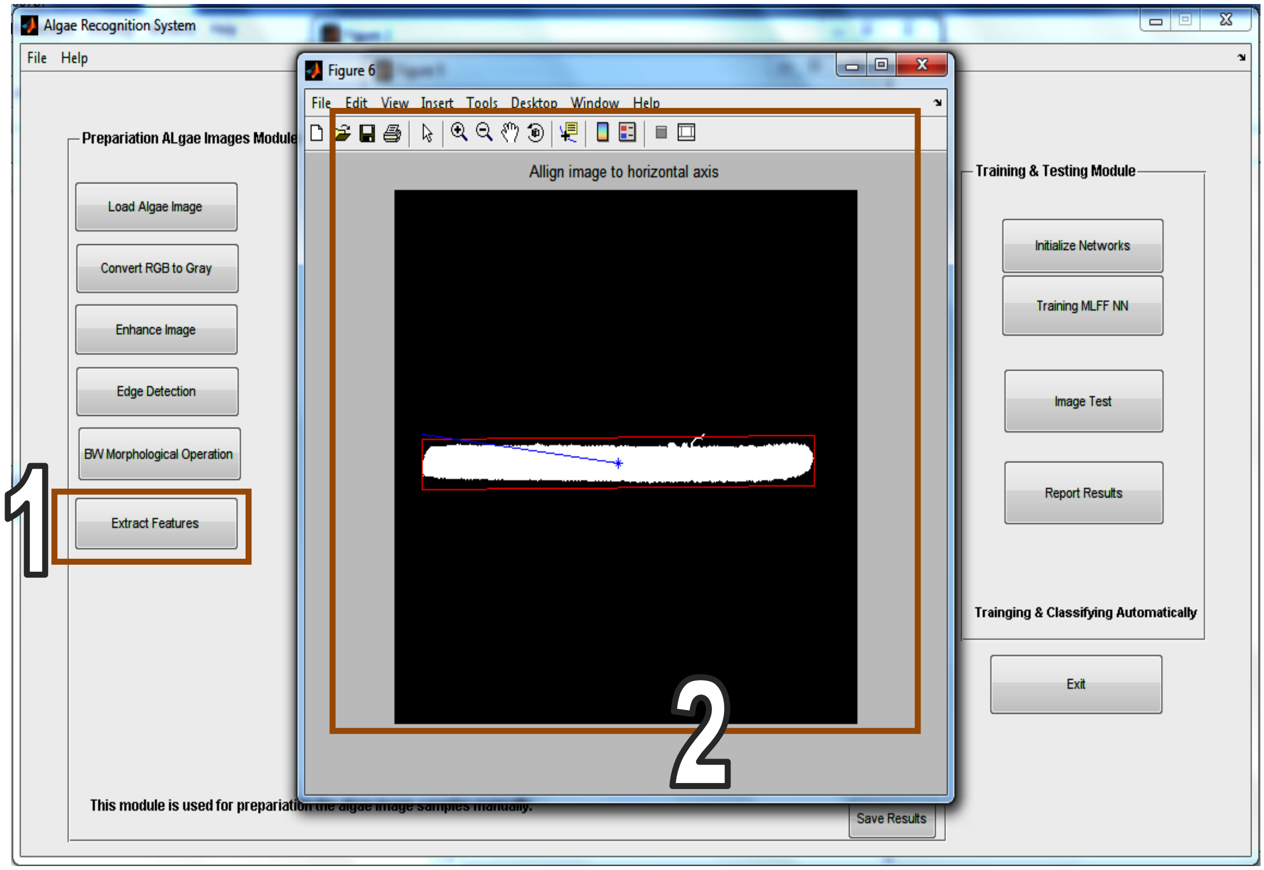


1 Morphological operation

2 We use this to remove small objects, close the open region of objects, and to fill in the object boundary

Appendix VIII

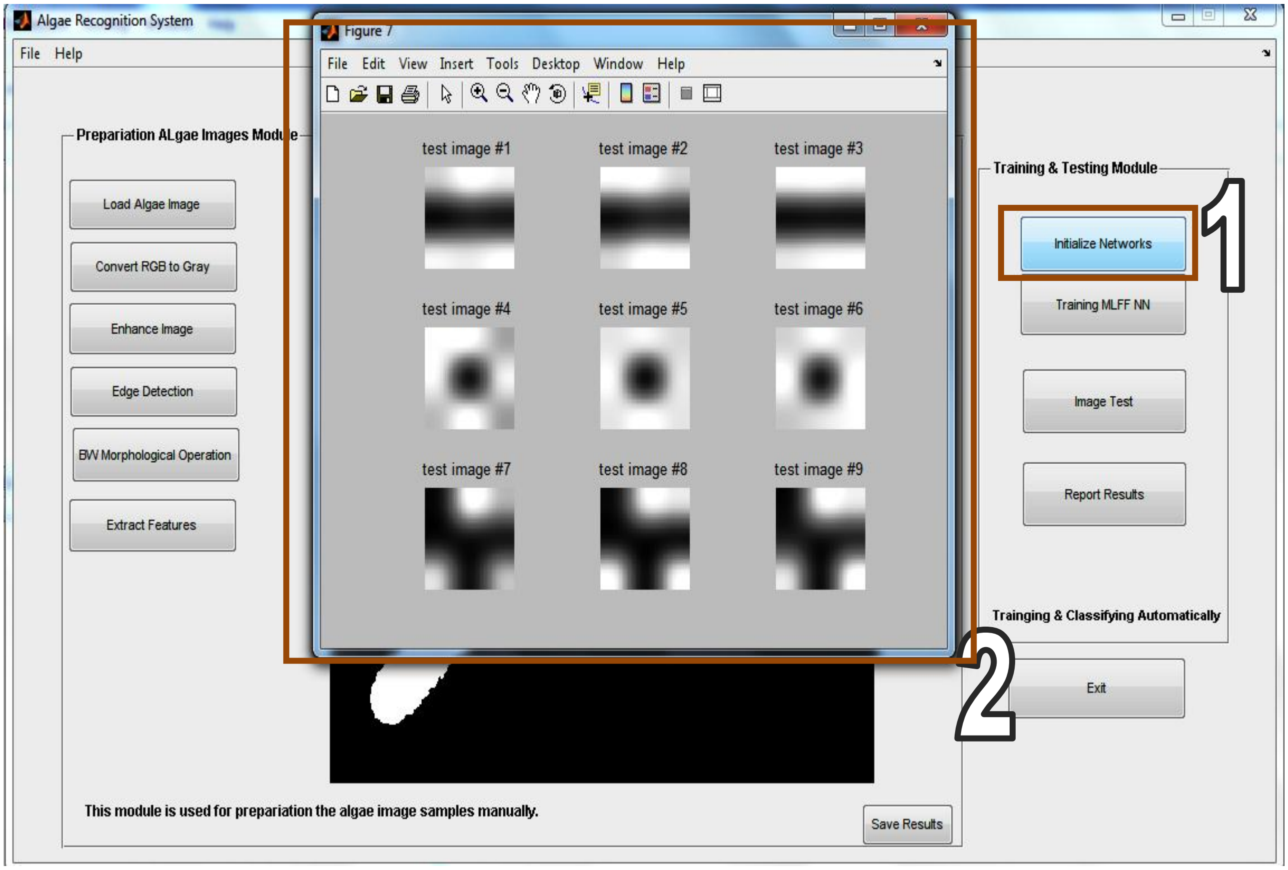
Procedure for feature extraction



1 Feature extraction: process of determining various attributes and properties associated within a region or objects. 2 Automatic rotation for objects apply in this step to be aligned with the horizontal axis

Appendix IX

Procedure for neural network initialization.

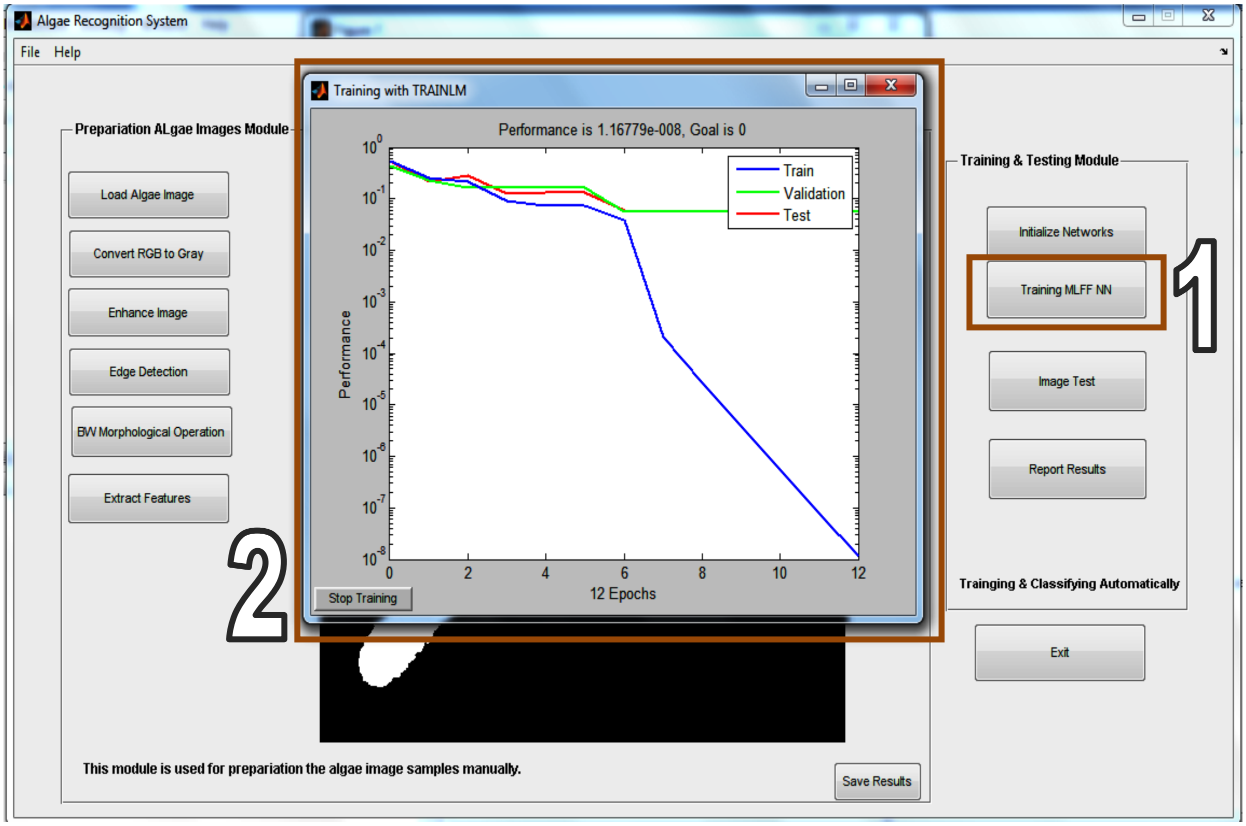


1 Initialize network

2 Display some features of images

Appendix X

Procedure for training MLFF NN

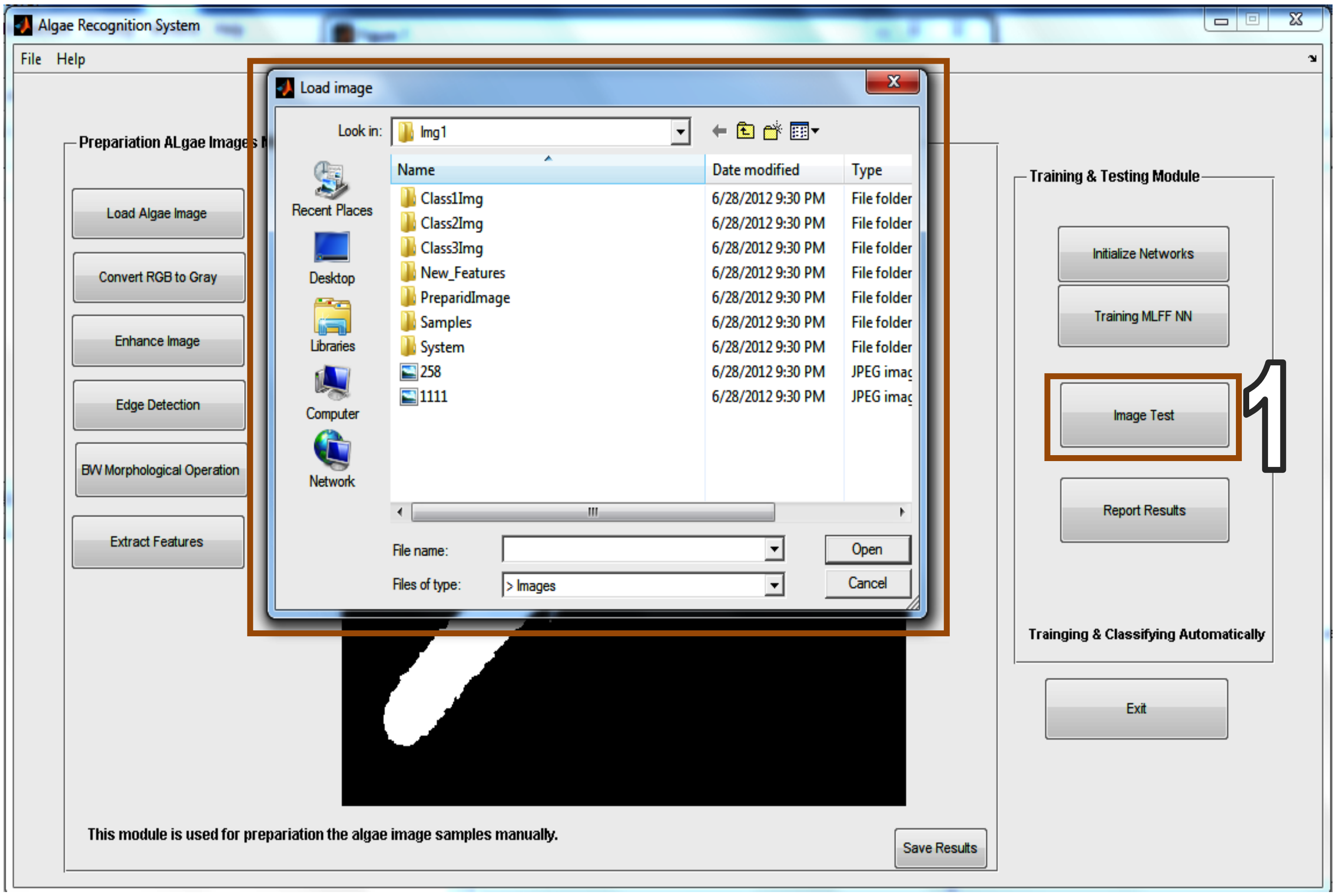


1 Training MLFF NN: it use for training and recognition process of selected algae image

2 The training process is implemented to upload the database file, which includes the normalized feature extracted from all the training image samples, and the training process set to stop when either the learning rate is achieved or when the epoch becomes larger than 400 steps.

Appendix XI

Procedure for image testing

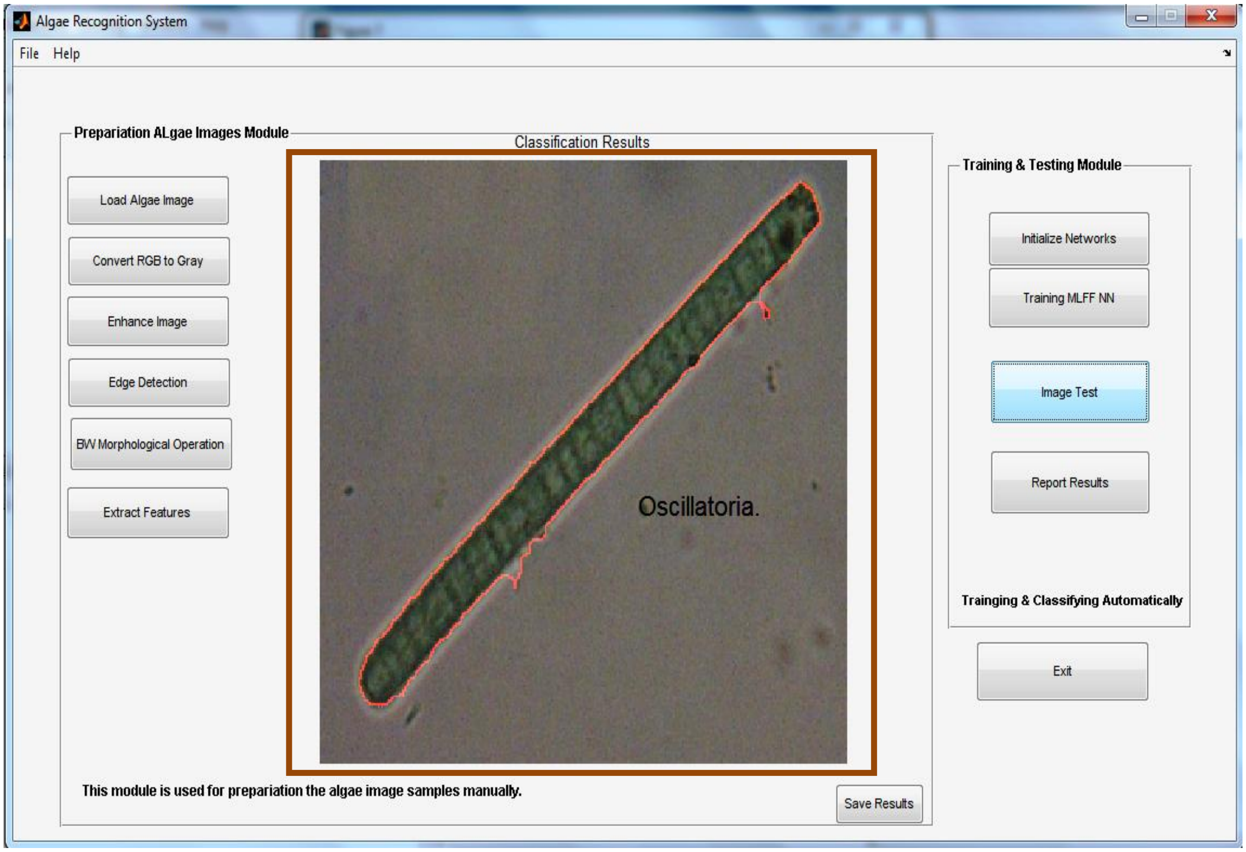


1 Test image

2 Select another algae image to test program result

Appendix XII

System output result display



Results obtained after training with algae classification name

System GUI illustrating classification results (testing mode).

Glossary

Eutrophic: A body of water acquires a high concentration of nutrients and high plant growth.

Mesotrophic: It is between eutrophic and oligotrophic, contain a narrow range of nutrients, principally phosphate and nitrate

Oligotrophic: lake has low nutrient concentrations and low plant growth.

Phytoplankton: are microscopic organisms that live in watery environments, both salty and fresh some phytoplankton are bacteria, some are protists, and most are single-celled plants. Among the common kinds are cyanobacteria, silica-encased diatoms, dinoflagellates, green algae, and chalk-coated coccolithophores

Automated recognition system: is a computer application for automatically identifying or verifying an object from a digital image by comparing selected object features from the image and a object database.

ANNs: is a mathematical model or computational model that is inspired by the biological neural networks. Its consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation.

Feature Extraction: it is a process of determining various attributes and properties associated within a region or objects.

Segmentation: It used to isolate the individual objects in binary image and divided the original image into several sub images based on the number of detected objects.

Matlab: MATLAB is a programming environment for algorithm development, data analysis, visualization, and numerical computation. Using MATLAB, you can solve technical computing problems faster than with traditional programming languages, such as C, C++, and Fortran.

Confusion Matrix:

principle Components Analysis (PCA): is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.

Histogram: maximize the image contrast by applying a gray level transform which tries to flatten the resulting histogram

Freshwater Algae: Algae are widely present in freshwater environments, such as streams, springs, and wetlands

System Module: A portion of a program that carries out a specific function and may be used alone or combined with other modules of the same program

Auto-Alignments: it is a novel function that developed in this study to align each algae with horizontal coordinates before extracted it is features.

Shape Index: it is a novel function developed in this study to make a new feature parameters hold the shape type of algae (e.g. value 1 is assigned when the shape is circular, -1 when shape spiral, etc).