The copyright © of this thesis belongs to its rightful author and/or other copyright owner. Copies can be accessed and downloaded for non-commercial or learning purposes without any charge and permission. The thesis cannot be reproduced or quoted as a whole without the permission from its rightful owner. No alteration or changes in format is allowed without permission from its rightful owner.

UUM
Universiti Utara Malaysia

TEMPORAL - SPATIAL RECOGNIZER FOR MULTI-LABEL DATA

UUM
Universiti Utara Malaysia

ASEEL MOUSA

DOCTOR OF PHILOSOPHY
UNIVERSITI UTARA MALAYSIA
2018

Awang Had Salleh
Graduate School
of Arts And Sciences

Universiti Utara Malaysia

PERAKUAN KERJA TESIS / DISERTASI
(Certification of thesis / dissertation)

Kami, yang bertandatangan, memperakukan bahawa
(We, the undersigned, certify that)

ASEEL MOUSA JAAFAR

calon untuk Ijazah PhD
(candidate for the degree of)

telah mengemukakan tesis / disertasi yang bertajuk:
(has presented his/her thesis / dissertation of the following title):

"TEMPORAL – SPATIAL RECOGNIZER FOR MULTI-LABEL DATA"

seperti yang tercatat di muka surat tajuk dan kulit tesis / disertasi.
(as it appears on the title page and front cover of the thesis / dissertation).

UUM
Universiti Utara Malaysia

Bahawa tesis/disertasi tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan, sebagaimana yang ditunjukkan oleh calon dalam ujian lisan yang diadakan pada : ***14 Januari 2018.***
That the said thesis/dissertation is acceptable in form and content and displays a satisfactory knowledge of the field of study as demonstrated by the candidate through an oral examination held on:
January 14, 2018.

Pengerusi Viva: *(Chairman for VIVA)*	Prof. Dr. Norshuhada Shiratuddin	Tandatangan *(Signature)*
Pemeriksa Luar: *(External Examiner)*	Prof. Dr. Khairuddin Omar	Tandatangan *(Signature)*
Pemeriksa Dalam: *(Internal Examiner)*	Dr. Nooraini Yusoff	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	Assoc. Prof. Dr. Yuhanis Yusof	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	Prof. Dr. Rahmat Budiarto	Tandatangan *(Signature)*

Tarikh:
(Date) January 14, 2018

Permission to Use

I am presenting this dissertation in partial fulfilment of the requirements for a Post Graduate degree from the Universiti Utara Malaysia (UUM). I agree that the Library of this university may make it freely available for inspection. I further agree that permission for copying this dissertation in any manner, in whole or in part, for scholarly purposes may be granted by my supervisor or in his absence, by the Dean of Awang Had Saleh school of Arts and Sciences where I did my dissertation. It is understood that any copying or publication or use of this dissertation or parts of it for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the university Utara Malaysia (UUM) in any scholarly use which may be made of any material in my dissertation.

Request for permission to copy or to make other use of materials in this dissertation in whole or in part should be addressed to:

Dean of Awang Had Salleh Graduate School
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok
Kedah Darul Aman

UUM
Universiti Utara Malaysia

Abstrak

Pengecaman corak merupakan satu tugas perlombongan data yang penting dengan aplikasi praktikal dalam pelbagai bidang seperti perubatan dan pengagihan spesis. Aplikasi tersebut melibatkan pertindihan data yang terkandung di dalam set data berbilang label. Oleh itu, terdapat keperluan bagi algoritma pengecaman yang boleh memisahkan pertindihan data untuk mengenal pasti corak yang betul. Kaedah pengecaman corak sedia ada adalah sensitif terhadap gangguan dan data yang bertindih kerana ia tidak dapat mengenali corak apabila terdapat perubahan pada lokasi data. Kaedah tersebut juga tidak melibatkan maklumat temporal dalam proses pengecaman dan ini membawa kepada kualiti kelompok data yang rendah. Dalam kajian ini, satu kaedah penambahbaikan pengecaman corak berdasarkan Daya Ingatan Temporal Hierarki (HTM) dicadangkan. Algoritma imHTM (HTM yang ditambahbaik) mengandungi penambahbaikan dalam dua komponennya; pengekstrakan fitur dan pengelompokan data. Penambahbaikan yang pertama dikenali sebagai algoritma $_T$S-Layer Neocognitron yang menyelesaikan masalah perubahan lokasi data dalam fasa pengekstrakan fitur. Dalam pada itu, komponen kedua iaitu pengugusan data mempunyai 2 dua penambahbaikan, $_T$FCM dan cFCM ($_T$FCM dengan metrik had-Chebyshev), membolehkan pertindihan data yang terdapat dalam corak dipisahkan dengan tepat ke dalam kelompok data yang berkaitan dengan menggunakan penglompokan temporal. Eksperimen ke atas lima set data telah dijalankan bagi membandingkan prestasi algoritma yang dicadangkan (imHTM) dengan kaedah pengecaman corak berasaskan statistik, templat dan struktur. Keputusan menunjukkan peratusan pengecaman yang berjaya adalah sebanyak 99% berbanding dengan kaedah pengecaman yang lain. Ini menunjukkan bahawa HTM yang ditambahbaik dapat membuat pengecaman corak yang optimum terutamanya bagi corak yang terdapat di dalam set data berbilang label.

Kata kunci: Model Temporal hierarki Daya Ingatan, Neocognitrion, Pengecaman Corak, Data berbilang label, Perlombongan Data

.

UUM Universiti Utara Malaysia

Abstract

Pattern recognition is an important artificial intelligence task with practical applications in many fields such as medical and species distribution. Such application involves overlapping data points which are demonstrated in the multi- label dataset. Hence, there is a need for a recognition algorithm that can separate the overlapping data points in order to recognize the correct pattern. Existing recognition methods suffer from sensitivity to noise and overlapping points as they could not recognize a pattern when there is a shift in the position of the data points. Furthermore, the methods do not implicate temporal information in the process of recognition, which leads to low quality of data clustering. In this study, an improved pattern recognition method based on Hierarchical Temporal Memory (HTM) is proposed to solve the overlapping in data points of multi- label dataset. The imHTM (Improved HTM) method includes improvement in two of its components; feature extraction and data clustering. The first improvement is realized as ${}_{T}$S-Layer Neocognitron algorithm which solves the shift in position problem in feature extraction phase. On the other hand, the data clustering step, has two improvements, ${}_{T}$FCM and cFCM (${}_{T}$FCM with limit-Chebyshev distance metric) that allows the overlapped data points which occur in patterns to be separated correctly into the relevant clusters by temporal clustering. Experiments on five datasets were conducted to compare the proposed method (imHTM) against statistical, template and structural pattern recognition methods. The results showed that the percentage of success in recognition accuracy is 99% as compared with the template matching method (Featured-Based Approach, Area-Based Approach), statistical method (Principal Component Analysis, Linear Discriminant Analysis, Support Vector Machines and Neural Network) and structural method (original HTM). The findings indicate that the improved HTM can give an optimum pattern recognition accuracy, especially the ones in multi- label dataset.

Keywords: Hierarchical Temporal Memory Model, Neocognitron, Pattern Recognition, Multi- label data, Data mining.

UUM
Universiti Utara Malaysia

Acknowledgement

First and Foremost, I would like to express my gratitude to Allah (S.W.T) for helping me to finish my thesis.

With highly appreciation, I would like to express my sincere gratitude to my supervisor Prof. Dr. Yuhanis Yusof for the continuous support of my PhD study and research, for her patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor for my PhD study.

On special note, I would like to thank my beloved family, may Allah save them: my mother, my children and my cousins, especially my cousin Zainab, who was very helpful. Big thanks with love to my husband, who strongly support me mentally and spiritually.

Last but not the least, I would like to thank my dear friends especially Shurooq Mohammed for continuous support throughout my life and study.

UUM
Universiti Utara Malaysia

Table of Contents

UUM Universiti Utara Malaysia

Universiti Utara Malaysia

UUM Universiti Utara Malaysia

UUM
Universiti Utara Malaysia

List of Tables

UUM Universiti Utara Malaysia

Universiti Utara Malaysia

Universiti Utara Malaysia

UUM
Universiti Utara Malaysia

List of Figures

UUM Universiti Utara Malaysia

UUM Universiti Utara Malaysia

Universiti Utara Malaysia

Universiti Utara Malaysia

Glossary of Terms

HTM	Hierarchical Temporal Memory
FCM	Fuzzy c-means clustering
PR	Pattern Recognition
RNN	Recurrent Neural Network
MPF	Memory Prediction Theory
MLT	Medical Laboratory Technician
MLD	Multi- label data
MLL	Multi- label learning
NCBI	National Center for Biotechnology and Information
ASR	Automatic speech recognition
ASIFT	Affine scale invariant feature transform
MRI	Magnetic resonance imaging
OSAD	Optimized Sum of Absolute Difference
SSD	Sum of Square Difference
NCC	Normalized Cross Correlation
SSTN	Sum Square T-distribution Normalized
SAD	Sum of Absolute Difference
PTM	Polyhedral template matching
PCA	Principal Component Analysis
DMSC	Dscriminative multi-scale sparse coding
SVM	Support Vector Machines
LDA	Linear Discriminant Analysis
NN	Group Method of Data Handling
PNN	Probabilistic Neural Networks

Universiti Utara Malaysia

PFCM	Possiblistic fuzzy c-means
CLA	Cortical Learning Algorithm
$_T$S-layer	Time S-layer Neocognitron algorithm
$_T$FCM	Temporal Fuzzy C-Means
imHTM	Improved Hierarchical Temporal Memory
HMM	Hidden Markov Model
$_C$FCM	Chebyshev Fuzzy C-Means

Universiti Utara Malaysia

CHAPTER ONE
INTRODUCTION

1.1 Background

Pattern recognition (PR) is an operation of detecting patterns in data sets and using this data to characterize new data (Sao, Hegadi, & Karmakar, et al.,2014). It is defined as a classification of input data via extraction important features from a lot of noisy data (Paul, Magdon-Ismail, & Drineas, et al.,2016). The identification or interpretation of the pattern in an image can be described effectively with the help of Pattern Recognition (PR) (Kaur & Kaur, et al.,2013). PR is a form of machine learning, which is a field in artificial intelligence (Wu & Toet, et al.,2014). Machine learning in turn is divided into two main groups: supervised and unsupervised learning (Shwartz & Ben-David, et al.,2014). In supervised learning, the computer system is trained by using a classis that are previously defined, and then using this classes to classify unknown objects depending on the patterns that were detected in training (Bova et al., 2016). In an unsupervised learning, the classes are not defined beforehand, and the computer system clusters the data using a group of general rules (Serb, Bill, Ali Khiat, Legenstein, & Prodromakis, et al.,2016). Unsupervised is equivalent to classification known as clustering, which groups the input patterns into clusters depending on the measures of similarity (the distance between input patterns). Other approaches to PR involve semi-supervised learning, which try to find new similarity relationships using previously defined classes to determine new groups (Shwartz & Ben-David, et al.,2014). On the other hand, reinforcement learning is an approach that improve the decisions iteratively, depending on the feedback technique and assigning a reward criterion (Duan, Chen, Houthooft, Schulman, & Abbeel, et al.,2016).

Universiti Utara Malaysia

Recognition of a pattern or image can be done based on statistical information that are extracted from the image or on available knowledge (Dornaika, Traboulsi, & Assoum, et al.,2016a). There are various applications of image recognition such as in face recognition (Yu, Dai, Ren, & Huang, et al.,2017), autonomous robot navigation (Assis, Soares, Coelho, & Van Baalen, et al.,2016), image tracking radar (Kim & Jeon, et al.,2014), intelligent transportation (Bommes, Fazekas, Volkenhoff, & Oeser, et al.,2016), biometrics (Khandelwal, Maheshewari, & Shinde, et al.,2016), license plate recognition (Azad, Azad, & Brojeeni, et al.,2014), fingerprints and character recognition (Gottschlich, et al.,2016).

There are several approaches for pattern recognition under three categories (Gabrielsson, Meibohm, & Weiner, et al.,2016): statistical approaches, which include principal component analysis (PCA), support vector machine (SVM), and linear discriminant analysis (LDA) (Min & Wenming, et al.,2015). Template matching approach fall under two categories of feature based and area based approaches (Larsen, Schmidt, & Schiøtz, et al.,2016). Structural approaches include Hierarchical Temporal Memory (HTM) (Peter Mahler et al., 2016). These pattern recognition approaches differ in performance against noisy and overlapping which could be found in data points. Template matching approach would fail if the patterns are distorted due to the imaging process, viewpoint change, or large intra class variations among the patterns. While for the statistical approach, patterns that are to be recognized should contain structural and relational information or else it is difficult to quantify the features in feature vector form. In structural approaches, on the other hand, the pattern should have clear structure, or it would be difficult to perform segmentation; hence this approach fails with noisy patterns.

Universiti Utara Malaysia

This research focuses on studying the structural approach in the identifying pattern contained in multi-label datasets and comparing among their performance. It investigates the idea of using Hierarchical Temporal Memory (HTM) and Neocognitron learning in recognizing the images and datasets under consideration (especially medical images and datasets).

Medical images are usually indexed with medical records (patient information, acquisition parameters, etc.). Digital medical images have become a key investigatory tool for medical diagnosis and pathology (Knoll et al., 2017). In the medical world, images are usually accompanied by metadata related to the patient, the image acquisition and the radiology department responsible for the acquisition. The complexity of image analysis algorithms is usually rather high and increases with images size (Zambrano, Guzmán-Ramírez, & Pogrebnyak, et al.,2013). Medical images are digitally represented in a multitude of formats based on their modality and the scanning device used. Different operations will display different features, which will cause errors in classifying the correct image (Zhang et al., 2017). Low resolution and strong noise are two common characteristics of medical images (Park; et al., 2015). Due to these characteristics, the visual content of medical images cannot be precisely extracted.

Universiti Utara Malaysia

1.1.1 Hierarchical Temporal Memory

The memory prediction theory has a practical implementation model of HTM, which is a model of a machine learning technology that has the goal of explaining the structural and algorithmic properties of the neocortex (George & Hawkins, et al.,2012). It is a neocortical bio-inspired algorithm of spatial and temporal pooling nodes arranged in a hierarchy for recognition of patterns, prediction of time series and controls of a network (Mnatzaganian, Fokoué, & Kudithipudi, et al.,2016). HTM is a very successful classifier for the case of spatial representation of a pattern when it is present at one point in time, such as image recognition (Kostavelis & Gasteratos, et al.,2012). However, HTM is less useful in cases in which patterns are constructed of a spatial structure evolving over time.

The secondary variation of pattern may cause different spatial representations. This is a problem in pattern recognition in terms of pixel densities. Many efforts by researchers have been made into developing methods that can overcome this problem, for example, tangent distance (Arel, Rose, & Karnowski, et al.,2010), or invariant feature extraction techniques, but only specific problems have been solved successfully (Maltoni & Lomonaco, et al.,2016). HTM exploits time continuity to overcome the problem of two representations (patterns) being related to the same object if they are close in time, even if they are different (Maciej Wielgosza & Pietron´, et al.,2016).

Universiti Utara Malaysia

1.1.2 Neocognitron

Neocognitron is a multi-layer hierarchical neural network proposed by Fukushima (1980). Its architecture was primarily proposed by neurophysiological resulted on the mammal’s visual systems by Hubel & Wiesel in 1962 and 1965. A neocognitron neural network consists of layers of S-cells, which resemble simple cells in the primary visual cortex. These layers are used as feature extraction layers. Layers of C-cells resemble complex cells. These layers of S-cells and C-cells are arranged alternately in a hierarchical manner. It has the ability to recognize the visual patterns robustly through learning. It has been applied in various recognition problems (Fukushima, et al.,2013a) (Fukushima, 2013a). The most recent application of Neocognitron is optical marks recognition (Gorokhovatskyi, et al.,2016), in which researchers used an improved version of Neocognitron neural network in order to improve the recognition speed and accuracy compared to traditional Neocognitron. In addition, Neocognitron has also been applied to handwritten digit recognition (Mariyama, Fukushima, & Matsumoto, et al.,2016) and parallel face recognition (Kunihiko, et al.,2016). A significant increase of performance has been seen on the application of face recognition.

1.1.3 Multi -label data

Multi-label data has more labels than features. The labels in multi-label data can be correlated. Moreover, frequent labels and rare labels can appear together in the same instances. The most basic information that can be obtained from multi- label data is the number of instances, attributes, and labels (Song, Zhu, Scully, & Price, et al.,2013).

Universiti Utara Malaysia

In multi-label learning tasks, however, where there are multiple target variables, it is not clear how stratified sampling could/should be performed (Sechidis, Tsoumakas, & Vlahavas, et al.,2011). In multi-label classification, examples are associated with a set of labels. Multi- label classification is mainly motivated by the tasks of text categorization and medical diagnosis (Tsoumakas & Katakis, et al.,2015). In medical diagnosis, for example, a patient may be suffering from diabetes and prostate cancer at the same time. An example of medical multi-label data is medical images, like blood type images, cancer images, tumor images, and medical datasets like biological data. Nowadays, modern applications increasingly require multi-label classification methods, such as protein function classification (Rehman, Azam, Yao, & Benso, et al.,2017) and semantic scene classification (Aksoy, et al.,2012). Classification errors occur when the classes overlap in the selected feature space, such problems arise in semantic scene and document classification and in medical diagnosis. In medical diagnosis, a disease may belong to multiple categories, and genes may have multiple functions, yielding multiple labels (Fan & Lin, et al.,2007).

1.2 Problem Statement

Today, machine learning techniques are found in object and face recognition in videos and images (Yu et al., et al.,2017), speech recognition (Tokuda & Zen, et al.,2016), and medical images (Wong, Summers, Kebebew, & Yao, et al.,2017). This research determines the problem statement as the limitation of traditional pattern recognition approaches. Traditional approaches of pattern recognition include statistical (e.g. Elementary decision theory and Discriminant functions) (Min & Wenming, et al.,2015), template matching (e.g. correlation and distance measure) (Kour, et al.,2017) and structural (e.g. rule based, and grammar based) (Peter Mahler et al., 2016). These non-hierarchical approaches do not reflect the hierarchical structure of the human brain; hence, they do not reflect the processing that occurs during the recognition operation in the brain. As a result, recognition accuracy and performance will be un accurate due to that nonhierarchical flow of data which causes the loose of data sequence between processing steps (Gabrielsson et al., et al.,2016).

Universiti Utara Malaysia

On the other hand, artificial intelligence approaches such as neural (Pouliakis et al., 2016) and convolutional (Fukushima & Shouno, et al.,2015) networks are based on spatial approaches and do not reflect the process of recognition in human brain (Mnatzaganian et al., 2016), which is spatial and temporal. The brain, referred to as the neocortex, contains virtually human memories, knowledge, skills, and experiences, so further it is responsible for memory prediction (Byrne, et al.,2015b). The neocortex stores sequences of patterns, recalls patterns auto-associatively stores patterns in a hierarchy. All memories are stored in the synaptic connections between neurons (Genzel et al., 2017). The regions of the neocortex that process auditory inputs. The mentioned processes that has been done by neocortex, are the task of pattern recognition in artificial intelligent.

The HTM, in particular, is a technology of structural and algorithmic properties of how the brain can recognize visual pattern, (Fan, Sharad, Sengupta, & Roy, et al.,2016). This indicates that it is suitable to reflect human brain structure. However, there are issues in HTM that need to be addressed, such as the shift in position of input pattern (Wielgosz et al., 2016). Shift in position means that a feature to be extracted in an input pattern is shifted to another location, meaning that the feature is not located at a position where it is supposed to be. This will cause that an input pattern may looks like another pattern because of shift of some pixels from their real position which leads to that the pattern may become another one (like (5) number it may have recognized as (s) letter if there is shift in some pixels). This leads to a different response (Mariyama et al., et al.,2016) in which an unknown pattern is recognized as a known object.

Universiti Utara Malaysia

In addition, existing HTM employs the Agglomerative Hierarchical Clustering (AHC) to cluster the data points, which is sensitive to noise (Mullner, et al.,2011). This problem is common in multi-label data that consists overlapped data points, this means that the pixel or a data point may intrinsically belong to more than one source, in other words, each data point is mapped to a more than one set of labels (Read, Martino, & Hollmén, et al.,2017). Hence, with the above-mentioned limitations, this research intends to improve the HTM by adapting Neocognitron (Jasmin Léveillé, Isao

Hayashi, & Fukushima, et al.,2014) as the learning algorithm and Fuzzy C-means (FCM) (Chen, et al.,2017) as the clustering algorithm.

1.3 Research Question

i. How to adapt the standard HTM in memorizing step to extract accurate features.

ii. How to design clustering approach that improve cluster quality based on temporal information to improve clustering quality of the multi- label data.

iii. How to adapt distance of clustering approach to obtain the optimal clusters

iv. How to evaluate the proposed approach.

1.4 Research Objectives

The main goal of this research is to develop an improved HTM model for pattern recognition. In order to achieve this goal, the following objectives are proposed:

i. To redesign the memorizing step of HTM for feature extraction based on S-layer of Neocognitron neural network with time factor to recognize the input pattern.

ii. To design a clustering algorithm based on FCM by including temporal information of the input data to improve clustering quality of the multi- label data.

iii. To redesign a distance measurement based on the Chebyshev distance to be used in the proposed enhanced FCM clustering algorithm (as in objective II) in order to obtain the optimal clusters.

iv. To evaluate the proposed HTM model by comparing its performance against state of arts methods.

Universiti Utara Malaysia

1.5 Research Scope

The research focuses on classification, clustering and recognizing multi-label datasets which are obtained from Mulan website. This dataset is of two types: medical datasets, blood type and cancer images and Genbase, STARKEY"93. It also includes Yeast dataset. These datasets were chosen because they have overlapped data points and it would be difficult to recognize the pattern.

Furthermore, in this research the focus is on the quality of the recognition process in terms of return in the overlapped data points to their real clusters.

1.6 Research Significance

In this research, an improved algorithm for pattern recognition is designed based on Hierarchical Temporal Memory (HTM). The recognition algorithm consists of an improved algorithm that suits overlapped data. The importance of the improved algorithm is demonstrated in the three objectives described in Section 1. 5. The first objective provides the means of extracting features from input patter. The second objective improves clustering quality using temporal information of the input pattern. The third objective is to provide an optimum cluster by identifying distance measurement to be used with clustering algorithm. The proposed algorithm can be realized in applications that are use multi-label datasets that can be used to recognize patterns, particularly in medical applications, ecological and evolutionary insights, and to predict distributions across landscapes to enhance recognition performance.

Universiti Utara Malaysia

1.7 Organization of the Thesis

This thesis is organized into eight chapters. Chapter One contains an introduction that discuss the pattern recognition methods, Hierarchical temporal memory and

Neocognitron neural network. This chapter also includes the problem statement, research objectives, research scope and research significance.

The second chapter presents the proposed category of pattern recognition methods. Existing literature on different categories of pattern recognition techniques is discussed as well as the significance of Hierarchical Temporal Memory in recognizing multi- label data.

Chapter Three clarifies the methodology used in execution this research. It includes the structure of the proposed Hierarchical Temporal Memory for multi- label data, which includes the components of the proposed method, data collection, feature extraction, hidden Markov model and temporal grouping using clustering process.

Chapter Four presents the realization of proposed feature extraction step using time S-layer Neocognitron neural network. The Neocognitron neural network is tested on the benchmarks datasets used in this research.

Universiti Utara Malaysia

Chapter Five presents the realization of proposed clustering algorithm using temporal Fuzzy c-Means algorithm. The temporal Fuzzy c-Means algorithm uses temporal information in clustering overlapping data. It is tested on the benchmarks datasets used in this research.

Chapter Six presents the distance metric that are used with proposed clustering algorithm presented in Chapter Five. The distance metric is used to raise the accuracy of clustering overlapped data points.

Later, Chapter Seven contains an elaboration on the proposed Hierarchical Temporal Memory is presented. A combination of the algorithms presented in Chapters Four, Five and Six is presented.

Finally, Chapter Eight gives concluding remarks on the proposed Hierarchical Temporal Memory. This includes the research contribution and recommendations for future research work and limitations relating to Hierarchical Temporal Memory.

1.8 Summary

As a summary, improving the performance of HTM model in recognition is important to both researchers and industry. Traditional approaches of pattern recognition do not reflect the process that occurs in the human brain when recognizing an image. Such a drawback is addressed by proposing the utilization of HTM and Neocognitron learning. It is hoped that the findings of the proposed research will contribute to theoretical modelling by modifying HTM learning in relation to medical image and multi-label datasets recognition.

Universiti Utara Malaysia

CHAPTER TWO
LITERATURE REVIEW

This chapter is aimed at situating the current research in the context of existing knowledge in image recognition and multi- label data recognition of HTM model, discovering the gap and providing rationale for the current study. The research focuses on recognizing medical images and multi- label data using HTM model and Neocognitron neural network as learning algorithm.

Image processing is the process of taking an image as an input and apply some processing actions to produce an enhanced output image for using it for further applications (Liu, Lu, Feng, & Zhou, et al.,2017). Figure 2.1 illustrate the image processing flow.

Universiti Utara Malaysia

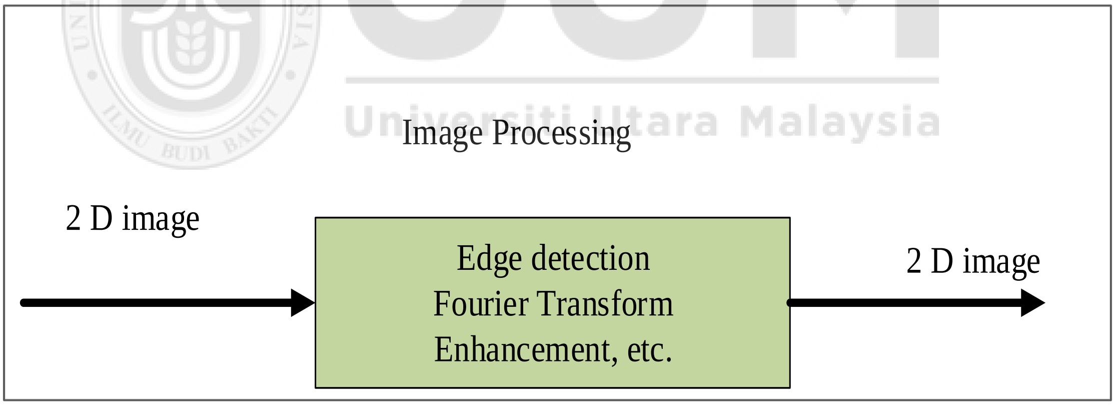


Figure 2.1. Image processing operation

Resource. (Zhou, et al.,2000)

The purpose of image processing could be divided into five groups   that includes (Boyat & Joshi, et al.,2015):

1. Visualization: observe the non-visible objects.

2. Image sharpening and restoration: create an enhancement image.

3. Image retrieval: search for interesting images.

4. Measurement of pattern: Measures the objects in an image.

5. Image Recognition: recognize the objects in an image.

This research focuses in recognizing medical images using HTM model and Neocognitron neural network as learning algorithm.

2.1 Pattern Recognition

First it is important to understand the biological realistic neocortex in recognizing patterns. The neocortex is in humans and other mammals, the center of intelligence. Physically, it is like a sheet. This sheet has thickness of several millimeters and has several layers of neurons. The organization of motor and sensory processing in the neocortex is topographical (Mokeev & Mokeev, et al.,2015). In general, it means that the nearby neurons represent related things. For example, the neurons of the cortex V1 that occurs in the visual areas are arranged in topographical manner and the neighboring neurons, on the retina, have neighbor receptive fields. The different regions are corresponding to different stages of information processing and different modalities, can be known on the sheet of neocortical (Bin Wang, Hiroki Yamamoto, Wu, & Ejima, et al.,2015).

Universiti Utara Malaysia

Decoding the cortical circuit's functional connectivity is difficult job and is correlating with the danger implicated in the complex system reverse engineering. Several things can be gained by comparing reverse engineering of an electronic circuit with it (Sao et al., et al.,2014). Although a single transistor can be used as an amplifier, for real world use, it is rarely build using a single transistor. A good structure includes a biasing circuitry, which makes sure that the transistor used for amplifying works properly even with changing in different characteristics and temperature conditions of device and feedback instabilities. If this situation is compared with the cortical sheet, it is found that a similar thing exists, where a multitude of neurons are implicated in biasing a canonical cortical circuit to function (Shahrour, et al.,2014).

When the circuit is not properly biased and it is tested for connectivity (for example, when the behaving condition is not in the normal of an animal), it could, at the end, some important connections are missing and some spurious ones are logging (Meyer & Fingscheidt, et al.,2016).

The biological circuit explains the mapping from the belief propagation computations intermediate values to column and lamina specific neural responses. In the same network, the neurons in V1 responding to misleading contours, and the event of higher-level recognition which decreasing the activity of lower levels. Modeling of these events can serve as examples of hypothesis driven physiological experiments can be derived circuits.

Universiti Utara Malaysia

Cortical areas have a laminar organization, which means that many different layers of neurons construct the neo-cortical sheet. Neurons in a given layer of the neo-cortical sheet fit to a regular pattern at the point of receiving connection and end of connection. In general, the neo-cortical sheet has six different laminae (layers). In spite of finding several exceptions, the six layers provide acceptability upon framework with which to explore models of cortical circuits (Min & Wenming, et al.,2015).

These layers are numbered, depending on its closest to or far from skull and white matter. Layer1 is closest to the skull but farthest from the white matter. Typically, the layers are horizontally described, with layer1 at the top and layer-6 at the bottom. With this description, layers 1-3 are identified as `superficial' layers or top layers. Layers 5 and 6 are the deep layers of cortex (Zytoon, et al.,2014).

From the previous discussion, it is clear that the regions in the visual cortex are connects in a hierarchical manner (Byrne, et al.,2015b). For example, the primary visual cortex, which is area V1, is the first cortical processing stage. The output of V1 is sent to cortical area V2 and then to cortical area V4. Physically, the different areas connections go through cortical sheet itself and through the white matter. It is recognized also that there is mutual connection between different regions of cortex (Karczmarek, Kiersztyn, Pedrycz, & Dolecki, et al.,2017).

Universiti Utara Malaysia

After explaining the structure of the brain and how it works when recognizing patterns, it is important to define how to recognize patterns in artificial brain. Alexander Arimond (Arimond, et al.,2010) defines the use of the computer algorithms to discover regularities in data. These regularities are used to construct models to represent real world circumstances. Based on these models, which is approximations of the real world, possible actions are the making predictions depending on data and classification of data into different categories.

The primary goal of pattern recognition is the classification, which in turns could be supervised or unsupervised (Wong, et al.,2017). Its classification or recognition consist of one of two tasks: supervised classification, example discriminant analysis.

Here the input pattern is identified as a member of a class that previously defined. In contrast, unsupervised classification, the pattern is related to a undefined class, example of this classification is clustering (Eyupoglu, et al.,2016). There are various application areas of pattern recognition such as document classification (Xia, Chen, Luan, & Song, et al.,2017) , image analysis (Tu et al., et al.,2017) , data mining (Ryu, Kim, Choi, Yu, & Lee, et al.,2017) , industrial automation (Amza & Cicic, et al.,2015) , biometric recognition (Wu & Deng, et al.,2015) , remote sensing (Zhu & Ma, et al.,2016) , handwritten text analysis (Bal & Saha, et al.,2016), GIS (Bin Wang et al., et al.,2015) , medical diagnosis (Zytoon, et al.,2014) , speech recognition (Meyer & Fingscheidt, et al.,2016) , and many more .

Universiti Utara Malaysia

There are six processes in pattern recognition (Dornaika, Traboulsi, & Assoum, et al.,2016b) . The first step is data acquisition. Which is the process of converting data from one form like speech, pictures, character into a form acceptable to the computing device pixels or binary for further processing (Manivannan et al., et al.,2016) . In general, the data acquisition is performed by several devises, such as digitizing machine, sensors and scanners. In this step, a functional mapping is done between the correct class labels and the feature vectors (Kravanja, Žganec, Žganec-Gros, Dobrišek, & Štruc, et al.,2016) . All the variations that are likely to be seen by the system , which are a field of data instances, are sampled by the training and test data (Kumar, et al.,2011).

Second step is data analysis. When the first step is finished, the task of analysis started. During this step the data learning is begin, then collect information about the pattern classes available in the data (Gabrielsson et al., et al.,2016) . Further processing is done on this information.

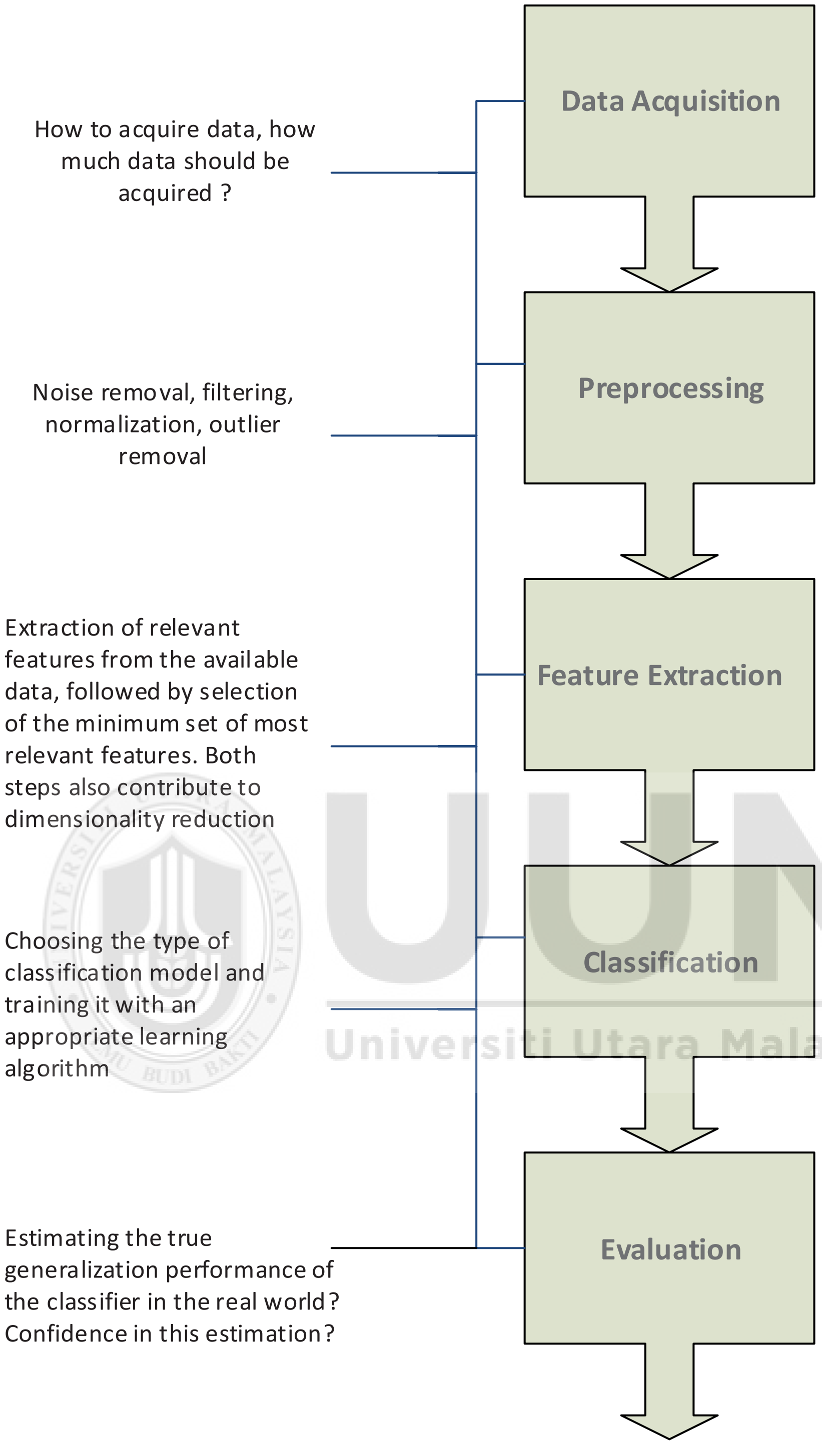
Third step in pattern recognition is the preprocessing. In this step removing of noise from various sources to the range possible is done (Hou, Feng, & Wang, et al.,2016).

Fourth step is the feature extraction. This step contains the discriminative and invariant features considering the rotation, translation and scale. This can be obtained by applying a mathematical transformation on data (Nai, Liu, Rempel, & Wang, et al.,2017) .

The fifth step is classification. Here an object is assigned to a category by using the feature vector that is extracted from the feature extraction step. There are several classification techniques, such as bayes classifier, naive bayes Classifiers, k-Nearest Neighbor (KNN) Classifier, and Neural Networks (Wu & Toet, et al.,2014).

Universiti Utara Malaysia

Lastly, is the sixth step which is evaluation and this includes estimating the generalization performance of the classifier in the real world (Sao et al., et al.,2014). Figure 2.2 shows the mentioned steps of pattern recognition in illustrative manner.


Data Acquisition
How to acquire data, how much data should be acquired ?
Preprocessing
Noise removal, filtering, normalization, outlier removal
Feature Extraction
Extraction of relevant features from the available data, followed by selection of the minimum set of most relevant features. Both steps also contribute to dimensionality reduction
Classification
Choosing the type of classification model and training it with an appropriate learning algorithm
Evaluation
Estimating the true generalization performance of the classifier in the real world? Confidence in this estimation?

UUM Universiti Utara Malaysia

Figure 2. 2. Pattern recognition process

Resource. (Arimond, et al.,2010)

Pattern recognition system is applicable in several areas like:

2.1.1 Speech Recognition

Is the decoding of the human voice into digitized speech which can be interpreted by the computer, computer software program or hardware device (Doulaty, Rose, & Siohan, et al.,2016). It is used to perform commands, operate a device or write without using a keyboard, press any buttons or using a mouse. This can be done on a computer today by using a software programs called automatic speech recognition (Badyalina, Shabri, & Samsudin, et). Christos Dimitrakakis and Samy Bengio focuses on how and whether can they use boosting and bagging for speech recognition. They use the data based on a previously segmented version of the OGI Numbers 95 (N95) data set. They compare their work to a single hidden markov method HMMs. The result shows that these methods, that are pass in favor of boosting, must be checked more closely as speech recognition useful learning technique (Tokuda & Zen, et al.,2016).

Universiti Utara Malaysia

2.1.2 Object Recognition

Which is an important function in computer vision and image processing It is the process that make the observers able to recognize three-dimensional objects even when it receives two-dimensional input that is greatly varies with respect to viewing conditions (Xie, Yang, Peng, Gao, & Tan, et al.,2016) . Reza Oji recognize some objects by joining two algorithms, a region merging algorithm affine and scale invariant feature transform (ASIFT). He is training the object in several image sides with different aspects to the fined best key points of it. Then, using a region merging algorithm to recognize and detect the full boundary object in the other images based on ASIFT key points and a measure of similarity for merging regions (Wielgosz & Pietro´n, et al.,2016) . Nikhil Gupta, Rahul Gupta, Amardeep Singh and Matt Wytock make a template matching for accurate constructing of 3D object. The research shows that if using more poses, this will improve the accuracy of the system. (Lopez-Perez, Ayala-Ramirez, & Hernandez-Belmonte, et al.,2016). Marek Bundzel and Shuji Hashimoto identify objects in dynamic images. They produce a system that enabling a mobile robot to recognize different types of objects and explore its environment without human supervision and elimination of uncertainty that linked with the variable speed of the robot. They do that by using HTM model structure with incremental learning (Pouliakis et al., et al.,2016).

2.1.3 Medical Image Processing

Which becomes one of the most important interpretation methods in biology and medicine over the past decade (Wong et al., et al.,2017). Nowadays a massive development of new and powerful tools for detecting, transmitting, storing, analyzing, and displaying medical images is happened (Zhang et al., et al.,2017). This massive development needs to improve and develop a huge number of application of digital processing techniques for solving medical problems (Tan et al., et al.,2017). Because of the poor quality of some images, the designing and implementation of complex medical systems needs for interconnection between physicians and engineers, otherwise it will leads to inaccurate in feature extraction, analysis, and recognition of medical application, which then causes an inaccurate diagnosis which cause sometimes to death (Dobrescu, Ichim, & Crişan, et al.,2013). Therefore, many researches have been done today towards improvement of image recognition (Zytoon, et al.,2014). Medical imaging contains:

UUM Universiti Utara Malaysia

Generalities: The discovery of X-rays and pioneered medical imaging by roentgen was an interesting matter in Europe and the United States that time (Zhang et al., et al.,2017) . The commercial projects and innumerable medical applications investigated and organize public demonstrations within a few months. The field of radiography was born with a bang (Soliman et al., et al.,2017) .

Imaging Modalities: Improvements have been done for imaging technology since the end of the 19th century. Developing of many imaging techniques which are used in clinical analysis take place, such as magnetic resonance imaging (MRI), computed tomography (Lopez-Perez et al., et), Doppler ultrasound, and various imaging techniques based on nuclear emission (PET (positron emission tomography)), SPECT (single photon emission computed tomography), and so on (Adhikari, Sing, Basu, & Nasipuri, et al.,2015). These techniques are not so far suited to specific organ or pathology, because they are based on different physical principles (Serb et al., et al.,2016). In practice, they couldn't be an original tool for medical analysis because they are complementary, that thy gives different results for sometimes same case. In medical imaging, these different imaging techniques are called modalities (Adjouadi, Zong, & Ayala, et al.,2005).

Universiti Utara Malaysia

In most developed countries blood cancer is the leading cause of death among all cancers developed (Dobrescu et al., et al.,2013). Any diagnosis tool helping to improve the sensitivity, or the specificity of blood cancer would be highly valued. Recognizing blood type accurately also important for Saving human life if someone need blood for some reason like bleeding too much or disease that need to exchange blood (Uga, Dan, Sano, Dan, & Watanabe, et al.,2014) . There are some difficulties and drawbacks of manual methods of medical image recognition that appears when acquire to use computational methods. Some of these difficulties are:

Time consuming: visual inspection of microscopic images taking time and are exhaustive. If there is an interrupt in counting process, the Medical Laboratory Technician (MLT) has to start again from the scratch (Dobrescu et al., et al.,2013).

Cell analysis is done by comparing the microscopic image with cell type images that they are familiar with. This process is realized by an experienced MLT. A non-professional MLT need to make many checking attempts with medical literature to realize the cell types to determine the count of a given sample. Thus these manual methods are liable to human exhaustions that can result in errors (Coutinho & Inoue, et al.,2016) .

After the analyzing process, the blood cell slides are put away. So the operation of retrieving the analyzed image for future reference becomes too slow compared with computerized system (Pablo, Menichini, Larese, & Riquelme, et al.,2015) .

Universiti Utara Malaysia

From the above discussion it can found that the expert physicians should depends on their own medical knowledge and experience companied with the software of image processing to manually describe and evaluate the images (Park; et al., et al.,2015)

In the past decade, medical imaging passes through a revolution in terms of faster, more accurate diagnosis. This accelerated improvement needs for software that in turn need for more development processing algorithms in the area of image and signal processing (Soliman et al., et al.,2017) .

Today, medical imaging is important and necessary in medical application. Pathologies can be seen and concluded directly from symptoms (Mussarat Yasmin, Sharif, & Mohsin, et al.,2013). Medical imaging techniques can be used also in performing surgery (Adhikari et al., et al.,2015) .

2.2 Approaches in Pattern Recognition

2.2.1 Template Matching

It is one of the simplest pattern recognition approaches. Patterns are represented by samples, models, pixels, curves, textures (Mahalakshmi, Muthaiah, & Swaminathan, et al.,2012). It is one of the most common techniques that are used to extract useful information from an image sequence (Jurie & Dhome, et al.,2002). It is a technique used to compare portions of images with another image in order to classify an object (Mahalakshmi et al., et al.,2012). The choice of matching based on the problem that should be solved and the nature of the image (Mahalakshmi et al., et al.,2012). There are two categories of template matching:

Universiti Utara Malaysia

2.2.1.1 Featured-Based Approach

Are applied, when the structural information are used instead of intensity information. The aim of this approach is to allocate the pair wise connection between template and reference pattern by using their spatial relations or features (Abdulfattah & Ahmad, et al.,2013). T. Mahalakshmi, R. Muthaiah and P. Swaminathan propose a method using cross correlation on feature- based approach. In this technique two dimensional FFTs (one direct and one inverse) and a complex multiplication are applied. The proposed method reduces the correlation computation by more than two orders of magnitudes with keeping the degree of accuracy. It has the limitation that it could be just used in real time applications, still have a problem when shifts and rotation is too large (Mahalakshmi et al., et al.,2012)

2.2.1.2 Area-Based Approach

Sometimes called correlation-like methods or template matching. Area based method is suitable for the templates with no strong features in image, since they directly operate on the block of values (Mahalakshmi et al., et al.,2012). An improved Template Matching algorithm is proposed to recognize license plate characters. Plate recognition system has two main parts: Segmentation of license plate characters and Recognition of characters from the plate. The result shows that the accuracy obtained is comparable to the best figures reported in literature. But it has the limitation of time (Chakraborty & Parekh, et al.,2015).

Universiti Utara Malaysia

Y. M. Fouda propose a 1-D template matching algorithm which is an alternative for 2-D full search block matching algorithms. It is robust to detect the target object with changes of illumination in the template also when the Gaussian noise added to the source image. It has huge execution time when the size of image is big (Fouda, et al.,2015).

M. Juvela, implement of a TM algorithm for map analysis. It provides a fast and still relatively robust way to identify elongated structures or other image features. The performance of the TM method is found to be comparable to that of RHT, but TM appears to be more robust regarding the input parameters. It implemented to Herschel surface brightness data (Juvela1, et al.,2016). Ghassan marwan abdulfattah and mohammad nazir ahmad improve the previous method by developing similarity measurements they use Optimized Sum of Absolute Difference (OSAD), Sum of

Square Difference (SSD) and Normalized Cross Correlation (NCC) for measuring the correlation between the template image and the input image. Also use two metrics (Sum Square T-distribution Normalized (SSTN) and Chi-Square distribution (Chi2)) in order to measure the matching between the two images (Abdulfattah & Ahmad, et al.,2013). The results show that the accuracy is increased from 98% and 95% for SAD and SSD respectively up to 100% on the Yale dataset and from 90% and 87% up to 100% on the MIT-CBCL dataset. On the other hand, the other metrics gives high localization results against only on expressions and poses and low results against small background changes and illumination. But it needs developing against clutter background. Peter Mahler Larsen, Søren Schmidt and Jakob Schiøtz, propose a new method, polyhedral template matching (PTM), which classifies structures according to the topology of the local atomic environment, without any ambiguity in the classification, and with greater reliability. Results shows that the reliably be used to identify structures even in simulations near the melting point (Peter Mahler, Søren, & Jakob, et al.,2016). Amandeep Kour, propose, a wavelet transformation for preprocessing the face image, extracting edge image, extracting features and finally matching extracted facial features for face recognition. Results shows increase the robustness of the existing face recognition algorithms (Kour, et al.,2017). Table 2.1 presents the summary of existing work in template matching approach.

Universiti Utara Malaysia

Table 2.1

Summary of existing researches in template matching.

Author	Year	Category	Technique	Advantage	Disadvantage	Domain
Swaminathan	2012	Feature based	FFTs and complex multiplication	Reduce the correlation computation	Shifts and rotation is too large	Real time
Ghassan Marwan	2013	Area based	(OSAD) (SSD) (NCC)	Increase the accuracy	Not good in small change in background and illumination	Face recognition
hakraborty & Parekh	2015	Area based	Segmentation & recognition	God accuracy	Time limitatiom	Image recognition
Fouda	2015	Area based	1-D template as 2-D.	robust to changes in illumination and noise	More execution time for big size	Face recognition
Juvela	2016	Area based	TM algorithm for map analysis	fast and robust	Low accuracy	Herschel surface brightness
Peter Mahler Larsen, Soren Schmidt and Jakob Schiotz	2016	Area based	classifies structures according to the topology of the local atomic environment	Reliable in identify structures near melting point	Not good with complex structures	Image recognition
Amandeep Kour	2017	Area based	wavelet transformation	increase the robustness of the existing face recognition algorithms	Cannot handle noisy case	Face recognition

Universiti Utara Malaysia

2.2.2 Statistical

In statistical technique patterns are represented as features. The discriminant function is the recognition function in this technique (Tabacchi et al., et al.,2013). The main goal of this approach is to select the features that allow the vectors of pattern relates to different categories to take up disjoint and compact regions in a d-dimensional feature space (Dornaika et al., et al.,2016a) . It is the most natural and general framework to formulate solutions to pattern recognition problems, which recognizes the probabilistic nature of both the information and the form that express the results (Bishop, et al.,1995). It has the ability to separate the patterns of different classes is the pointer of effectiveness of the representation space (feature set) (Tabacchi et al., et al.,2013).

Among the most utilized statistical methods in image recognition are:

2.2.2.1 Principal Component Analysis (PCA)

Is one of the most successful techniques that have been used in image recognition and compression (Huan, Caramanis, & Sanghavi, et al.,2012). It uses an orthogonal transformation as a mathematical operation to convert a set of correlated patterns into a set of uncorrelated patterns or values (Reddy & Raju, et al.,2016) . It can be considered as a statistical method from the consideration of the factor analysis concept. It is the most widely used subspace projection technique for face recognition (Khandelwal et al., et al.,2016). PCA is suitable in the linear domain, for the reason of that it is considered as a classical technique, it is used for applications with linear models, such as signal processing, system and control theory, 2D-image processing, communications and so on (Li & Khashanah, et al.,2015) .

Universiti Utara Malaysia

K. Sai Prasad Reddy and Dr. K. Nagabhushan Raju use the Eigen faces system based on Principal Component analysis (PCA) to recognize facees. A small set of significant features are used to describe the variation between face images. Results shows that the recognition rate increases with the number of training images per person, if the minimum distance between the test image and other images is zero, the test image entirely matches the image from the training base otherwise it is an unknown person (Reddy & Raju, et al.,2016). Muzameel Ahmed and V.N. Manjunath Aradhya uses PCA for 2D shape recognition using shape features extraction. The work has two steps, in the first step the Ridgelet transform is used to feature extraction, in the second step PCA is used for dimensionality reduction. The proposed method is tested on the standard MPEG-7 dataset. The results shows the efficiency of the proposed method in recognition of 2-D shapes (Ahmed & Aradhya, et al.,2016).

Universiti Utara Malaysia

Fujiao Ju, Yanfeng Sun, Junbin Gao, Simeng Liu, Yongli Hu, Baocai Yin proposes a mixture of bilateral-projection probabilistic principal component analysis model (mixB2DPPCA) on 2D data. A Bayesian inference scheme has been proposed based on the variation of EM (Expectation-Maximization) approach for learning model parameters. Results show that the performance of mixB2DPPCA has been largely improved, resulting in more accurate reconstruction errors and recognition rates than the existing PCA-based algorithms (Ju et al., et al.,2016).

In (Murali, et al.,2015), PCA is used for face recognition for feature identification in large data sets and to highlight their similarities and differences is more essential step in face recognition. It uses Yale database B, he proves that the proposed method reduces the dimensions of process data.

Yu-Feng Yua, Dao-Qing Daia, Chuan-Xian Rena, Ke-Kun Huangb, proposes a discriminative multi-scale sparse coding (DMSC) model to address this problem. They use PCA in learning intra-class variant dictionary, the method applied on benchmark databases and the results shows that the proposed method is more robust and has higher breakdown point in dealing with the SSPP problem for face recognition (Yu et al., et al.,2017)

2.2.2.2 Support Vector Machines (SVM)

It is originally a binary classification method developed by Vapnik and colleagues at Bell laboratories (Das, Pal, Ballester, & Blumenstein, et al.,2014). The recognition function is usually a correlation or distance measure (Marqu´es, et al.,2010).

Guangyi Chen, Tien. D. Bui and Adam Krzyżak introduce another term to the objective function of the standard SVM to achieve a sparse representation. The proposed method realize higher classification rates because of sparser support vectors and is more robust to outliers in the datasets (Chen, Bui, & Krzyżak, et al.,2016) . Min Zhang and Wenming Cheng introduces an intelligent hybrid model for recognizing the mixture CCPs that includes three main aspects: feature extraction, classifier, and parameters optimization. In the feature extraction, statistical and shape features of observation data are used in the data input to get the effective data for the classifier. Results demonstrate that the proposed approach is able to effectively recognize mixture CCPs (Min & Wenming, et al.,2015).

Shahrour, use Multi-class Support Vector Machine (SVM) Advanced Pattern Recognizer in presence of leakages in various sizes and locations to detect and classify leakages. It improve the day to day operational decision making process by detecting and classifying the different stages of pipelines leakages and breakages according to their severity (Shahrour, et al.,2014) .

Saurabh Paul, Malik Magdon-Ismail and Petros Drineas gives two provably accurate feature-selection techniques for the linear SVM. The algorithms run in deterministic and randomized time respectively. They demonstrate that this method is competitive and often better than prior state-of-the-art, for which there are no known provable guarantees. Algorithms can be used in supervised or unsupervised setting. They also prove margin is preserved to within ε-relative error in the full feature space (Paul et al., et al.,2016).

Vasileios Mygdalis, Alexandros Iosifidis, Anastasios Tefas, Ioannis Pitas Introducing the Graph-Embedded One-Class SVM and Graph Embedded SVDD methods. They are exploiting data relationships in the optimization process, Adopting generic geometric class information acts as a regularizer. They prove that the regularized solutions for both One-Class Support Vector Machine and Support Vector Data Description are equivalent to applying the original methods in a transformed (and shared) feature space, adopting generic geometric class information acts as a regularizer (Mygdalis, Iosifidis, Tefas, & Pitas, et al.,2016). Zhi He, Lin Liu, Suhong Zhou and Yi Shen Proposes a GSLR method to learn structured dictionary for HSI. They Apply fast super-pixel segmentation method to gain spatial groups and add group-based sparse and low-rank regularizations for dictionary learning then update representation matrix by IALM and dictionary by BCD at the end Classify representation matrices of test samples by linear SVM (He, Liu, Zhou, & Shen, et al.,2016)

Universiti Utara Malaysia

2.2.2.3 Linear Discriminant Analysis (LDA)

Is a method that seeks to keep the discriminatory of classes information, while reduce its dimensionality. This approach is also capable of distinguishing the variation of image that are identical from variation related to other sources such as expression and illumination (Flores, Linguraru, & Okada, et al.,2010). Telgaonkar Archana H and Deshmukh Sachin uses LDA to reduce the dimension on high dimensional data set UMIST, COIL and YALE which consists of images of objects and human faces. Classify the objects using KNN classifier and naive Bayes classifier to compare performance of these techniques. Difference between supervised and unsupervised learning is also inferred using these results. Results shows that semi supervised learning is better in big dimension images (H & Sachin, et al.,2015). A. V. Mokeev and V. V. Mokeev found that in processing large sets of images, this will reduce the complexity of computation of principal components, this is proposed to use the linear condensation method and principal component synthesis (Mokeev & Mokeev, et al.,2015). Aytug Onan, Hasan Bulut and Serdar Korukoglu uses an improved ant clustering algorithm, where two novel heuristic methods are proposed to enhance the clustering quality of ant-based clustering. In addition, the latent Dirichlet allocation (LDA) is used to represent textual documents in a compact and efficient way. The clustering quality of the proposed ant clustering algorithm is compared to the conventional clustering algorithms using 25 text benchmarks in terms of F-measure values. The experimental results indicate that the proposed clustering scheme outperforms the compared conventional and met heuristic clustering methods for textual documents (Aytug Onan, Bulut, & Korukoglu, et al.,2016). Radu Dobrescu and his partner use fractal dimension and lacunarity to distinguish among benign and malign cases of mammary tumors. They try to discriminate among benign and

Universiti Utara Malaysia

malignant tumors in order to develop new tools investigation in cancer diagnosis. The classification accuracy improved. They prove that the Combinations of shape factors and texture measures would certainly be more effective in the classification of breast cancer than any type of features on its own. But extend the database with medical images gives more relevant statistics they use test/training error and Leave K Out Cross Validation (LKOCV) as performance metric (Dobrescu et al., et al.,2013).

2.2.2.4 Neural Network (NN)

Can be defined as parallel computing systems that it is constructing of many interconnection among huge number of simple processors (Jiang, Trundle, & Ren, et al.,2010). Neural networks considered as an extension of conventional techniques in statistical pattern recognition (Bishop, et al.,1995). Neural network models use some organizational principles, such as adaptively fault tolerance, learning generalization, and distributed representation, in a network consists of nodes, which are artificial neurons, connected with each other's in a weighted directed graph by directed edges (with weights). In last decade, the employment of neural networks in a medical imaging applications opening a wide spaces for researchers (Mussarat Yasmin et al., et al.,2013). The most effective and popular proposals focusing neural networks are:

1. Group Method of Data Handling (GMDH) based algorithm: Is the group of methods for data handling proposed by Kondo and Kondo (Badyalina et al., et al.,2014) for identification of medical images. GMDH use feedback loop methodology to detect the properties of patterns more accurately. For 3D modeling of the heart, GMDH algorithm uses neurons based function for organization of selectable neural network. Medical images for heart are analyzed and processed using revised GMDH approach. The selection of

Universiti Utara Malaysia

neural network is automatically done by the algorithm. Image densities of neighboring regions have better statistics for extraction of image features. In this approach only, the image features which can be useful are selected. Results of testing showed that the revised GMDH neural network technique accurate methods for image recognition. Especially for heart image recognition revised GMDH outperformed other conventional methods of image recognition and analysis. Another approach of identification is based on neural network which is using the feedback GMDH. This approach discusses about the minimization of prediction error criterion. The intentional application of this algorithm is to be applied over images of brain. It is observed that this method of image recognition works perfectly for image analysis of brain in all dimensions. Another approach cancer detection using GMDH extract the regions of lungs where possibility of cancer occurs. This algorithm works on multi-detector row images and two phases approach is followed. In first phase the lungs themselves are detected and in second phase areas where cancer is detected are extracted (Kondo, et al.,2011). Vaibhav Garg investigates the suitability of an inductive group method of data handling polynomial neural network (GMDH-NN) technique in estimating sediment yield. The data on various meteorological and geomorphological features—namely, river length, watershed area, erodible area, average slope of watershed, annual average rainfall, and drainage density—from 20 sub watersheds of the Arno River Basin in Italy were used for model development. The results of this study show that the inductive GMDH-NN can efficiently capture the trend of sediment yield with a coefficient of correlation of 0.975 (Garg, et al.,2014).

Universiti Utara Malaysia

2. Fuzzy neural networks: Information of medical image processing usually deals with huge set of information this information is usually passed through the system but not brought in use of simplifying and updating the process of image processing. A method of image classification using soft information fusion is proposed in Hariton Costin and Cristian Rotariu (Zhang & Shen, et al.,2014). Sudip Kumar Adhikari, Jamuna Kanta Sing and Dipak Kumar Basu presents a spatial fuzzy C-means (SpFCM) algorithm for the segmentation of MRI images. The algorithm utilizes spatial information from the neighborhood of each pixel under consideration and is realized by defining a probability function. A new membership function is introduced using this spatial information to generate local membership values for each pixel. Finally, new clustering centers and weighted joint membership functions are presented based on the local and global membership functions. The resulting SpFCM algorithm solves the problem of sensitivity to noise and intensity inhomogeneity in MRI data and thereby improves the segmentation results (Adhikari et al., et al.,2015). Esmaeil Mehdizadeh, Amir Golabzaei and Kun Chen propose a new heuristic fuzzy clustering algorithm based on electrical rules. The laws of attraction and repulsion of electric charges in an electric field are conducted the same as the target of clustering. Use the electrical rules in electric fields and Coulomb's law to obtain the better and the realest partitioning, having respect to the maximum separation of clusters and the maximum compactness within clusters. Computational results show that our proposed algorithm in comparison with FCM algorithm as a well-known fuzzy clustering algorithm have good performance (Mehdizadeh, Golabzaei, & Chen, et al.,2016). In (Chen, et al.,2017), the authors develops an ICKFCM

Universiti Utara Malaysia

method based on kernelized fuzzy -means clustering with ICA analysis for extracting regions of interest in MRI brain images. The proposed method first removes the skull region using a skull stripping algorithm. Through ICA, three independent components are then extracted from multimodal medical images containing T1-weighted, T2-weighted, and PD-weighted MRI images. As MRI signals can be regarded as a combination of the signals from brain matters, ICA can be used for contrast enhancement of MRI images. Finally, the three independent components are utilized as inputs by KFCM algorithm to extract different brain tissues. Results show that the proposed method is capable of accurately extracting the complicated shapes of brain tissues and remaining robust against various types of noises.

UUM Universiti Utara Malaysia

3. Probabilistic neural networks: PNN is supervised learning algorithm, with not weighted hidden layers. Instead the action on the hidden node is going to be the weights of that node, these are not adjusted at all. Alexander Serb, Johannes Bill, Ali Khiat, Radu Berdan, Robert Legenstein and Themis Prodromakis presents an approach that can be exploited for processing unlabelled data and can adapt to time-varying clusters that underlie incoming data by supporting the capability of reversible unsupervised learning. The potential of this work is showcased through the demonstration of successful learning in the presence of corrupted input data and probabilistic neurons, thus paving the way towards robust big-data processors (Serb et al., et al.,2016)

   Many algorithms in probabilistic domain have been proposed to handle the issue of creating proper shape of the organ. One of these algorithms is proposed by Yuyin Zhou, Lingxi Xie, Wei Shen, Elliot Fishman and Alan Yuille propose a coarse-to-fine approach to deal with this problem, they train two deep neural networks using different regions of the input volume. The first one, the course-scaled model, takes the entire volume as its input. It is used for roughly locating the spatial position of the pancreas. The second one, the fine-scaled model, only sees a small input region covering the pancreas, thus eliminating the background noise and providing more accurate segmentation especially around the boundary areas. The results show that it improve the segmentation accuracy by a little (about 0.3%) (Zhou, Xie, Shen, Fishman, & Yuille, et al.,2016) . In (Pouliakis et al., et al.,2016). But after an image is segmented into regions; the resulting aggregate of the segmented pixels is represented and described for further computer processing. The neural network is under probabilistic neural is the Neocognitron neural network.

Nicola Bova, Viktor Gál, Óscar Ibáñez, and Óscar Cordón propose an alternative approach for traditional optimization-based deformable model adjustment. This involves a general purpose, machine learning-based image segmentation framework that translates the available information into a different type of decision process. They test it against a large set of state-of-the-art segmentation algorithms over two well-known image datasets with different image modalities and target structures. The results shows that the proposed framework is not intended for a specific segmentation problem, its implementation is competitive or even outperforms most of the state-of-the-art algorithms specifically designed for the segmentation tasks at hand (Bova & G, et al.,2016) .

Universiti Utara Malaysia

2.2.2.5 Recurrent Neural Network (RNN)

Recurrent Neural Networks (RNNs) are networks that operate in discrete time using feedback connections (Olah & Carter, et al.,2016). An RNN has a set of layers, in each time step each one taking a real value, and a set of weighted connections between them. The input layer works by setting it from the environment and the output one set by using the weights that are connects the hidden layer with the output layer(et)(et). The difference between RNN and feed forward architectures is that it isn't operates on an input space only, but also on an internal state space, which is trace of processed items by the network (Olah & Carter, et al.,2016).

There are several types of formal RNN models. Discrete-time models which are represented mathematically as maps iterated over discrete time steps n = 1, 2, 3. Continuous-time models which are represented by the differential equations whose solutions are defined over a continuous time t. The continues model is usually used for the purposes of biological modeling, which describes activation signals on the level of individual action potentials (spikes) (Lipton & Berkowitz, et al.,2015) . The time in RNN is used as a counter of input stimuli, (each input has specific time sequence). So, the RNN can be describe as a neural network that operates in time (Figure 2. 3).

Universiti Utara Malaysia

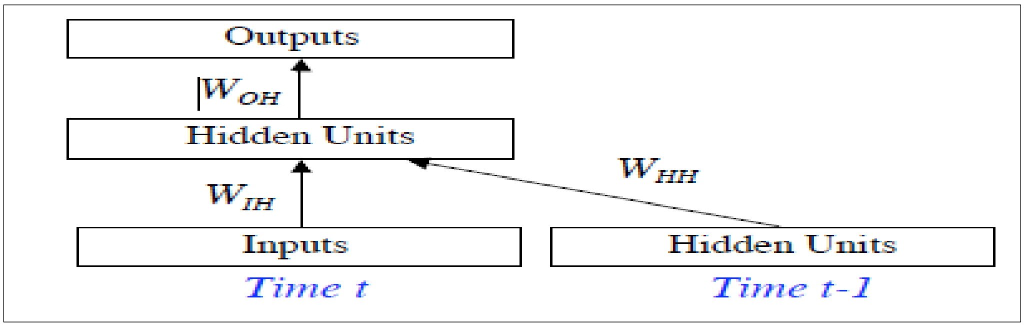


Figure 2. 3. Simple recurrent network

Resource. (Lipton & Berkowitz, et al.,2015)

Time for artificial systems can be any time step size suitable for the problem given. A delay unit here is used as simply delays the signal or activation until the next time step (Soltani & Jiang, et al.,2017) . However, the RNN suffer from some problems such as:

The recurrent structure make the training of RNNs difficult, since the gradient is distributed through the time instead of instantaneously (Bradbury, Merity, & Socher, et al.,2017) . So, any training algorithm for RNNs based on gradient suffers from the problems of gradient decay through the time and the network topology will slow the training and increase the computational complexity.

Other problem is the complexity of performance surfaces and containing of many local minima. Also the system becomes unstable during training because of the recurrent connections (Olah & Carter, et al.,2016) .

Universiti Utara Malaysia

From the discussion above it is seen that RNN using time as a counter for input steps (Bradbury et al., et al.,2017) . This means that the next input takes the time of current input time plus one-time unit. Time in HTM are used as supervisor for matching i.e. inputs are belongs to same group if they come in sequence or the same time.(Lipton & Berkowitz, et al.,2015) . Table 2.2 presents the summary of existing work in statistical approach.

Table 2.2

Summary of existing researches in statistical approach.

Author	Year	Technique	Advantage	Disadvantage	Domain
Sai Prasad Reddy and Dr. K. Nagabhushan Raju	2016	Eigen faces system based on PCA	Increase recognition rate with the number of training images per person	If distance is not zero, it will not be recognized	Face recognition
Muzameel Ahmed and V.N. Manjunath Aradhya	2016	Shape feature extraction based on PCA	More efficient	Cannot handle noisy case	MPEG-7 dataset
Fujiao Ju, Yanfeng Sun, Junbin Gao, Simeng Liu, Yongli Hu, Baocai Yin	2016	A Bayesian inference scheme	More accurate reconstruction errors and recognition rates	Not suitable for large database	Medical images
Murali	2016	PCA	Reduce the dimensions	Handling noisy case	Face recognition
Yu-Feng Yua, Dao-Qing Daia , Chuan-Xian Rena, Ke-Kun Huangb	2017	DMSC	More robust and has higher breakdown point	Handling noisy case	Face recognition
			SVM		
Shahrour	2014	Multi-class Support Vector Machine	Improve the day to day operational decision-making process	Long -term response strategy	pipelines leakages

Universiti Utara Malaysia

Table 2.2 continued

Min Zhang and Wenming Cheng	2015	Feature extraction, classifier, and parameters optimization	Effectively recognize mixture CCPs	Overlapped data	Medical images
Guangyi Chen, Tien. D. Bui and Adam Krzyżak	2016	SVM	Higher classification rates	Difficult with overlapping	Medical images
Saurabh Paul, Malik Magdon-Ismail and Petros Drineas	2016	SVM	prove margin is preserved to within ε-relative error in the full feature space	No known provable guarantees.	Medical images
Vasileios Mygdalis, Alexandros Iosifidis, Anastasios Tefas, Ioannis Pitas	2016	Graph-Embedded One-Class SVM and Graph Embedded SVDD		Overlapped data	Medical images
Ariffanan Mohd Basri					
			LDA		
Radu Dobrescu	2013	Fractal dimension and lacunarity	Improve classification accuracy	Big dimension	Medical images
Telgaonkar Archana H and Deshmukh Sachin	2015	KNN and Nave Bayes clssifier	Reduce the dimension on high dimensional data set	Not good with small dimention	Object and face images

UUM Universiti Utara Malaysia

Table 2.2 continued

Mokeev & Mokeev	2015	LDA+PCA	Reduce the complexity of computation	Not handle noisy cases	Face images
Aytug Onan, Hasan Bulut and Serdar Korukoglu	2016	Ant clustering algorithm	Improve performance	Big data	Text document
Radu Dobrescu, Loretta Ichim, Daniela	2013	Fractal dimension and lacunarity	Non-regularities. Classification accuracy improved	Small data set	Breast cancer
Vaibhav Garg	2014	GMDH-NN	Capture the trend of sediment of correlation	Not efficient in Small dataset	geomorphological
Kumar Adhikari, Jamuna Kanta Sing and Dipak Kumar Basu	2015	Spatial fuzzy C-means	Solve sensitivity to noise and intensity	Low accuracy	Brain image
Esmaeil Mehdizadeh, Amir Golabzaei & Kun Chen	2016	Heuristic fuzzy clustering	Good performance	Increases the convergence rate	White blood cells
Chen	2017	Kernelized fuzzy - means	Removes the skull region, robust against noises	Noisy images	Brain image

Universiti Utara Malaysia

Table 2.2 continued

Alexander Serb, Johannes Bill, Ali Khiat, Radu Berdan, Robert Legenstein & Themis Prodromakis	2016	Probabilistic neural	Successful learning of corrupted input data	Low performance with big data.	Face recognition
Nicola Bova, Viktor Gál, Óscar Ibáñez, and Óscar Cordón	2016	Probabilistic neural	competitive segmentation	Not intended for a specific segmentation	Organ images
Zhou, Xie, Wei, Fishman & Yuille	2016	Probabilistic neural	Eliminating the background noise.	accurate segmentation is small	Organ images

Universiti Utara Malaysia

2.2.3 Structural

Referred to as syntactic pattern recognition in some cases because it used to recognize formal language theory, depends on syntactic grammars to separate among data that related to different depending on the morphological interrelationships found within the data (Zipfel, et al.,2014) . One of the problems of structural pattern recognition systems are the difficulties of applying it to new domains because it require domain knowledge from the experts to implement of both description and classification (Pelillo, Elezi, & Fiorucci, et al.,2016) . Poonam Sao, Rajendra Hegadi, Sanjeev and Karmakar describes about the analysis of electrocardiogram (ECG) signals using different pattern recognition approaches, template matching, statistical classification, syntactic or structural matching, and neural networks (Sao et al., et al.,2014) . Megdiel Jimenez Guarneros, Jesus Ariel Carrasco Ochoa and Jose Fco. Martinez Trinidad use structural approach to overcome the problem of selection of suitable set of prototype graphs that better describes the whole set of graphs. The method based on clustering for graph embedding, which select border prototypes and some non-border prototype. Results shows that the selected method better in accuracy and runtime (Guarneros, Ochoa, & Trinidad, et al.,2015) .

Universiti Utara Malaysia

2.2.3.1 Hierarchical Temporal Memory (HTM)

Hierarchical temporal memory is a machine learning model found by Jeff Hawkins and Dileep George of Numenta, Inc (George & Hawkins, et al.,2005b). HTM technology has been developed by a private company, called Numenta, who make it available to researchers and practitioners (Numenta, et al.,2009). It models some of the algorithmic and structural properties of the neocortex, which has been applied by a many researches that approve its effectiveness and priority over other pattern recognition methods because of temporal and spatial processing of pattern (George &

Hawkins, et al.,2005a). Hierarchical Temporal Memory (HTM) provides a flexible and biologically accurate framework for solving prediction, classification, and anomaly detection problems for a broad range of data types (Ahmad & Hawkins, et al.,2015). It is similar to Bayesian Networks, but differs in using the time, hierarchy, action and attention, it is a memory-prediction network that use the revision techniques and Bayesian belief propagation as learning algorithm (Rozado, et al.,2011). The authors consider the two most important capabilities of HTM, its ability to discover and infer causes (Hawkins & George, et al.,2006).

HTM differs ultimately from traditional neural networks as a simulation of neocortex (William, et al.,2012). In HTM, reversion of features is depend totally on the temporal information of pattern, while in neural network it depends on spatial information as a basic recognition item, specifically Neocognitron which is hardwired and in Convolutional Network and HMAX a simple spatial operators such as max or average is applied (Ahmad & Hawkins, et al.,2015) .

UUM Universiti Utara Malaysia

The HTM network simulate the biological cortical regions and sub regions of the neocortex (Ahmad & Hawkins, et al.,2015) . Its nodes train the data that serves for extracting the world’s causes invariant representations, by its temporal relations among data and by its hierarchical structure (Byrne, et al.,2015b) . HTMs are organized as a tree-shaped hierarchy of nodes (Hawkins & George, et al.,2006), where each node implements a common memory Function and learning algorithm (Figure 2.4). The HTM has the ability to resolving difficult pattern recognition problems that not yet be solved (Fan et al., et al.,2016). There are several research fields of the HTM, for example, in the field of PR problems addressed up to now using the NuPIC tool, in the field of control the learning and inference processes optimum values of the HTM nodes. This task is limited and completely depending on the user's trial-and error computer experimentation (Byrne, et al.,2015a). There are some characteristics that made Hierarchical temporal memory better than deep learning in pattern recognition (Hawkins & Ahmad, 2015), these characteristics are explained in table 2.3

Table 2.3

Summary of the characteristics of the two different approaches.

Simple Neural Network	**Biological Neural Network**
Deep Learning	Hierarchical Temporal Memory (HTM)
Large datasets	Data streams
Derived from labeled databases	Derived from unlabeled data streams
Classification	Prediction Anomaly detection Classification
Batch	Continuous
Requires labeled data	No
Hard	Easy
Simple	Realistic
No	Yes

Universiti Utara Malaysia

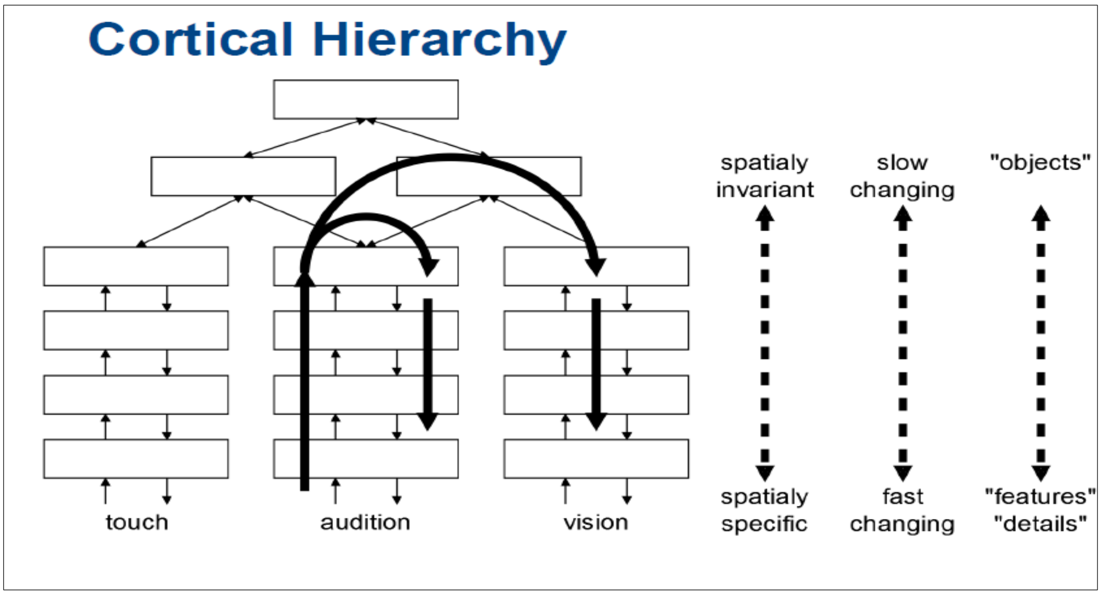


Figure 2.4. Simple Hierarchical Temporal Memory

Resource. (George & Hawkins, et al.,2005b)

Universiti Utara Malaysia

HTM described as a biologically inspired computational framework, it is a practical implementation of the memory-prediction theory of brain function presented by Hawkins (George & Hawkins, et al.,2005a). A node works in two modes, training mode (discovering causes) and sensing mod (inferring the causes). In training mode, the input data is examined in order to recognize patterns that repeat in both spatial and temporal dimensions (Jaros, George, Hawkins, & Astier, et al.,2014). The node consecutively make a groups of spatial patterns and identifies frequently observed temporal patterns (Diao & Kang, et al.,2014) . The grouping of spatial patterns is performed by an algorithm for cluster analysis (William, et al.,2012). Spatial patterns are assigned to groups based on their spatial similarity that are, small number of similar patterns than the possible patterns, so resolution in space is lost in each hierarchy level (Kalmar & Vida, et al.,2013). A mechanism must be provided to determine the probability that a new input belongs to the identified spatial groups (Ahmad &

Hawkins, et al.,2015) . Information on the membership of the inputs to the spatial groups is recorded in a time sequence which provides a basis for temporal patterns using a temporal data Mining algorithm (Wielgosz & Pietro´n, et al.,2016).

Regardless of its position in the hierarchy, each node in an HTM apply the same algorithm, in spite of that the hierarchical organization of the network cause that the nodes operate on different phenomenon regardless of the level that node is located (Mnatzaganian et al., et al.,2016). The nodes in lower levels recognize simple features that change quickly and require smaller spatial areas. Nodes in higher levels have complex features which are recognized by lower level nodes, so they are influenced by a greater range of data (George, et al.,2008). The higher level nodes deal with the complex causes, called patterns of patterns, which are less variant and exist in larger spatial areas (Jaros et al., et al.,2014) . This characteristic is similar to that of Slow Feature Analysis, which involves employing the temporal hierarchy as principle for learning to learn invariant features (Otahal, Najman, & Stepankova, et al.,2016) .

Universiti Utara Malaysia

For learning operation in a node, the input to the node is a sequence of patterns. Each node receives one input pattern at a time. For each input pattern, the node has three operations (George, et al.,2008) as follows:

1-Memorization of patterns

In this operation the patterns that are submitted to the input nodes of the first layer are compared with the patterns already stored in the memory (Wielgosz & Pietro´n, et al.,2016). If the comparison failed, then the pattern is added to the memory and remarked with a discriminating number (Byrne, et al.,2015b) .This comparison between input pattern and stored one done in supervised manner, which leads to time consuming (Jasmin Léveillé et al., et al.,2014), incorrect features extraction because any change in input pattern, for example change in pixel position due to viewpoint change, imaging process, or large interclass variations among the patterns, will cause incorrect matching with the stored patterns, which in turn leads to extract incorrect feature, furthermore leads to incorrect recognizing of input pattern (Kalmar & Vida, et al.,2013) .

Researchers improve this step by using different methods for memorizing input stimuli. Ioannis Kostavelis and Antonios Gasteratos apply the logo polar instead of Cartesian mapping of input image to the first level (Kostavelis & Gasteratos, et al.,2012). This results in enhancement recognition performance, but still has a problem with noisy data. Ali Nouri and Hooman Nikmehr apply a simple stochastic gradient descent learning algorithm in an unsupervised manner to learn and predict a multi-scale spatio-temporal patterns of the input signals to provide real-time prediction and robust of future inputs (Otahal et al., et al.,2016) , Still has a problem with noisy image. The second problem is the consideration in this research, which will be overcome by using the s-cell layer of Neocognitron neural network for its characteristic of robustness against shift in position (Jaros et al., et al.,2014). Figure 2.5 shows a block diagram of the S-layer structure.

Universiti Utara Malaysia

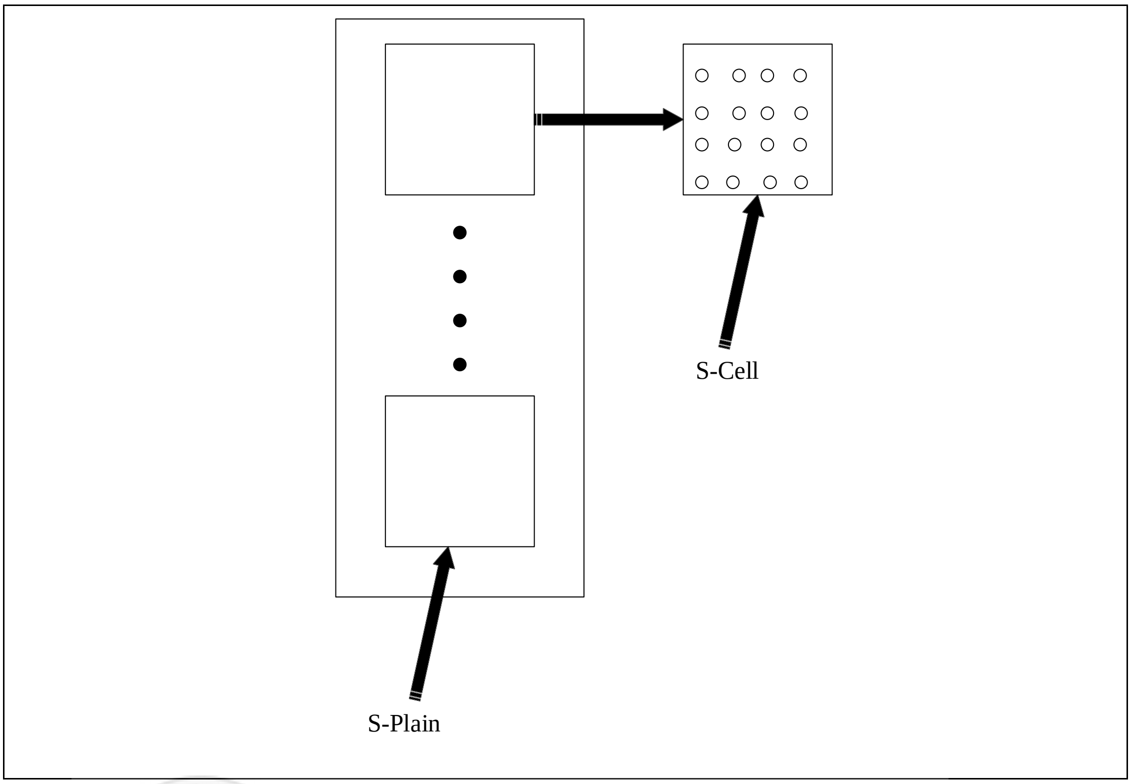


Figure 2.5. Block diagram of S-layer

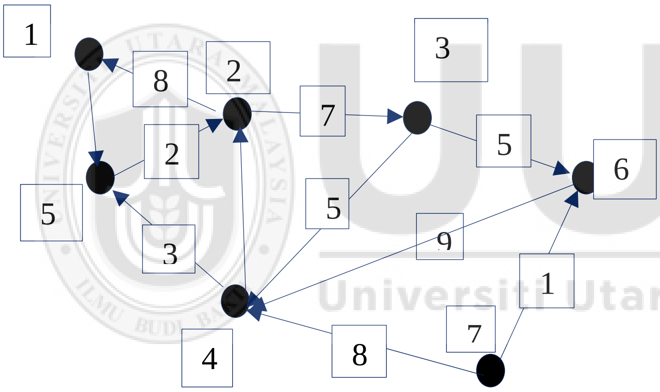
UUM Universiti Utara Malaysia

2- Learning transition probabilities

The main goal of the learning of HTM is correctly recognizing an invariant representations of the world's causes, this operation is done depending on the temporal relations among training data (Ahmad & Hawkins, et al.,2015) . To achieve this goal, innovators of the HTM model evaluate a frequency of transition events, this means, co-occurrences of the memorized patterns occurred in an adjacent time.

Then a first order Markov graph is applied to describe the temporal relations in a form of the vertices that are represent the memorized patterns and the links between these vertices which represents the transitions between patterns in time (Mnatzaganian et al., et al.,2016) . The vertices of the graph represent to the stored patterns. Each vertex is labeled by a number which represents that vertex. Adding new pattern to the memory contents, causing to add a new vertex to the Markov graph. This means that when input pattern (*i*) is followed by input pattern (*j*) for the first time, then introducing a link between the vertices (*i*) and (*j*), after that, number 1 is set on that transition. For every time input pattern (*i*) followed by input pattern (*j*), the number on the link is increased by one and so on. The Markov graph is normalized so that the probability of the number of transition is estimated (Asgar & Salehi, et al.,2015) .

Figure 2. 6 gives an example of Markov graph. The normalization operation done by dividing the number of transition of the outgoing links of each vertex over the total number of transition from that vertex (Figure 2.7).



UUM
Universiti Utara Malaysia

Figure 2.6. Markov graph

Numbers on the black circuits (vertices) is a random number given to the input pattern. Numbers on the connections between vertices is the number of transitions between every two vertices, when these vertices have followed each other in time.

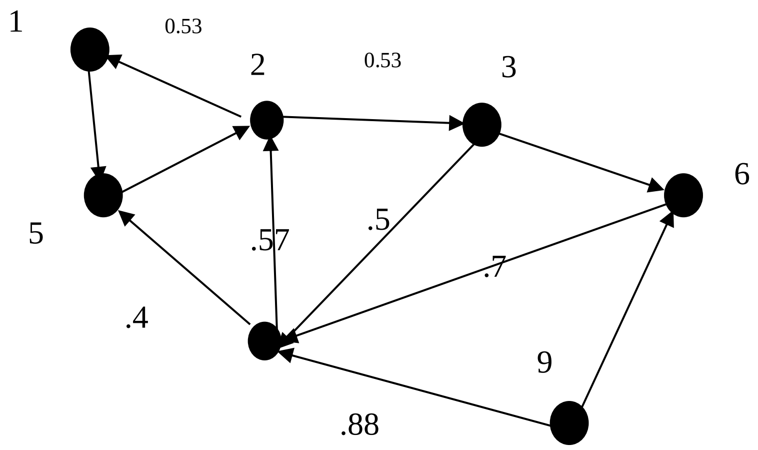


Figure 2.7. Normalizes Markov graph

Numbers on the links between vertices are the normalizing of the transitions (dividing the numbers of the transitions outgoing from each vertex by total number of transitions for the vertex).

3-Temporal grouping

The last step of the learning process is to analyze the normalized Markov graph into a set of temporal groups (Maciej Wielgosza & Pietron´, et al.,2016) . The main purpose of this operation is to make groups of the patterns (the vertices of the Markov graph) that are highly likely follow one another in time which means that they are related to the same cause (object) (Fan et al., et al.,2016) . Each temporal group is a subset of the set of vertices of the Markov graph.

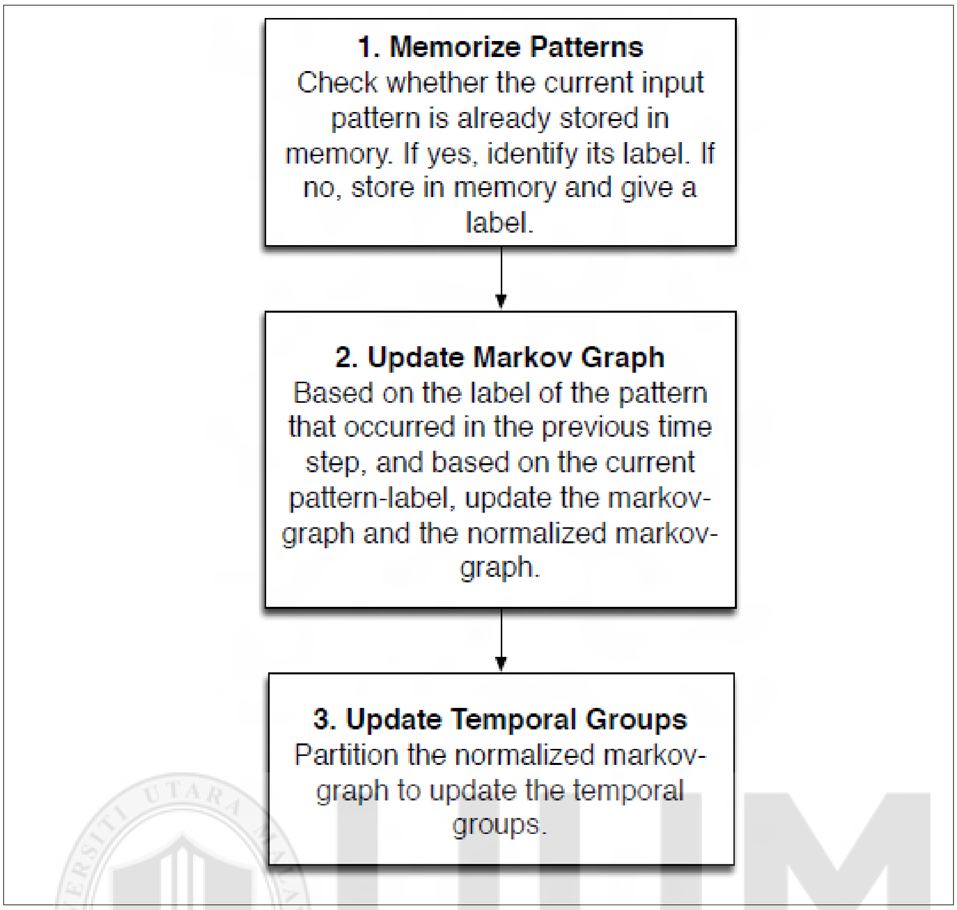
This operation is done in a way that the number of transitions between links is high enough that its vertices is of the same temporal group (Thornton, Main, & Srbic, et al.,2012). Then a clustering method is applied to form a temporal group (the original HTM use the Agglomerative Hierarchical Clustering algorithm (AHC)) by measuring the similarity between nodes (Diao & Kang, et al.,2014) . The nodes that are follow

Universiti Utara Malaysia

one another in time are putted in the same group, these groups are the temporal groups (Kostavelis & Gasteratos, et al.,2012). There is an output vector for each input pattern, that represents the degree of membership of the input pattern in its temporal groups (Diao & Kang, et al.,2014) . Degree of membership of the input pattern is calculated by how much the closeness of the vertices to each other within the temporal group. This produce an output vector the same length of the number of temporal groups (Abdel-Azim, et al.,2016). This algorithm has some disadvantages that affect the accuracy of recognition, like unable to handle overlap in data, sensitivity to noise and outliers, difficult to handle clusters of different size and convex shapes, breaking large clusters, and Sometimes the difficulty to identify the correct number of clusters by the dendrogram (Abdel-Azim, et al.,2016). Summary of the operations of the node at learning phase is illustrate in Figure 2.8.

Universiti Utara Malaysia

In sensing mode it involves classifying objects in the input data that belongs to a cause that previously discovered , then attempt to determine the most likely cause of the input after a novel input is encountered (Asgar & Salehi, et al.,2015). For every input pattern a node generate an output (Kostavelis & Gasteratos, et al.,2012). Each temporal group is a set of patterns with no respect to their transition probabilities within-temporal-group (Maltoni, et al.,2011). Figure 2.8 illustrates the operation of HTM node.



1. Memorize Patterns
Check whether the current input pattern is already stored in memory. If yes, identify its label. If no, store in memory and give a label.

2. Update Markov Graph
Based on the label of the pattern that occurred in the previous time step, and based on the current pattern-label, update the markov-graph and the normalized markov-graph.

3. Update Temporal Groups
Partition the normalized markov-graph to update the temporal groups.

Figure 2.8. Operations of node performs every time step during learning

Universiti Utara Malaysia

Resource. (George, et al.,2008).

The first disadvantage is under consideration in this research. This done by using fuzzy c-means clustering algorithm (FCM). FCM was proposed originally by Dunn and modified by Bezdek, it is one of the most common fuzzy clustering techniques, (Bezdek, et al.,1981). It has advantage of giving best result for overlapped data set (Chen, et al.,2017).

There are several clustering algorithms which can be used with pattern recognition. These are categorized into five categories: partitional clustering, hierarchical clustering, density based clustering, grid based clustering and model clustering (Das & De, et al.,2016).

In partitional clustering, the most popular clustering algorithm is K-means algorithm and fuzzy c-means algorithm, while in hierarchical there are two types: Agglomerative and divisive clustering. In density based clustering, density based spatial clustering of application with noise is the most popular algorithm, while in grid based, several approaches are found like STING and CLIQUE algorithms (Xu & Wunsch, et al.,2005).

Partitioning Based is suitable for large number of variables, may be faster than hierarchical method, it may produce constricted clusters than hierarchical clusters and does not require one to specify the number of clusters in advance. Can find arbitrarily shaped cluster.

Universiti Utara Malaysia

Fuzzy clustering methods produce overlapping clusters by assigning the degree of the membership to the clusters for each point. Traditional partitioning clustering methods, such as K-Means, and hierarchical methods produce separated clusters, which means that each data point is assigned to only one cluster. which is due to the tendency to partition a data set into a fixed number of clusters. hierarchical algorithms, produce a set of solutions with different numbers of clusters, which are then presented by a hierarchical graphical structure called dendrogram (Peng & Liu, et al.,2018).

FCM has the ability of determining and iteratively updating the values of membership of a data point in number of clusters that are defined previously (Chattopadhyay, Pratihar, & Sarkar, et al.,2011). So, any data point can be related to all clusters with its membership value (Adhikari et al., et al.,2015), i.e. each data shared a membership within each of the cluster centers that are defined previously. This algorithm try to assign membership to each data point corresponding to each cluster center (S. Das, et al.,2013). FCM is depends on a calculation of mean distance between each data point and group centroid point (Mehdizadeh et al., et al.,2016). The standard FCM apply Euclidean distance, which has the weakness of data correlation, which means that it is difficult to establish when there is a correlation in data points (Cai, Chen, & Zhang, 2010). The standard FCM also, do not utilize the spatial and temporal relationship of an image (Kannan, Ramathilagam, & Pandiyarajan, et al.,2011). Here More the data is near in space to the cluster center, more its membership to it (Zhang & Shen, et al.,2014). The objective function of clustering algorithm, which will be minimized, is defined as (Chen, et al.,2017) :

Universiti Utara Malaysia

$$J(u,v) = \sum_{i=1}^{n} \sum_{j=1}^{c} {\mu_{ij}}^{m} \left\| x_i - v_j \right\|^{q} \qquad (2.1)$$

where $\boldsymbol{m}$ is defined to any real number that greater than 1, $\boldsymbol{\mu_{ij}}$ is refer to a degree of membership of xj in cluster j, $\boldsymbol{x_i}$ is the ith of d-dimensional measured data, $\boldsymbol{v_j}$ is the dimension centroid of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the centroid, $\boldsymbol{c}$ represents the number of cluster center, n is the number of data points, $\left\|\boldsymbol{x_i} - \boldsymbol{v_j}\right\|$ is the distance between i^{th} data to j^{th} cluster center.

Beside an optimization of the objection function in each iteration, FCM algorithm also update $\boldsymbol{\mu_{ij}}$ membership matrix and centre of clustering $\boldsymbol{v_j}$ as(Chattopadhyay et al., et al.,2011):

$$\mu_{ij} = 1/\sum_{k=1}^{c}(d_{ij}/d_{ik})^{(\frac{2}{m-1})} \qquad (2.2)$$

$$v_j = (\sum_{i=1}^{n}(\mu_{ij})^m x_i)/(\sum_{i=1}^{n}(\mu_{ij})^m), \ \forall_j = 1,2,3 \dots . c \quad (2.3)$$

Where “d_{ij}” represents the Euclidian distance between i^{th} data to j^{th} cluster center Euclidian distance used with the FCM has the weakness of that measures can unequally weight the factors of algorithm (Lai & Garibaldi, et al.,2013). Also, in FCM, random choose is done for the cluster centers, this means that they are virtual, this make a specific cluster may be out of the data set (Singla & Khera, et al.,2016).

Universiti Utara Malaysia

In order to show different metrics, the following two points are used: point P has coordinate (1, 2, 3, 4) and point Q has coordinate (5, 6, 7, 8).

The Euclidean distance between point P and Q is:

Dpq = (2 − 3)2 + (3 − 5)2 + (4 − 7)2 + (5 − 9)2 =5.5

The Manhattan distance between point Pand Q is:

Dpq = |2−3| + |3−5| +|4−7| + |5−9| =10

The Chebyshev distance between point P and Q is:

Dpq = max {|2−3|, |4−7|, |5−9|} = max {1,2,3,4} =4

The Minkowski distance of order 3 between point P and Q is:

Dpq = (2 − 3 3 + 3 − 5 3 + 4 − 7 3 + |5 − 9|3)1/3 =4.6

Table 2.4 presents the distance measures and the application suitable for each.

Table 2.4

Distance measures and their applications

Measure	**Application**
Euclidean distance	K-means with its variants
Manhattan distance	Fuzzy ART, clustering algorithms
Chebyshev distance	Fuzzy C-means clustering
Minkowski distance	Fuzzy C-means clustering

It is obvious that the best metric is Chebyshev which have the minimum value between two points that relates to the same class.

Abbas Rammal, Eric Perrin, Valeriu Vrabie, Isabelle Bertrand and Brigitte Chabbert propose a weighted-covariance factor fuzzy C-means clustering method combined with bootstrapping. The algorithm can classify spherical and nonspherical clusters, in contrast to classic fuzzy C-means, which is only adapted to spherical clusters. Bootstrapping enables resampling of the available spectra to generate several datasets on which the classification is performed. (Rammal, Perrin, Vrabie, Bertrand, & Chabbert, et al.,2017). The results produce a better result than NIR spectra for the initial characterization and for dynamic samples, while MIR spectra acquired on raw samples, without soluble extraction, provided better classification than wet chemistry. Sunanda Das and Sourav De propose a modified genetic (MfGA) algorithm is proposed to improve the performance of FCM to oercome the drawbck of Convergence to local minima. the optimized class levels derived from the MfGA are

UUM Universiti Utara Malaysia

employed as initial input to FCM for finding global optimal solutions in a large search space. Results establishes the superiority of the proposed approach (Das & De, et al.,2016).

Zanaty proposes a new fuzzy c-means method for improving the magnetic resonance imaging (MRI) segmenta- tion. The proposed method called "possiblistic fuzzy c-means (PFCM)" which hybrids the fuzzy c-means (FCM) and possiblistic c-means (PCM) functions. It is realized by modifying the objective function of the conventional PCM algorithm with Gaussian exponent weights to produce memberships and possibilities simultaneously, along with the usual point prototypes or cluster centers for each cluster. The membership values can be interpreted as degrees of possibility of the points belonging to the classes, i.e., the compatibilities of the points with the class prototypes. For that, the proposed algorithm is capable to avoid various problems of existing fuzzy clustering methods that solve the defect of noise sensitivity and overcomes the coincident clusters problem of PCM (Zanaty, et al.,2013).

Universiti Utara Malaysia

As a result, HTMs provide the following four basic functions whatever the problem it applies to solve it. The first two are basic functions, the other two are optional (William, et al.,2012).

1) Discover causes in the world

2) Infer causes of input

3) Produce predictions

4) Direct behavior

There are researchers who work on this step to improve HTM performance try to use maybe different clustering algorithm according to type of pattern they work on. Subutai Ahmad and Jeff Hawkins examines Sparse Distributed Representations

(SDRs), the primary information representation strategy in Hierarchical Temporal Memory (HTM) systems and the neocortex. They derive several properties that are core to scaling, robustness, and generalization. They use the theory to provide practical guidelines and illustrate the power of SDRs as the basis of HTM. The goal is to help create a unified mathematical and practical framework for SDRs as it relates to cortical function (Ahmad & Hawkins, et al.,2015). Fergal Byrne presents a detailed description of HTM's Cortical Learning Algorithm (Assis et al., et), including for the first time a rigorous mathematical formulation of all aspects of the computations. Prediction Assisted CLA (paCLA), a refinement of the CLA is presented, which is both closer to the neuroscience and adds significantly to the computational power (Byrne, et al.,2015b).

On the other hand, Fan D, Sharad M, Sengupta A, Roy K. propose that the computing blocks for HTM can be mapped using low-voltage, magneto metallic spin-neurons combined with an emerging resistive crossbar network, which involves a comprehensive design at algorithm, architecture, circuit, and device levels. Simulation results show the possibility of more than 200× lower energy as compared with a 45-nm CMOS ASIC design (Fan et al., et al.,2016).

Jianhua Diao and Hyunsyug Kang proposes an Integrated Hierarchical Temporal Memory (iHTM) network for real-time continuous multi-interval prediction (RCMIP) based on the hierarchical temporal memory (HTM) theory. The iHTM network is constructed by introducing three kinds of new modules to the original HTM network. The iHTM network make sure newly arriving data is processed and RCMIP is provided. Performance evaluation shows that the iHTM is efficient in the memory and time consumption compared with the original HTM network in RCMIP (Diao & Kang, et al.,2014).

Marek Otahal, Michal Najman and Olga Stepankova integrate artificial neural network (probably of BICA type) with a real biological network (ideally in the future with the human brain) in order to extend or enhance cognitive- and sensory- capabilities. They propose to design such neuro-module using Hierarchical Temporal Memory (HTM) which is a biologically-inspired model of the mammalian neocortex. A complex task of contextual anomaly detection was chosen as the case-study (Otahal et al., et al.,2016).

Wielgosz present a method to classify objects in video streams using a brain inspired Hierarchical Temporal Memory (HTM) algorithm. present a method to classify objects in video streams using a brain inspired Hierarchical Temporal Memory (HTM) algorithm. The system achieves the highest F1 score of 0.95 (Maciej Wielgosza & Pietron´, et al.,2016).

Universiti Utara Malaysia

James Mnatzaganian Brings together all aspects of the spatial pooler (Singh & Nair, et), a critical learning component in HTM, under a single unifying framework. They optimize hardware designs. They verifies that given the proper parameterizations, the SP may be used for feature learning (Mnatzaganian et al., et al.,2016).

Recently, N.E. Osegi presents a first step in realizing useful HTM like applications specifically for mining a synthetic and real-time dataset based on a novel intelligent agent framework and demonstrate how a modified version of this very important computational technique will lead to improved recognition. The develop will provide a good starting point for developing structured object-oriented implementations of cortical learning microcircuits in real-time embedded applications (Osegi, et al.,2016). Table 2.4 illustrates summary of HTM in pattern recognition. Table 2.5 presents the summary of existing work of applying HTM with pattern recognition.

UUM Universiti Utara Malaysia

Table 2.5

Summary of HTM in pattern recognition

Author	Year	Technique	Advantage	Disadvantage	Domain	Problem to solve
Jianhua Diao, Hyunsyug Kang	2014	Integrated HTM	Efficient in the memory and time consumption s	Noise sensitive	Real-time application	Time and storage consuming in RCMIP
Subutai Ahmad, Jeff Hawkins	2015	Sparse Distributed Representations (SDRs)	Generalization, robustness	Relies on number of properties	Brain image	un unified mathematical and practical framework for SDRs
Fergal Byrne	2015	Cortical Learning Algorithm (Assis et al., et)	Significant computational power	Relays on high speed network	Medical image	CLA slow for applications need high speed computations
Fan D, Sharad M, Sengupta A, Roy K	2016	Mapping HTM to low-voltage, magnetometallic spin-neurons	200× lower energy	Depends on capacity of CMOS ASIC	Hardware design	High energy in power devices

Universiti Utara Malaysia

Table 2.5 continued

Marek Otahal, Michal Najman and Olga Stepankova	2016	Neural network with biological network	Successful anomaly detector for complex contextual patterns.	Low speed and sensitivity	Brain	Low cognitive- and sensory- capabilities of human brain
N.E. Osegi	2016	Improve first step HTM	Good for developing structured object-oriented	Not good with variation in input	Real time	Mining a synthetic and real-time dataset
Mnatzaganian	2016	spatial pooler (Singh & Nair, et), a critical learning component in HTM, under a single unifying framework	Good feature learning	Noise sensitive	Real time	Optimization of hardware designs
Wielgosz	2016	HTM	Achieves the highest F1 score of 0.95	Low speed	Video stream	Classify video stream

UUM Universiti Utara Malaysia

2.3 Multi- Label Classification in HTM

In multi- label there are three terms; multi- label data (MLD), multi- label classification (MLC) and multi- label learning (MLL). The MLD has more labels than features. Some MLDs have only a few labels per instance, while others have much of them. Most MLDs are imbalanced, which means that some labels are very frequent while others are rarely represented (Z. Zhang, Zhao, & Chow, et al.,2013). The labels in an MLD can be correlated or not. Moreover, frequent labels and rare labels can appear together in the same instances. The most basic information that can be obtained from an MLD is the number of instances, attributes and labels (Song et al., et al.,2013).

For any MLD containing |D| instances, any instanceD_i, $i \in \{1... |D|\}$ (Lou, Li, & Liu, et al.,2012) will be the union of a set of attributes and a set of labels (Xi, Yi), $Xi \in X^1\ x\ X^2 x \ldots\ldots x\ X^f$,Yi ∈ L, where f is the number of input features and X^j is the space of possible values for the $j-th$ attribute, j $\in \{1..f\}$. L being the full set of labels used in D, Yi could be any subset of items in L (Nasierding & Kouzani, et al.,2012).

On the other hand, multi- label classification assigns to each sample a set of target labels. This can be thought as predicting properties of a data-point that are not mutually exclusive, such as topics that are relevant for a document. A text might be about any of religion, politics, finance, or education at the same time or none of these. Each instance x_i is associated with a subset of labels Yi ⊆ L. Existing methods for multi-label classification can be grouped into two main categories; a) problem transformation methods, and b) algorithm adaptation methods. Multi-label classification problems can be found in various domains (Cabral, De la Torre, Costeira, & Bernardino, et al.,2011) including classifications of text document (Cabral

Universiti Utara Malaysia

et al., et al.,2011), bioinformatics data Yeast (Cabral et al., et al.,2011), and Genbase (Klimt & Yang, et al.,2004), Emotions related musical data (Han, Kamber, & Pei, et al.,2011), scene images (Streich, et al.,2010), textual email messages Enron (Klimt & Yang, et al.,2004) image and video annotation (S. Das, et al.,2013). The multi-label learning is a form of supervised learning where the classification algorithm is required to learn from a set of instances, each instance can belong to multiple classes. Multi-label datasets (MLDs) are generated from text documents (Klimt & Yang, et al.,2004), sets of images (Duygulu, Barnard, Freitas, & Forsyth, et al.,2002), music collections, and protein attributes (Diplaris, Tsoumakas, Mitkas, & Vlahavas, et al.,2005) among other sources. For each sample, a set of features (input attributes) is collected, and a set of labels (output label set) is assigned.

There is some work in the field of classification of multi- label data under taken using HTM. In (Mnatzaganian et al., et al.,2016), the author propose a mathematical framework in first step in HTM (Asgar & Salehi, et al.,2015) train a HTM network to recognize handwritten digits and letters taken from the well-known hoda dataset. They use multi -class SVM classifier. They achieved better performance than other methods. They achieved 99.12% recognition rate using modified HTM.

In using HTM to feature extraction field, several works have been repeated. In (Wielgosz & Pietro´n, et al.,2016), the authors use HTM–based system for object classification in video streams, They improve the feature extraction to achieve classification accuracy of the system. In (Maciej Wielgosza & Pietron´, et al.,2016) HTM is used for object classification in noisy video streams. This is done by improving the spatial pooler of the HTM, they do so by using SVM classifier and

Universiti Utara Malaysia

histogram calculation module. They achieve approximately 12 times the noise reduction for a video signal with 13% distorted bits. In (Xie et al., et al.,2016) authors presents an optimized HTM method for multiple-object recognition and classification. They use locomotive object recognition; the proposed network also has a temporal module for feature extraction. They increase the network speed for identifying the input patterns, where the recognition rate was approximately 91.4%. (Duan, Yao, Wang, Bai, & Yue, et al.,2016) presents an integrative approach of intrinsic time-scale decomposition ITD and hierarchical temporal memory for gearbox diagnosis under variable operating conditions. A total of two modules are emphasized including a feature extraction method based on ITD and an integrative feature fusion and classification model. The proposed method obtained features that are insensitive to variable operating conditions including speed and load changes. (Dornaika et al., et al.,2016a) proposes a novel discriminant semi-supervised feature extraction method for generic classification and recognition tasks. This method, called inductive flexible semi-supervised feature extraction, is a graph-based embedding method that seeks a linear subspace close to a non-linear one. They extend the proposed method to the case of non-linear feature extraction through the use of kernel trick. This latter allows to obtain a nonlinear regression function with an output subspace closer to the learned manifold than that of the linear one. Obtained results demonstrate a significant improvement over state-of-the-art algorithms that are based on label propagation or semi-supervised graph-based embedding (Dornaika et al., et al.,2016a). Table 2.6 presents the summary of existing work of HTM with multi- label datasets.

Universiti Utara Malaysia

Table 2.6

Summary of HTM with multi- label classification and feature extraction

Author	**Year**	**Technique**	**Advantage**	**Disadvantage**	**Problem to solve**
(Mnatzaganian et al., et al.,2016)	2016	mathematical framework for first step in HTM	Dimensionality reduction	Time consuming with big datasets	Recognition of handwritten digits
Asgar & Salehi	2016	Multi- class SVM classifier	Optimum recognition rate	More classes lower recognition	Recognition of handwritten digits
Wielgosz & Pietro´n	2016	Improved feature extracting	Improve classification accuracy	Noise sensetive	Feature extraction in stream data
Maciej Wielgosza & Pietron	2016	using SVM classifier and histogram calculation module	Noise reduction and distorted bits	Unsuitable for big dataset	Object classification in noisy video streams
XIE, YANG, PENG, GAO, & TAN	2016	optimized HTM method for multiple-object recognition	Increase the network speed and recognition rate was approximately 91.4%.	Not good for big datasets	Multiple object recognition
Duan, Yao, Wang, Bai, & Yue	2016	ITD and hierarchical temporal memory	features that are insensitive to variable operating	More features, more complex.	Gearbox diagnosis under variable operating conditions
F. Dornaika, Y. El Traboulsi and A. Assoum	2016	Seeks a linear subspace close to a non-linear one	obtain a nonlinear regression function than that of the linear		non-linear feature extraction through the use of kernel trick

Universiti Utara Malaysia

2.4 Neocognitron Neural Network

Neocognitron neural network is a multi-layer hierarchal neural network proposed by Fukushima in 1979 (Fukushima, et al.,2013a) its architecture was initially suggested by neurophysiological findings on the visual systems of mammals (Kunihiko, et al.,2016). It is a hierarchical multi-layered network and acquires the ability to recognize robustly the visual patterns over learning. It consists of layers of S-cells, which resemble simple cells of the visual cortex, S-cell layers has variable input connection which are modified over learning and they are feature-extracting cells. And layers of C-cells, which resemble complex cells, its input connections are fixed and unmodified, exhibit an approximate invariance to the position of the stimuli presented within their receptive fields (Kunihiko, et al.,2016). The C-cells is the recognition cells at the highest stage of network, which indicate the result of the pattern recognition. These layers of S-cells and C-cells are arranged alternately in a hierarchical manner as in Figure 2.9 (Fukushima & Shouno, et al.,2015). Behavior of S-cell is illustrated in Figure 2.10. The Neocognitron network is capable of recognizing input patterns robustly, with some portion of effect from deformation, shift in position or change in size. For this reason it is a powerful learning algorithm for image recognition (Fukushima, et al.,2013b).

Universiti Utara Malaysia

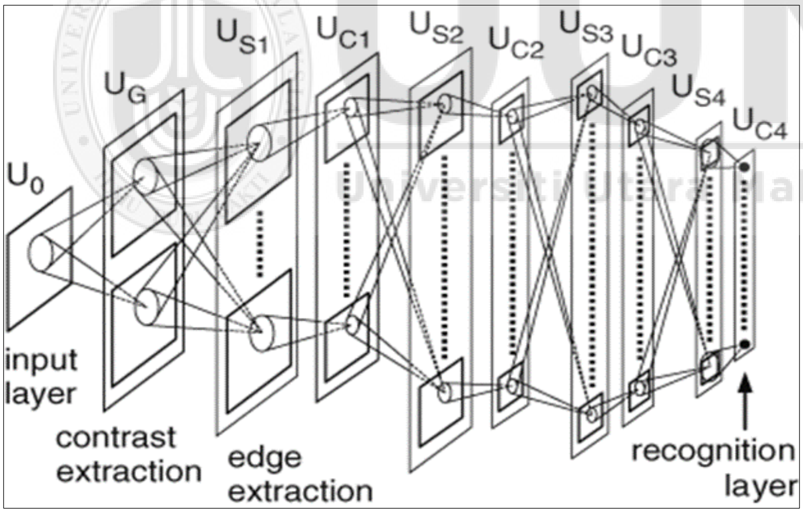


Figure 2.9. Structure of Neocognitron neural network

Resource. (Fukushima & Shouno, et al.,2015)

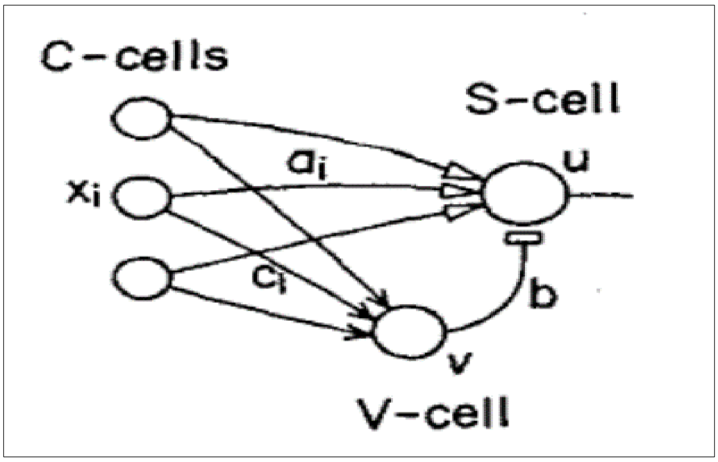


Figure 2.10. Structure of S-cell

Resource. (Fukushima, et al.,2016)

Every cell in S-cell as in Figure 2.10, behave as follows: Let *ai* be the strength of the excitatory variable connection to an S-cell from the *ith* C-cell, whose output is *xi*; and b be the inhibitory variable connection from the V-cell whose output is v, Also

Universiti Utara Malaysia

$$u_{Sl}(k,\mathbf{n}) = \frac{\theta l}{1-\theta l}\cdot\varphi\left[\frac{1+\sum_{k\ =1}^{Kc_{l-1}}\sum_{v<AS_l} as_l(v,k,K)\cdot u_{Cl}(k\ \ ,\mathbf{n}+v)}{1+\theta\cdot bs_l(k\ \ )\cdot v_l(\mathbf{n},\mathbf{k})}-1\right] \tag{2.4}$$

$$\varphi[x] = \begin{cases} x & if \quad x \geq 0 \\ 0 & if \quad x < 0 \end{cases}$$

Where $\varphi[]$is a function defined by $\varphi[x]$=max$(x, 0)$, Parameter $asl(v, k, K)$ is the strength of variable excitatory connection coming from C-cell of the preceding stage For l=1; however, $u_{cl-1}(n, k)$stands for, $u_G(n, k)$; It should be noted here that all cells in a cell-plane share the same set of input connections, hence $asl(v, k,)$ is independent of n. asl denotes the radius of summation range of n; that is, the size of spatial spread of input connections to a particular S-cell. θ is a constant (0<θ<1) determining the threshold of the S-cell.

Parameter $b_l(k)(\geq 0)$; is the strength of variable inhibitory connection coming from the V-cell given by the equation

$$b = \sqrt{\sum \frac{a^2}{c_i w_i}} \tag{2.5}$$

$c_i w_i$ is the strength of the excitatory connection to the inhibitory V-cell from the *ith* C-cell, whose output is v.

UUM Universiti Utara Malaysia

The positive constant θl is the threshold of the S-cell and determines the selectivity in extracting features.

The output of s-cell passes to the next c-cells. This output will be passed to the s-cells of the next layer, and so on to the last layer. The Neocognitron algorithm has several improvements, these improvements accelerate the recognition network. In 2013, the author use add-if-silent rule in order to make the learning simpler and stable and reduce the computational cost for learning. With respect to the add-if-silent rule, shown in Figure 2.11 input connections to a cell are created only at the moment when the cell is generated. The input connections are not modified any more after generating the cell and added to the network, never mind what is the training stimuli are given to the network.

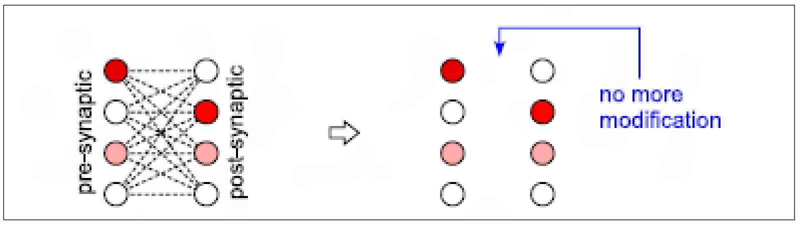


Figure 2.11. The add-if silent learning rule

Resource. (Fukushima, et al.,2013a, et al.,2016)

In the conventional Neocognitron, the inhibitory signals from V-cells are work in a divisional manner. In the most improved Neocognitron, on the other hand, the inhibition works in a subtractive manner, this will make Neocognitron much more robust against the noise of background (Fukushima, 2011).

UUM
Universiti Utara Malaysia

Here the output U of the S-cell is given by:

$$U = (1/1 - \theta\ ).\varphi[\ \textstyle\sum_n a_n x_n - \theta v] \qquad (2.6)$$

Where $\varphi[]$is a function defined by $\varphi[x]$=max$(x, 0)$, the strength of the inhibitory connection is θ, which determines the threshold of the S-cell (0 ≤ θ < 1). The response of the V-cell is given by:

$$v = \sqrt{\textstyle\sum_n c_n x_{n^2}} \qquad (2.7)$$

Where c_n, is the strength of the fixed excitatory connection from the nth C-cell.

Fukushima use Neocognitron in recognizing handwritten digits with its improvement versions to achieve high recognition rate. He uses supervise learning at highest stages, self-organization of line-extracting. This allowed to remove the accessory circuits that were appended to the previous versions (Romanuke, et al.,2016) . He uses Neocognitron to be able to incremental learning to speed up the recognition rate, by using sequential construction of layers. In 2015, David Guichard use interpolating vectors algorithm to reduce the error rate from 1.52% to 1.02% (Guichard., et al.,2015). In 2012, John H. Heinbockel produce another version of Neocognitron for removing noise in background of handwritten digits. He use subtractive inhibition to S-cells from V-cells with a root-mean-square. This improvement make the recognition much robust against background noise (Heinbockel, et al.,2012).

Universiti Utara Malaysia

Oleksii Gorokhovatskyi, use simplified Neocognitron neural network model as a tool for practical recognition of handwritten mark images. Simplification of Neocognitron structure from only two stages and fixed number of feature-extraction planes is proposed, the overall stages of solving practical image processing problem are described. Recognition properties of simplified net are investigated, it is shown that for particular task simple structure of net may be the same effective (up to 98% of correct recognition) as traditional four-stages one (Gorokhovatskyi, et al.,2016).

Kunihiko Fukushima and Hayaru Shouno improve Neocognitron neural network to reduce the computational cost. They propose to search the nearest plane, not among all possible combinations of three reference vectors, but only among trios that contain the nearest reference vector. For reducing the computational cost, it is also important to represent the large number of training vectors accurately with a compact set of reference vectors. To create a compact set of reference vectors, the learning is carried out in two steps. The effectiveness of the proposed method for recognizing hand-written digits is demonstrated and proved (Fukushima & Shouno, et al.,2015).

Toshisada Mariyama, Kunihiko Fukushima, proposes a method to automate small network design for a regression problem based on the Add-if-Silent (Alqaisy, Yassen, Khalifa, & Salim, et) function used in the Neocognitron. Because the original AiS is designed for image pattern recognition, this study modifies the intermediate function to be Radial Basis Function (RBF). This study shows that the proposed method can determine an optimized network structure using the Bike Sharing Dataset as one case study (Mariyama et al., et al.,2016)

Fukushima use an improved version of Neocognitron by using add-if-silent rule in the middle stages. In the highest stages, he uses interpolating vector. He get higher recognition rate with a smaller scale of the network (Fukushima, et al.,2013b) for recognizing handwritten digits.

Universiti Utara Malaysia

In 2014 Jasmin Léveillé, Isao Hayashi, and Fukushima use generalization of the WKL rule within Neocognitron neural network to learn from high-dimensional data. The result of their work shows that using this rule leads to a small reduction in overall accuracy, and a major reduction in the number of coding nodes in the network. This permits incremental learning and retain most of the useful structure in the input data (Léveillé, Hayashi, & Fukushima, et al.,2014).

Poletaev (Poletaev, Pervunin, & Tokarev, et al.,2017)presents a method of bubbles images identification based on a modern technology of deep learning called convolutional neural networks (CNN). Neural networks are able to determine overlapping, blurred, and non-spherical bubble images. Results show that They can increase accuracy of the bubble image recognition, reduce the number of outliers, lower data processing time, and significantly decrease the number of settings for the identification in comparison with standard recognition methods developed before (Poletaev et al., et al.,2017).

Zhang and his partners develop HiCPlus, a computational approach based on deep convolutional neural network, to infer high-resolution Hi-C interaction matrices from low-resolution Hi-C data. they compared the overlap between significant interactions identified in three interaction matrices with the ChIA-PET identified interactions. 67% of ChIA-PET-predicted loops can be recovered by interactions identified in the real and HiCPlus-enhanced matrices, respectively, while only 9% of the ChIA-PET interactions can be recovered by the down-sampled Hi-C matrix (Zhang et al., et al.,2018).

Peng, Liwen and Liu, Yongguo consider feature dependence and feature interaction simultaneously, and they propose a multi- label feature selection algorithm as a preprocessing stage before MLC. In practice, the instances of multi- label datasets are distinguished in a single cluster by such frameworks. In practice, the instances of multi- label datasets are distinguished in a single cluster by such frameworks (Peng & Liu, et al.,2018). Table 2.7 presents the summary of existing work of Neocognitron in pattern recognition.

Universiti Utara Malaysia

Table 2.7

Summary of Neocognitron in pattern recognition

Author	Year	Technique	Advantage	Domain
John H. Heinbockel	2012	Subtractive inhibition to S-cells	Removing noise	Handwritten digit
Fukushima	2013	Add-if-silent interpolating vector in the middle stages.	Higher recognition rate with a smaller scale of the network	Handwritten digit
Kunihiko Fukushima	2013	Add-if-silent rule, interpolating	Higher recognition rate with a smaller scale of the network	Handwritten digits.
Jasmin Léveillé, Isao Hayashi, Kunihiko Fukushima	2014	WKL rule	Small reduction in accuracy	Handwritten digits.
David Guichard	2015	Interpolating vectors	Reduce the error rate	Handwritten digit
Oleksii Gorokhovatskyi	2016	Two step Neocognitron	Up to 98% of correct recognition	Handwritten mark images
Kunihiko Fukushima and Hayaru Shouno	2015	Search the nearest plane	Reduce the computational cost	Handwritten digit
Toshisada Mariyama, Kunihiko Fukushima	2016	Improve Add-if-Silent function to RBF	Optimized network structure	Bike sharing dataset
Poletaev	2017	deep learning	increase accuracy, reduce the number of outliers	Image recognition

UUM
Universiti Utara Malaysia
UNIVERSITI UTARA MALAYSIA BUDI BAHASA

Zhang	2018	deep convolutional neural network	high-resolution Hi-C interaction matrices	Pattern recognition
Peng	2018	feature selection	Multi- label datasets are distinguished in a single cluster	Pattern recognition

2.5 Summary

Pattern recognition is a classification tool for a specific object. It is part of machine learning of artificial intelligent. The researchers claimed that the almost techniques of the three traditional approaches of pattern recognition, template matching, statistical and structural suffers from handling noisy images and overlapped data. Since its innovation on 2004, many researchers try to use the memory prediction theory and its practical implementation Hierarchical Temporal Memory in pattern recognition area. The idea behind these researches is to build intelligent machine that mimic the mammalian brain in recognizing patterns. The basic two feature of the brain is hierarchy and time(Hawkins & George, et al.,2006). The HTM model meet these features. The basic three operations of HTM node is memorizing, which include the store of input stimuli which is not occurred in memory. Learning transition probabilities, which construct a Markov graph for the stored inputs. Temporal grouping, which apply a clustering algorithm to relate each input to its group (George, et al.,2008). S-cells are feature-extracting cells, which have variable input connections and are modified over learning, so it is used in this research to redesign memorizing step of HTM. And C-cells, which resemble complex cells, which have fixed input connections and unmodified. The C-cells are recognition cells which appear in the

Universiti Utara Malaysia

highest stages, which indicate the result of the pattern recognition, so it is not used in first step of HTM. The researcher applies different learning rules and use different patters to reach the optimum recognition. The most common problems with the use of HTM with image processing field is shift in position and capacity of memory.

Universiti Utara Malaysia

CHAPTER THREE
RESEARCH METHODOLOGY

3.1 Introduction

The purpose of this chapter is to describe the methodology used in conducting this research. Based on experimental methodology used in many fields such as medical applications, machine learning, and neural networks (Abdel-Azim, et al.,2016). The main elements in the methodology include redesign of two components of the original HTM, namely the memorizing step (Feature Extraction) and clustering algorithm. The block diagram of the component of the proposed HTM is illustrated in Figure 3.1, while Figure 3.2 shows the experimental steps.

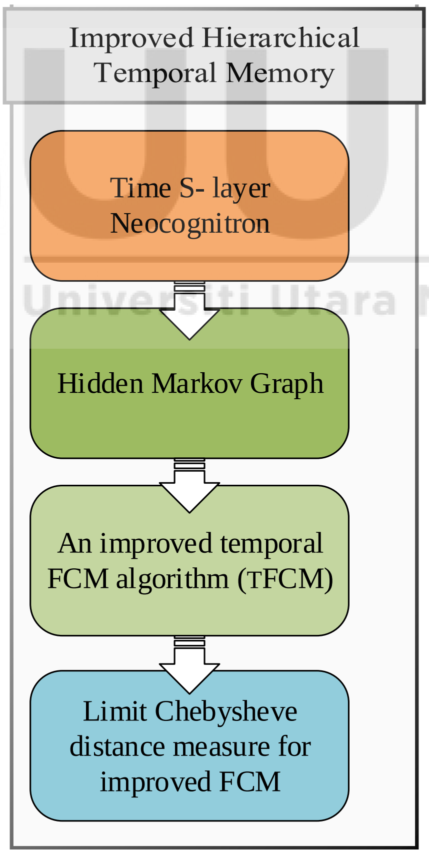
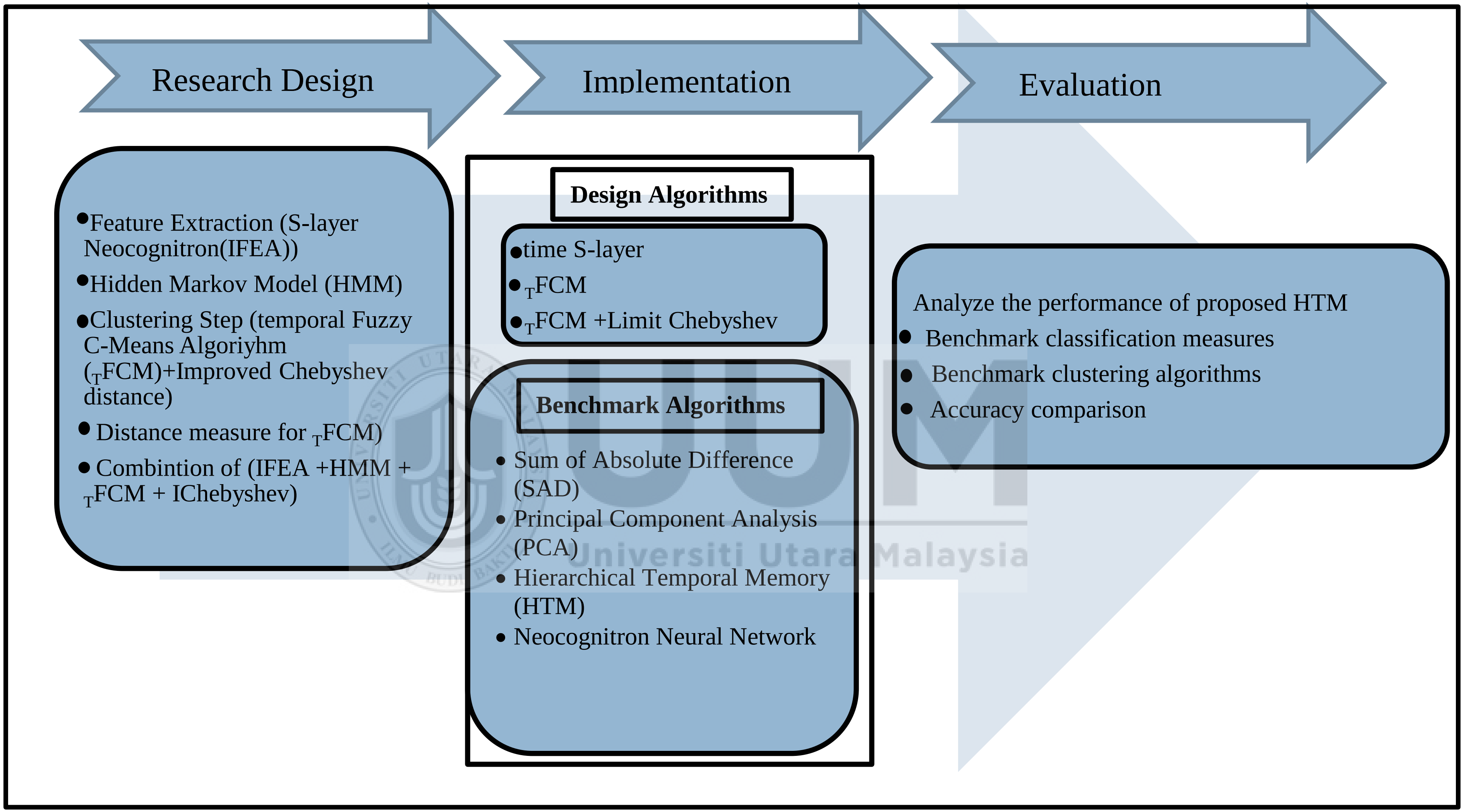

Improved Hierarchical Temporal Memory
Time S- layer Neocognitron
Hidden Markov Graph
An improved temporal FCM algorithm (TFCM)
Limit Chebysheve distance measure for improved FCM

Figure 3.1. Component of proposed Hierarchical Temporal Memory for multi- label data.

Universiti Utara Malaysia


Research Design
Implementation
Evaluation
●Feature Extraction (S-layer Neocognitron(IFEA))
●Hidden Markov Model (HMM)
●Clustering Step (temporal Fuzzy C-Means Algoriyhm ($_{T}$FCM)+Improved Chebyshev distance)
● Distance measure for $_{T}$FCM)
● Combintion of (IFEA +HMM + $_{T}$FCM + IChebyshev)
Design Algorithms
●time S-layer
●$_{T}$FCM
●$_{T}$FCM +Limit Chebyshev
Benchmark Algorithms
• Sum of Absolute Difference (SAD)
• Principal Component Analysis (PCA)
• Hierarchical Temporal Memory (HTM)
• Neocognitron Neural Network
Analyze the performance of proposed HTM
● Benchmark classification measures
● Benchmark clustering algorithms
● Accuracy comparison

UUM Universiti Utara Malaysia

Figure 3.2. Experimental research steps.

3.2 Research Design

This research proposes a Hierarchical Temporal Memory with two new algorithms (Feature Extraction and Clustering Algorithm) and one new distance measure.

The proposed HTM improves the accuracy of recognition, with two algorithms: S-layer Neocognitron neural network and FCM clustering algorithm with improved Chebyshev distance. For the S-layer Neocognitron neural network, the research proposes an S-layer with a new objective function based on the time factor in calculating the response of S-cells for feature extraction. This will help in reducing the shift in position that could be found in patterns; by using time rather than space in the extraction of features. This will improve the accuracy of feature extraction and enhance the following steps. Additionally, time factor will be used thereafter in temporal FCM on the last step of HTM for temporal groups.

For clustering, this research proposes a membership function that incorporates temporal information for use in FCM. The proposed FCM minimizes the objective function with the lowest iteration count in order to enhance the clustering of overlapped data.

The proposed hierarchical temporal memory contains three steps: memorizing (feature extraction) step, Hidden Markov, and clustering step. Figure 3.3 illustrates the steps of the proposed HTM

Universiti Utara Malaysia

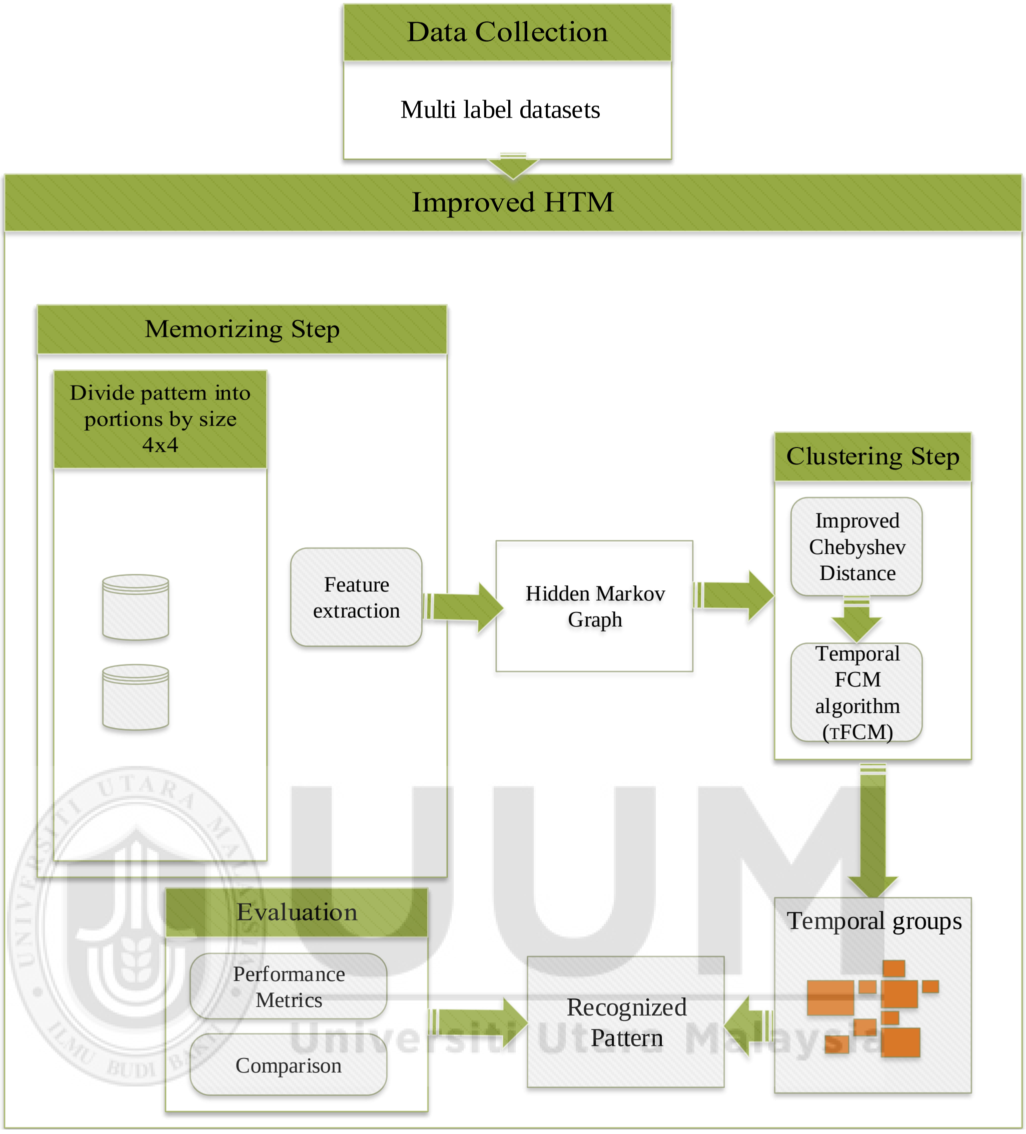


Figure 3.3. *Steps of proposed HTM*

3.2.1 Data Collection

In this research, multi- label datasets are used, which are blood type images and blood cancer images (leukemia), STARKEY'93, Yeast and Genbase datasets. In the following sections, this will be explained in detail.

Universiti Utara Malaysia

3.2.1.1 Blood Type

The dataset of blood type images was obtained from the National Center for Biotechnology and Information (NCBI). The dbRCB database provides an open, publicly accessible platform for clinical data related to human Red Blood Cells (RBC) (https://www.ncbi.nlm.nih.gov/). It combines the well-established Blood Group Antigen Gene Mutation Database (BGMUT) with tools and interlinked resources developed at the NCBI, images are sized 32x32; 2) publicly supervised. They saved as grayscale JPEG. Each input level nodes represents to one 4x4 portion of a vector of 16 components 64 nodes are arranged in an 8x8, i.e. each image is of 64 portions, each of 4x4 pixel size. Figure 3.4 gives examples of blood type images(Pablo et al., et al.,2015).

Universiti Utara Malaysia

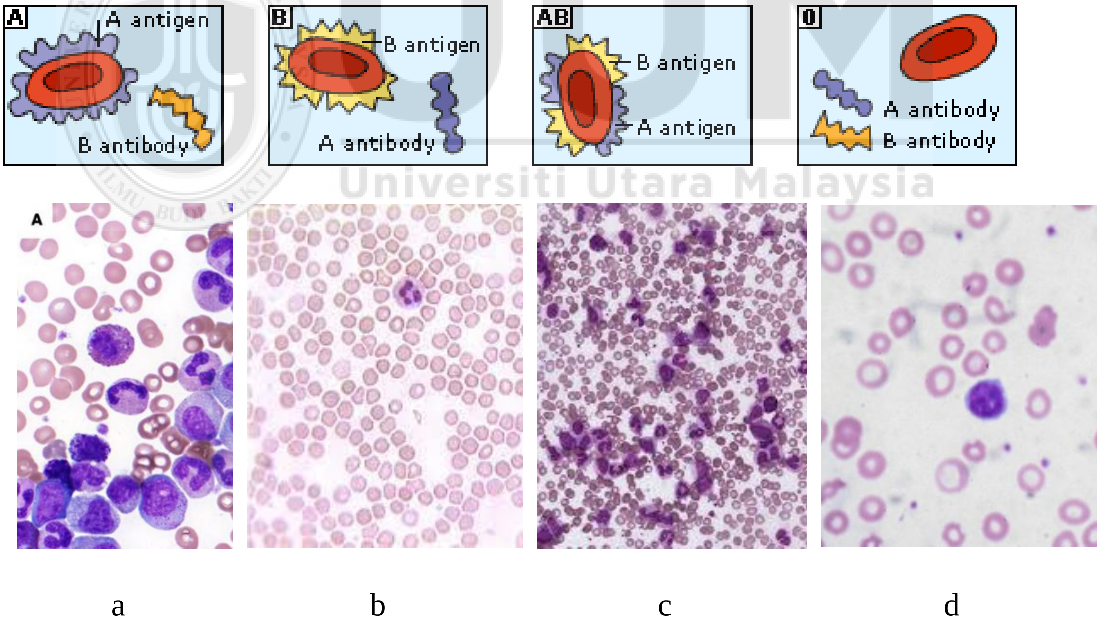
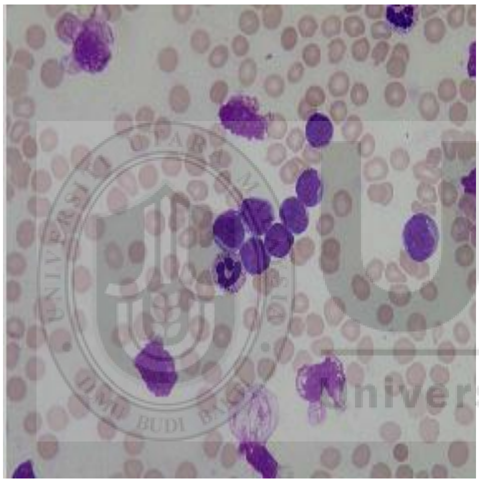


Figure 3.4. Four different types of blood groups; (a): type A, (b): type B, (c): type AB, (d): type O.

3.2.1.2 Blood Cancer

The dataset of blood cancer used in this research is the public supervised image datasets (ALL-IDB) (t "http://www.dti.unimi.it/fscotti/all," et) which has been provided to test and compare algorithms for cell classification of the cancer disease. There are two types of datasets available. The ALL-IDB1 can be used both for testing the classification systems and image preprocessing methods and ALL-IDB2 has segmented WBCs to test the classification of blast cells. A total of 108 peripheral blood smear images were obtained from the public dataset, Figure 3.5 gives an example of an images (Fathima, et al.,2013).



UUM Universiti Utara Malaysia

Figure 3.5 Cancer image example

As mentioned above, image is divided into 4x4 pixel portions, each portion is an input pattern that will be submitted to the input layer of HTM. A normalized Markov graph will be constructed then the improved clustering algorithm will be used to recognize the image (Figure 3.5).

3.2.1.3 Genbase, STARKEY”93 and Yeast

Three multi- label datasets have been used in this research, the STARKEY”93 dataset, contains 88 trajectories (corresponding to 33 elk, 14 deer, and 41 cattle), but each trajectory is very long. The whole dataset has 79,987 (x, y, t) observations, with an average of 909 observations per trajectory. The undertaken experiment was on datasets 1–4 from Nepusz (Rowland et al., et al.,1998), which contain pre-computed sequence similarities together with the ground-truth clustering. D1 contains 669 sequences and 5 ground-truth clusters, while D2 contains 587 sequences and 6 clusters, D3 contains 567 sequences and 5 clusters, and D4 contains 654.

The second dataset is the Genbase dataset, which is a biology dataset. it has 662 objects and 27 labels with a density of 0.046.

The third is the Yeast biology dataset. It has 2417 objects and 14 labels with 0.303 density and 4.237 cardinality. All of the employed datasets are available at the website: http://mulan.sourceforge.net/datasets.html.

Table 3.1 summarizes the properties of the datasets including the domain of the dataset; number of instances; number of attributes, and Label Cardinality (L_{CARD}) a measure of multi-labeled-ness‖ of a dataset introduced by (Tsoumakas, Katakis, & Vlahavas, 2010) which quantifies the average number of labels per example in a dataset; and Label Distinct Combinations (L_{DC}) a measure representing the number of distinct combinations of labels found in the dataset (Kafrawy, Mausad, & Esmail, et al.,2015b).

Tables 3.2, 3.3, 3.4 demonstrate sample of data.

Table 3.1

Description of multi- label datasets

Datasets	**Domain**	**Instances**	**Attributes**	**labels**	**L_{CARD}**	**L_{DC}**
Yeast	Biology	4417	72	14	4.237	198
Genbase	Biology	662	1186	27	1.252	32
STARKEY”93	Text	79987	2450	67	3.568	1230

Universiti Utara Malaysia

Table 3,2

STARKEY"93 dataset portion

UTMGri	UTMGr	UTMG	Id	Time	GMDate	GMTim	LocDate	LocTi	RadNu	UTM	UTMN	Yea	Grensu	Obs
374145	5014200	374145	5014200	06:02:02	890424E06	02:08:28	1993050	18:08:2	59	E	374137	93	12:34:00	1.62
374265	5010180	501018	21228E09	07:02:02	921228E09	1993050	1993050	00:55:5	46	E	501574	93	12:34:00	1.68
377475	5015730	377475	5015730	08:55:55	890222E01	23:02:02	19930507	00:55:55	40	E	377477	93	12:34:00	1.43
377055	5008470	377055	5008470	01:21:12	910313E37	17:21:12	19930507		67	E	19930506	93	12:34:00	1.73
376665	5008710	500871	376665	02:01:15	900219E15	18:01:15	1993050	18:01:1	78	E	376679	93	12:34:00	1.74
378165	5015280	378165	5015280	02:05:55	890418E01	18:05:55	1993050	12:34:0	117	E	378177	93	12:34:00	1.65
375105	5014500	375105	5014500	02:15:53	910315E15	18:15:53	19930507	18:15:53	113	E	19930506	93	12:34:00	1.39
376935	5008320	376935	5008320	02:25:51	910312E12	18:25:51	1993050	18:25:5	127	E	199305	93	12:34:00	1.62

UUM
Universiti Utara Malaysia

Table 3.3

Genbase dataset portion

Relation	Protein - Weka	Filters	Unsupervised	Instance	Randomize -S73	Attribute	Protein
O00060	O00139	O02741	O08424	O12984	O12988	O13224	O13351
0.47	0.44	0.4	0.74	0.68	0.54	0.67	0.36

Table 3.4

Yeast dataset portion

MIT	NUC	CYT	ME1	ME2	ME3	EXC	VAC	POX	ERL
0.61	0.58	0.51	0.78	0.86	0.41	0.74	0.37	0.44	0.78
0.47	0.44	0.4	0.74	0.68	0.54	0.67	0.36	0.38	0.74
0.47	0.57	0.42	0.17	0.39	0.39	0.39	0.46	0.56	0.48

3.2.2 Memorizing Step of HTM

This section includes the flow of proposed feature extraction S-layer Neocognitron in memorizing features of the input pattern. In the procedure, the S-layer operation of Neocognitron neural network is applied.

Universiti Utara Malaysia

Neocognitron consists of two types of cell, simple and complex. Simple cell works as feature extraction cells, which extract simple features of input pattern. Complex cells are the recognition cells at the highest stage of network, which indicate the result of the pattern recognition. The structure of S-cells is illustrated in figure 2.10 in Chapter two.

The features that are extracted by improved S-cells Neocognitron are combined with a time factor created with each input pattern to be presented to input layer of HTM. In the first step of the learning, patterns seen in the receptive field of the node are memorized. The memorization of patterns is as vector quantization process of the input data.

When the input pattern is presented to the input layer, the creation of a time matrix will take place. Each element of the time matrix corresponds to one input vector. An S-cell becomes a winning cell if the same pattern is presented to the input layer. Ones the winner cell is chosen from a cell-plane, modification of the input connections of that cell take place; otherwise, a new sub layer is created in the S-layer and the pattern is stored in a new cell of this sub layer. The data flow of the memorizing step is presented in Figure 3.6.

The patterns are stored in a node as the rows of a matrix. Each row of the matrix is of a different pattern. For example, say that the patterns' matrix is A, the individual patterns will be referred to as a_1; a_2 etc., depending on the rows of patters that are stored. The time matrix T element's, corresponding to each pattern are t_1, t_2 etc., while t_1 corresponds to a_1 and so on. The following presentation gives an example of how a matrix is constructed:

Universiti Utara Malaysia

$A = \begin{bmatrix} 10110010 \\ 00100110 \end{bmatrix} \begin{matrix} a_1 \\ a_2 \end{matrix}$ $\quad T = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ time matrix corresponding to input pattern

The resulting matrix:

$$B = \begin{bmatrix} 10110010 & 1 \\ 00100110 & 2 \\ . & . \\ . & . \\ . & . \end{bmatrix}$$

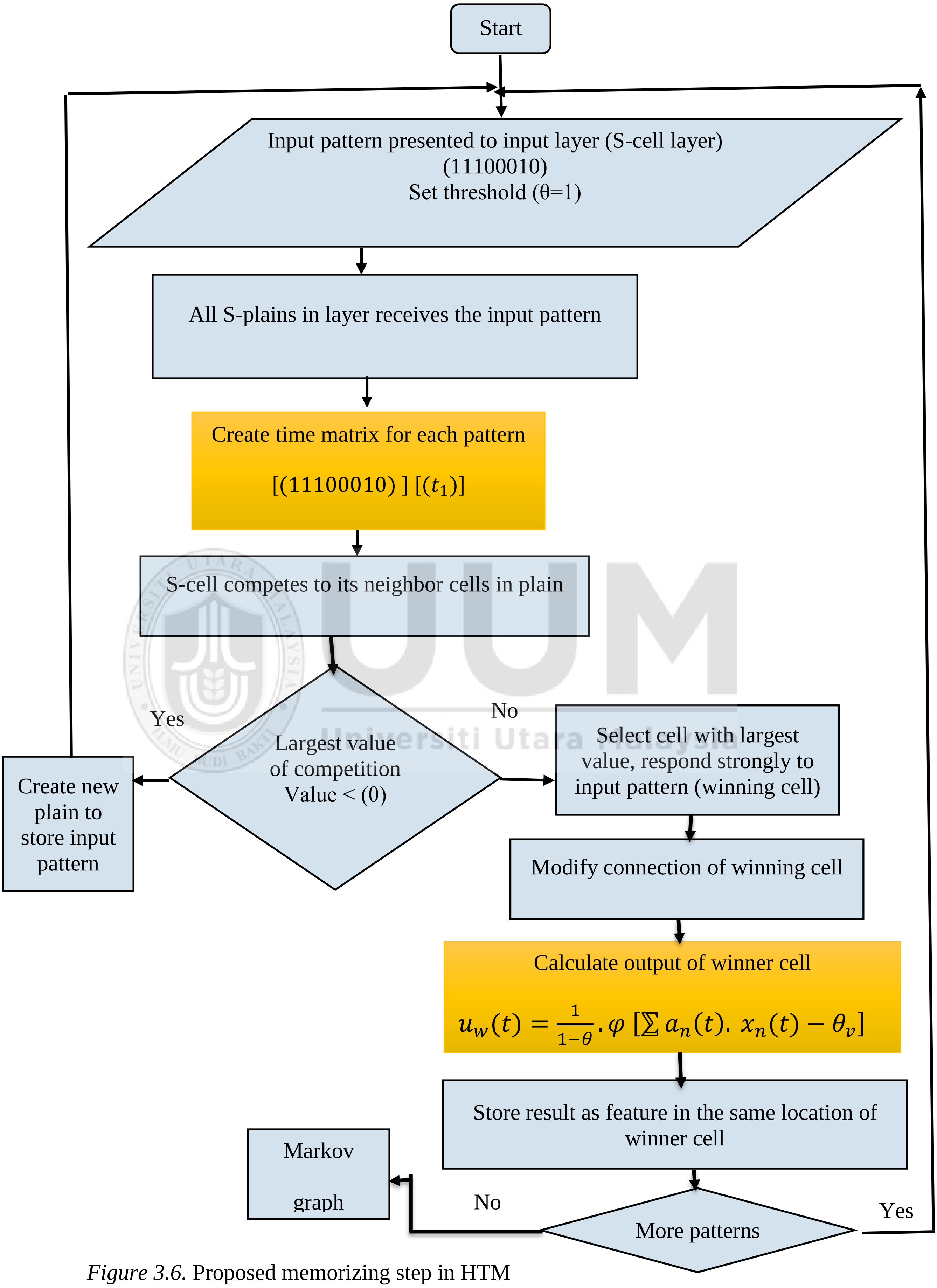
Output of S-cell becomes:

$$u_w(t) = \frac{1}{1-\theta} . \varphi \left[\sum a_n(t).\ x_n(t) - \theta_v\right] \tag{3.1}$$

Where $\boldsymbol{u_w(t)}$ is the output of s-cell for time (t) for the input φ[] is a function defined by $\varphi[x]$=max$(x, 0)$, the strength of the inhibitory connection is θ, which determines the threshold of the S-cell (0 ≤ θ < 1). $\boldsymbol{a_n(t),}$ is the strength of variable excitatory connection coming from the preceding stage at time t, $\boldsymbol{x_n(t)}$ is the current input.

In the previous HTMs, (as mentioned in Chapter two section 2.3), memorizing step is time consuming (Jasmin Léveillé et al., et al.,2014), and incorrect feature extraction may occur because any change in input pattern, for example change in pixel position due to imaging process, viewpoint change, or large interclass variations among the patterns, will cause incorrect matching with the stored patterns. This in turn leads to extraction of incorrect features, furthermore leading to incorrect recognizing of input pattern (Abdel-Azim, et al.,2016; Kalmar & Vida, et al.,2013) . A simple cell in the neural network remembers a configuration of features from the input level, while an HTM node remembers the configuration of the Markov chain "features". This means that the features are extracted at Hidden Markov graph step, while in this research HTM extracts features in the feature extraction step by using improved Neocognitron neural network. The Markov step here will be used to support the previous feature extraction step necessary to increase the accuracy, this will be done by using Markov chain to relates each feature to its pattern, that is each vertex in Markov graph represents one feature and the links between these vertices represent the transitions between patterns in time, it will be used as similarity factor between features. The features will be grouped together into two steps, first by using Markov chain then by using clustering algorithm in the next step of HTM.

Universiti Utara Malaysia

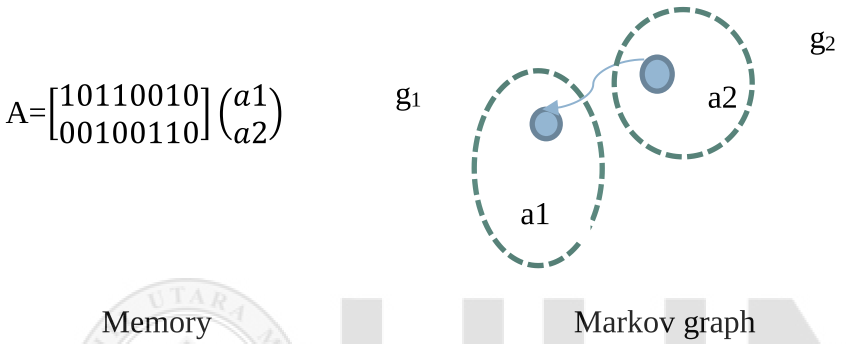

Start
Input pattern presented to input layer (S-cell layer)
(11100010)
Set threshold (θ=1)
All S-plains in layer receives the input pattern
Create time matrix for each pattern
[(11100010)] [(t_1)]
S-cell competes to its neighbor cells in plain
Largest value of competition Value < (θ)
Yes
No
Create new plain to store input pattern
Select cell with largest value, respond strongly to input pattern (winning cell)
Modify connection of winning cell
Calculate output of winner cell
$u_w(t) = \frac{1}{1-\theta}.\varphi\, [\sum a_n(t).\, x_n(t) - \theta_v]$
Store result as feature in the same location of winner cell
More patterns
No
Yes
Markov graph

UUM Universiti Utara Malaysia

Figure 3.6. Proposed memorizing step in HTM

3.2.3 Clustering in HTM

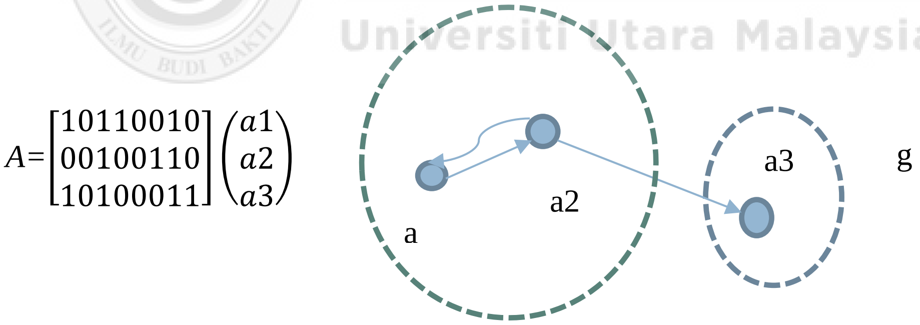
The winner cell u_w of the S-cell layer is combined with time matrix, T, and sent to the creation of Markov graph as shown in Figure 3.7. As mentioned in chapter two, building a Markov graph requires the creation of groups of input vectors depending on probability of following input vectors to each other's. the input pattern is assumed as matrix, A, has input vectors as described in Figure 3.8.



$$A=\begin{bmatrix}10110010\\00100110\end{bmatrix}\begin{pmatrix}a1\\a2\end{pmatrix}$$

Figure 3.7. Operation of building Markov graph

UUM
Universiti Utara Malaysia



$$A=\begin{bmatrix}10110010\\00100110\\10100011\end{bmatrix}\begin{pmatrix}a1\\a2\\a3\end{pmatrix}$$

Figure 3.8. How a new node is added

In Figure 3.8 another pattern is added to the node memory, this will make the total number of patterns equal to three. This will cause a new vertex to be added to the Markov graph, which corresponds to the new pattern. The step will produce two temporal groups, with two vertices in one group and one vertex in the other. The data obtained from Markov graph includes the probability of transition between two patterns. This probability is used as a similarity measure between patterns for clustering algorithm to create clusters known as temporal groups.

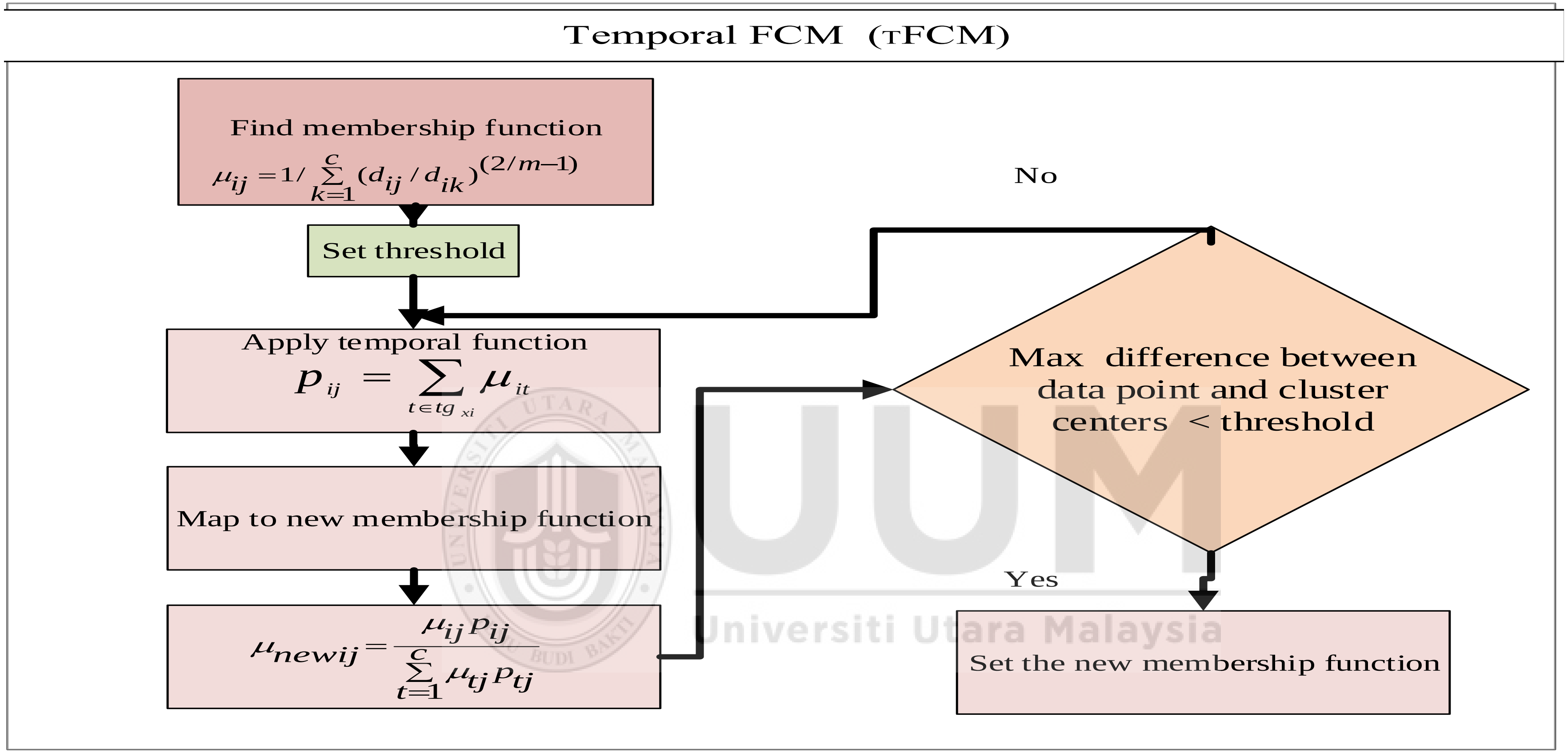
The clustering algorithm used is FCM, it is a method of clustering which allows one piece of data to belong to two or more clusters, it is frequently used in pattern recognition as mentioned in Chapter two (section 2.2.3.1). The standard FCM algorithm used in existing HTM does not utilize the temporal relationship contained in data, as mentioned in Chapter two, section 2.3. This means that the neighboring pixels in time are highly correlated, hence possesses same feature value, means that the probability that they belong to the same cluster is great. This temporal relationship is important in clustering. To exploit temporal information, a temporal function is defined as follows:

$$p_{ij} = \sum_{t \in tg_{x_i}} \mu_{it} \qquad (3.2)$$

$$\mu_{new_{ij}} = \frac{\mu_{ij} p_{ij}}{\sum_{t=1}^{c} \mu_{tj} p_{tj}} \qquad (3.3)$$

Where $\boldsymbol{p_{ij}}$ in Eq.13**,** is the probability of that the pixel $\boldsymbol{x_i}$ belongs to the ith cluster, $\boldsymbol{tg}$ is the square matrix centered on xi in the time domain, $\boldsymbol{\mu_{it}}$ represents the membership of pixel of time *t* in the *i*th cluster. In Eq.14, *c* is the number of cluster centers. The clustering start at applying the standard FCM to calculate the membership function, then the temporal function is calculated as in Eq.13, 14. Figure 3.9 illustrates the improved FCM.

Universiti Utara Malaysia


Temporal FCM (τFCM)
Find membership function
$\mu_{ij} = 1/\sum_{k=1}^{c}(d_{ij}/d_{ik})^{(2/m-1)}$
Set threshold
Apply temporal function
$p_{ij} = \sum_{t \in tg_{xi}} \mu_{it}$
Map to new membership function
$\mu_{newij} = \frac{\mu_{ij} p_{ij}}{\sum_{t=1}^{c} \mu_{tj} p_{tj}}$
Max difference between data point and cluster centers < threshold
No
Yes
Set the new membership function

Universiti Utara Malaysia

Figure 3.9. Proposed improved FCM steps

Each cluster contains a set of patterns. These patterns follow each other in term of time manner. The Fuzzy c-Means is a clustering method that permits one data item to be related to more than one cluster.

3.2.4 Distance Measurement for Clustering in HTM

A metric or distance function is a function which defines a distance between the elements of a set. A set with a metric is called a metric space(Kahkashan Kouser1 & Sunita, et al.,2013). In clustering, measuring similarity between objects has become an important part. Normally, the task is to define a function Sim *(X, Y)*, where X and Y are two objects or sets of a certain class, and the value of the function represents the degree of "similarity" between the two. So the use of the suitable metric is important, there are several distance metrics like Euclidean, Manhattan, Chebyshev and correlation coefficient, Table 3.2 gives the distance measures and their applications (Grabusts, et al.,2011).

Universiti Utara Malaysia

As mentioned in Chapter Two (section 2.3), the Chebyshev distance metric is suitable for the FCM clustering algorithm. Hence it is used with FCM instead of Euclidean distance (Ng, Li, Huang, & He, et al.,2007). In this research, the Chebyshev distance has been improved to avoid the infinity, which causes the value of distance to become larger, a data point far away from its specific cluster, this is done by using the limit. The limit is used in the two basic problems of calculus: the tangent line problem and the area problem (Horvath & Khoshnevisan, et al.,2010). The expression $\lim_{p \to \infty} f(p) = L$ indicates that if the value of p is close but not equal to ∞, then $f(p)$ will be close to the value L; moreover, $f(p)$ gets closer and closer to L as p gets closer and closer to ∞ (Guichard., et al.,2015).

For $\lim_{p \to \infty} f(p) = L^{1/p}$ means that $1/p$ approachs 0 as p approaches infinity (Heinbockel, et al.,2012), which minimizes the distance value.

The goal of improving the distance metric is to minimize the distance between any point and cluster center of the class. This requires the value of the distance metric to be close to zero. The distance between two points is computed as follows:

$$Dxy{=}lim\ (max_k(|x_{ik} - x_{jk}|)^p\ ) \quad \infty < \ p > 2 \qquad (3.4)$$

The following example is done by applying the improved HTM. The dataset used in this example is blood type image example:

Figure 3.10 presents the response of S-layer. Figure 3.11 on the other hand, represents the features that are extracted by S-layer for one pattern.

UUM
Universiti Utara Malaysia

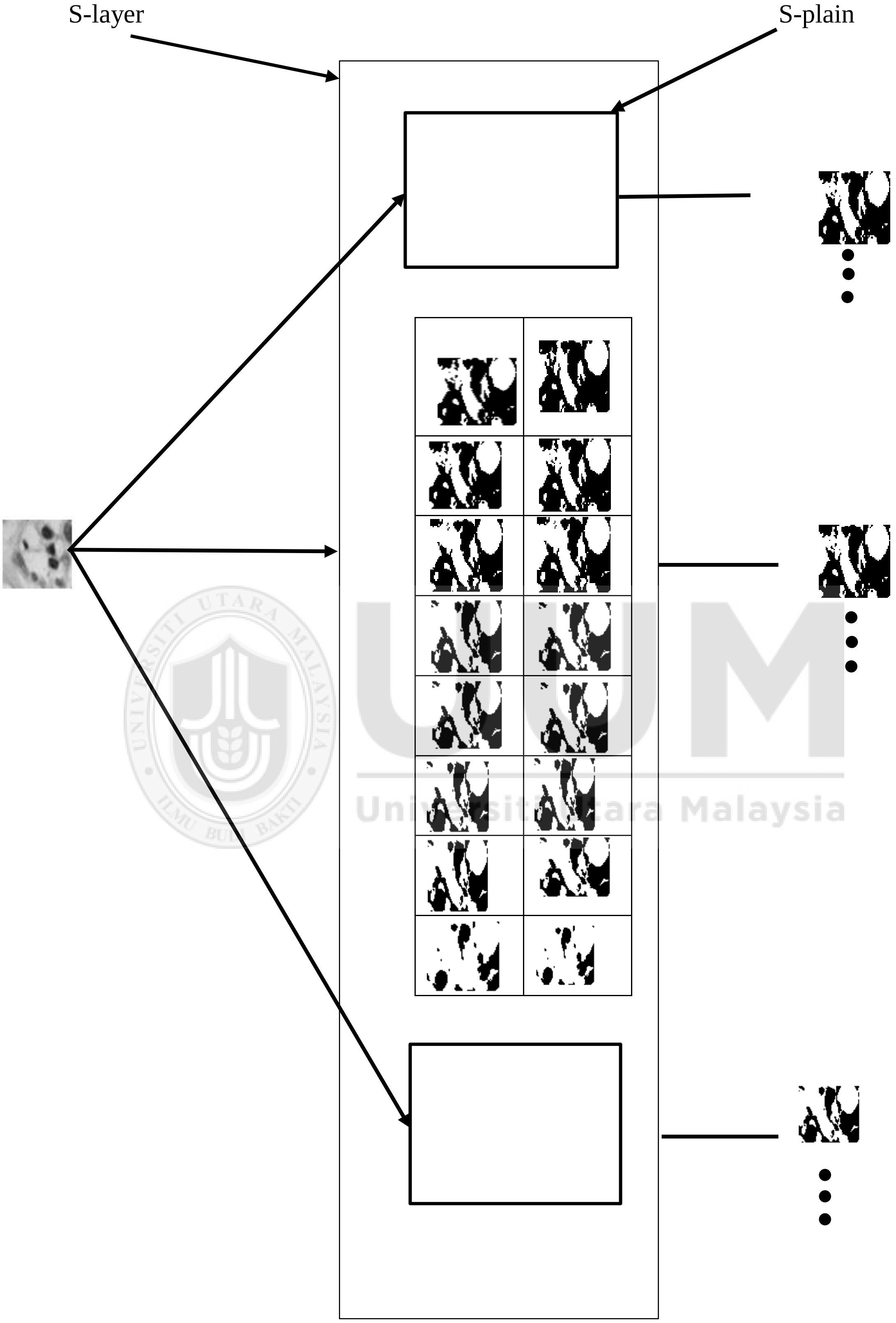


Figure 3.10. Single node features extraction in level-1 HTM

Universiti Utara Malaysia

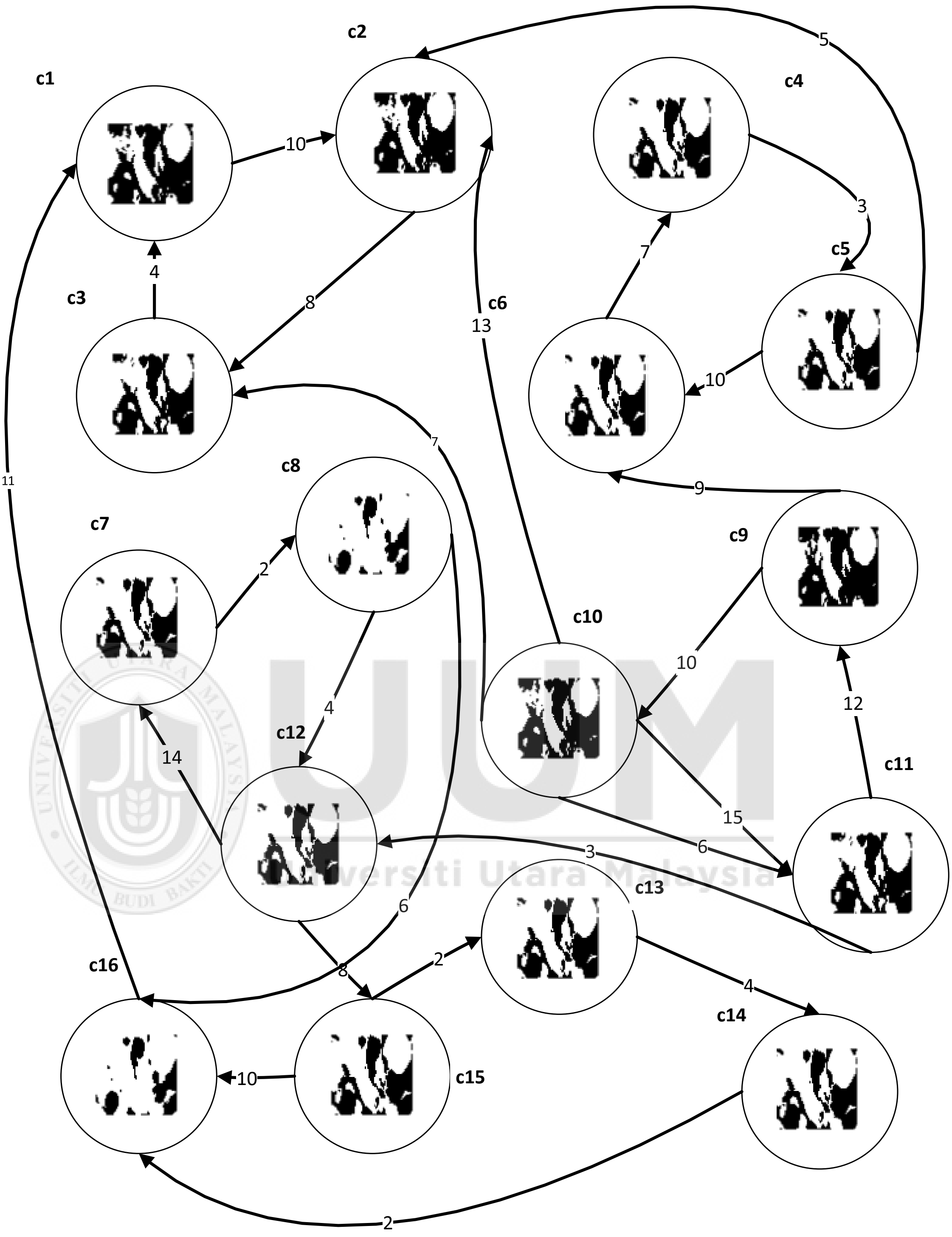
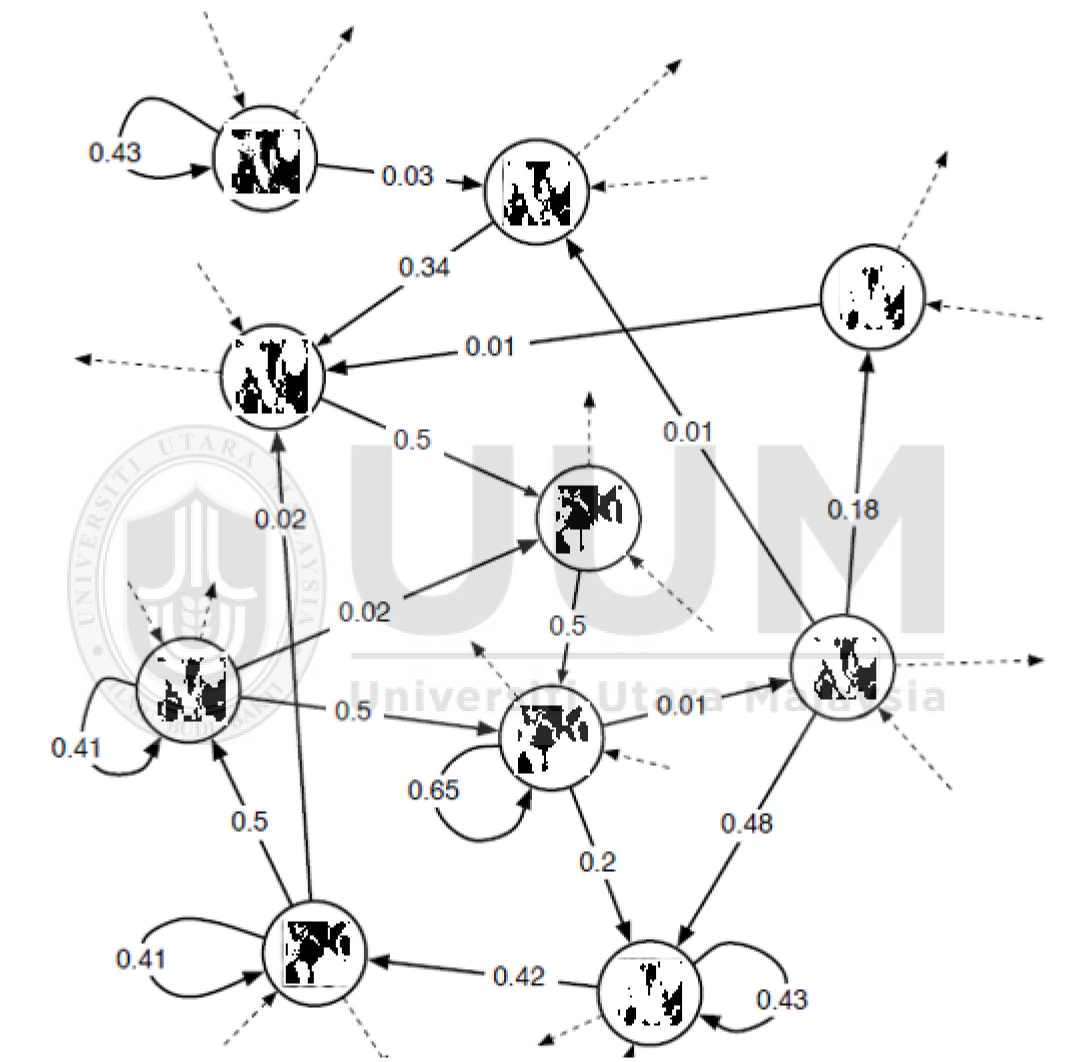


Figure 3.11. Markov graph for the portion pattern of figure 3.10. The vertices are shown as circles, the patterns corresponding to these vertices are shown within the circles as 4x4 pixel arrays.

Universiti Utara Malaysia

Figures 3.12 illustrates Markov graph for the extracted features for one portion pattern respectively. Figure 3.13 presents portion of normalized Markov graph for whole image; the original Markov graph has 896 vertices obtained at the end of the training period.



UUM
Universiti Utara Malaysia

Figure 3.12. Portion of the normalized Markov graph learned in a level-1 node. The portion shown here has ten vertices. The vertices are shown as circles. For ease of interpretation, the patterns corresponding to these vertices are shown within the circles as 4x4 pixel arrays.

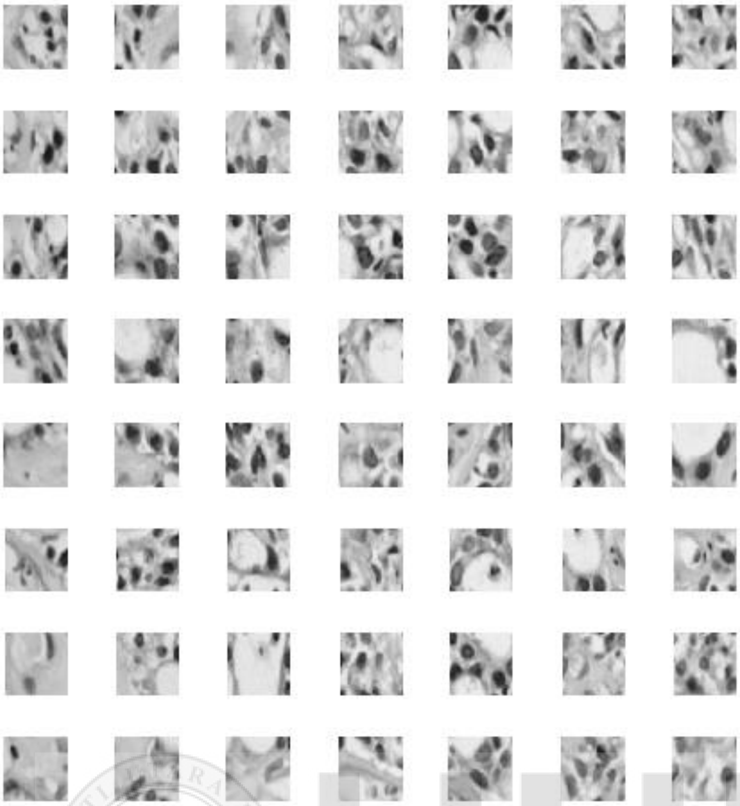


Figure 3.13. Patterns memorized by a level-1 HTM of figure 7.1. Level-1 memorize 56 patterns from 64 input patterns, while Figure 3.14. gives the temporal groups resulted from this step.

UUM
Universiti Utara Malaysia

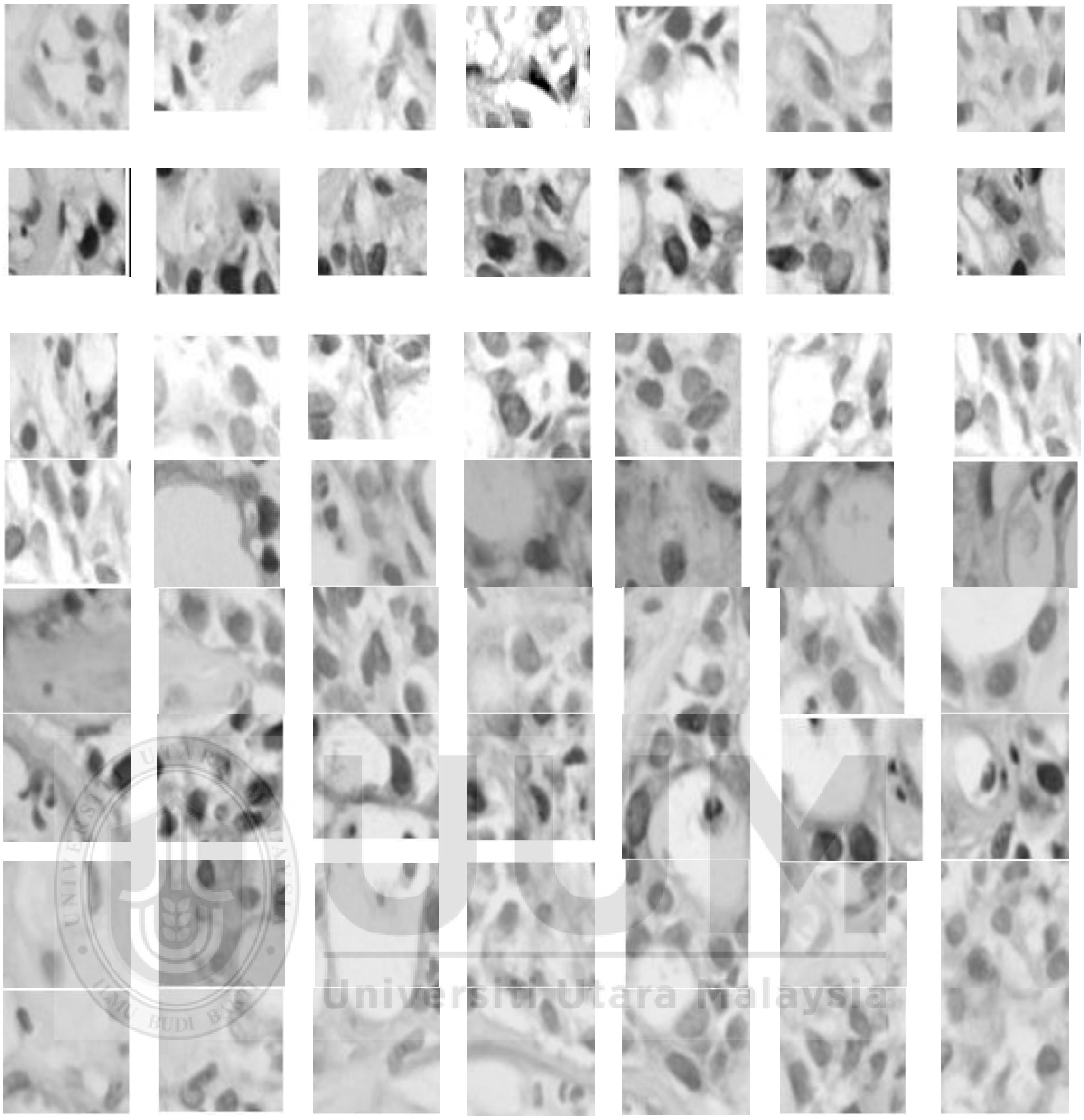


Figure 3.14. Final temporal groups obtained from $_T$FCM.

Universiti Utara Malaysia

3.3 Development

Original Hierarchical Temporal Memory (HTM), Neocognitron neural network, Principal Component Analysis (PCA), Sum of Absolute Difference (SAD) and original Fuzzy c-Means (FCM) with Euclidean Manhattan Minkowski Chebyshev distances, improved S-layer Neocognitron, and improved FCM with improved Chebyshev distance have been implemented on the same dataset for the purpose of comparison with the proposed model. The implementation is carried out using MATLAB V.13 on a Dell Inspiron Core i -5 laptop

3.4 Evaluation

This section includes information on how the proposed HTM is evaluated. It is divided into two parts; comparison methods and the performance measurement. These evaluations are conducted as follows: **Comparison Methods**

Universiti Utara Malaysia

Evaluation of the proposed HTM model will be made by comparing its performance against the traditional pattern recognition methods to see the effectivity of proposed method in recognition process, also comparison made against FCM clustering with four distance metrices under consideration and the proposed distance metric to find the powerful of clustering part of the proposed method against the original selected clustering algorithm:

3.4.1.1 Statistical Techniques

Comparisons between improved HTM, Neocognitron Neural Network, Principal Component Analysis (PCA) and Sum of Absolute Difference (SAD) are as reported in Chapter Four.

3.4.1.2 Structural Techniques

Comparisons between improved HTM and original HTM, and Hidden Markov model are as reported in Chapter Seven.

3.4.1.3 Clustering Techniques

Comparisons between the improved HTM (${}_{T}$FCM) and Fuzzy C_Means clustering algorithm (FCM) are as reported in Chapter Five.

Comparison between improved HTM (${}_{T}$FCM) and Fuzzy C_Means clustering algorithm (FCM) with Euclidean, Manhattan, Minkowski Chebyshev and Improved Chebyshev (limit Chebyshev) distances are as reported in Chapter Six.

3.4.2 Performance Measurements

The metrics involved in the evaluation include the ones for classification and clustering.

3.4.2.1 Metric for Classification

Rankings-based and label- based classification metrics are used. The Hamming Loss, accuracy, F1 score, Subset accuracy are rankings-based, while Macro-F1 and Micro-F are label based.

Hamming Loss: refers to loss function which calculates the percentage of misclassified labels to the total number of labels (Wong, et al.,2017) and is given by the formula illustrated in Equation. 3.5:

$$Hamming\ Loss = \frac{1}{K}\sum_{i=1}^{K}\frac{predicted_i\ \Delta True_i}{L} \tag{3.5}$$

where K is the number of examples/instances, L is the number of classes, Δ is the XOR operation or difference between the two sets $predicted_i$and $True_i$, while $predicted_i$ is the prediction for instance i and $True_i$ is the true/target value for instance i. Since it is a loss function, the smaller the value, the better the performance of the algorithm.

Universiti Utara Malaysia

Accuracy is the number of correct predictions made as a ratio of all predictions made (Azad et al., et al.,2014). The formula is in Equation 3.6 as follows:

$$\text{ACC} = \frac{N_c}{N_t} \quad (3.6)$$

where N_c Is number of correctly classified testing examples

N_t Is number of testing examples

Subset Accuracy: is a strict accuracy measurement which consider a classification as correct if and only if all the labels have been correctly classified (Kafrawy, Mausad, & Esmail, et al.,2015a). It is given in Equation 3.7:

$$\text{Subset Accuracy}= \frac{1}{K}\sum_{i=1}^{K} predicted_i = True_i \quad (3.7)$$

where $predicted_i$ is precision for classifier, K is the number of classifiers.

F1 score: is a harmonic mean of Precision and Recall (Taha & Hanbury, et al.,2015). The F1score is given as follows:

Universiti Utara Malaysia

$$\text{F1score}= \frac{1}{K}\sum_{i=1}^{N} predicted_i * Recall_i \quad (3.8)$$

Where $predicted_i$ is precision for classifier *i* and $Recall_i$is recall for classifier *i* and N is the number of classifiers.

Macro-F1 calculates the F1 for each of the *m* labels and averages them (Wong et al., et al.,2017)**,** Equation 3.9 illustrates the Macro-F1:

$$\text{F1}_{macro}(P, G) = \frac{1}{m}\sum_{1}^{m} F1(P_{ij},\ G_{ij}) \quad (3.9)$$

Micro- F is the harmonic mean between micro-precision and micro-recall (Kafrawy et al., et al.,2015b).

Micro-F1 is defined as Equation 3.10:

$$\text{F1 micro} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n} Y_i^j Z_i^j}{\sum_{j=1}^{k}\sum_{i=1}^{n} Y_i^j + \sum_{j=1}^{k}\sum_{i=1}^{n} Z_i^j} \qquad (3.10)$$

3.4.2.2 Metrics for Clustering

Five measures are used in this research to evaluate clustering performance; purity, entropy, minimum objective function, iteration count and average distance.

Purity is measured by counting the number of correctly assigned image and dividing by N (N is the number of clusters) (Deepa & Revathy, et al.,2012) as follows:

$$purity\ (\Omega, \mathbb{C}) = \frac{1}{N}\sum_k \max_j \left|w_k \cap c_j\right| \qquad (3.11)$$

$\Omega = \{w1, w2, \ldots, w_k\}$ is the set of clusters and

Universiti Utara Malaysia

$\mathbb{C} = \{c1, c2, \ldots, c_j\}$ is the set of classes.

ωk is cluster k, and cj is class j.

Bad clustering has purity values close to 0, where a perfect clustering has a purity of 1

Entropy is the measure of randomness and good quality of clusters (Das & De, et al.,2016). The formula is given as follows:

$$\boldsymbol{H}(\Omega) \underline{= - \sum_k \boldsymbol{p}(\boldsymbol{w_k}) \log \boldsymbol{p}(\boldsymbol{w_k})} \qquad (3.12)$$

$$\underline{= \sum_k \frac{|w_k|}{N}\ log \frac{|w_k|}{N}}$$

H is the symbol for entropy, Ω is a vector of zero-indexed symbols, and P means "probability of." The log2 function (log to base 2) assumes that log2 (0) = 0.0 rather than the true value of negative infinity.

If the purity is high and the entropy value is low, this tends to indicate a good clustering. Other values indicate bad clustering.

An Objective Function is a function such that the clusters are obtained when the minimum/maximum is reached are homogeneous. This allows the choice of good set of features and the appropriate number of clusters to generate a good partition of the data (Y.-T. Chen, et al.,2017). Equation 3.13 gives the formula.

$$J(u,v) = \sum_{i=1}^{n} \sum_{j=1}^{c} \mu_{ij}{}^{m} \left\| x_i - v_j \right\|^{q} \tag{3.13}$$

where $\boldsymbol{m}$ is defined to any real number that greater than 1, $\boldsymbol{\mu_{ij}}$ is refer to a degree of membership of xj in cluster j.

Universiti Utara Malaysia

Iteration Count is based on minimizing an objective function that represents the distance from any given data point to a cluster center weighted by that data point's membership grade (Rammal et al., et al.,2017).

Average Distance is the distance between all pairs of subjects in the two clusters. It is the evaluation of compactness of clustering output, the smaller the value of average distance the more compact output (Abdel-Azim, et al.,2016). The formula is illustrated in Eq. 3.14

$$AD = \sum_{j=1}^{K} \frac{\frac{\sum_{i=1}^{ni} D(Oi,Di)}{ni}}{K} \tag{3.14}$$

where, K is the number of clusters, *ni* is the number of data points in cluster *i*, *Oi* the center of cluster I, and D is distance.

3.5 Summary

In this chapter, the steps to develop the proposed HTM model is presented. Key points of the model are to overcome the shift in position and overlapped data found in medical images. To achieve this, two steps of HTM are improved upon. First, the classification step uses an improved S-cell layer Neocognitron neural network as feature extraction step of HTM. This is done by adding the time factor for further processing in higher level of HTM. The aim is to overcome the shift in position found in input patterns.

Second, the clustering step of HTM is improved by using an improved FCM clustering algorithm to handle the temporal relationship of the data. The improved FCM clustering algorithm will use an improved Chebyshev distance metric to handle these two improvements and improve upon noisy medical images and multi- label datasets. The goal is to obtain more accurate clusters to reach the optimum recognition process.

Universiti Utara Malaysia

The proposed HTM algorithm is later compared against state-of-art pattern recognition algorithms, such as Sum of Absolute Difference (SAD) method of template matching, Principal Component Analysis (PCA) statistical approach and Hidden Markov Model (HMM) structural model. The evaluation of the proposed HTM is based on performance metric, which are classification and cluster.

CHAPTER FOUR
FEATURE EXTRACTION USING TIME- S-LAYER NEOCOGNITRON

4.1 Introduction

As mentioned previously (see Chapter One, section 1.2) original HTM suffers from shift in position of input data. Such a drawback leads to an unknown pattern being recognized as known object. To ensure the accuracy of HTM, the memorizing step (feature extraction) is redesigned using improved Neocognitron neural network. The design includes improvement of the first S-layer Neocognitron, which is the feature extraction layer, to make this step more accurate.

Universiti Utara Malaysia

A neocognitron neural network is a spatial approach to classify a pattern; hence, it will not reflect the process of recognition in human brain which is spatial and temporal. Hence, temporal factor is added (i.e. time factor) to the S-layer output function in order to use this factor as a similarity measure in feature extraction step as mentioned in Chapter Three (section 3.2.1).

The algorithm of improved feature extraction is illustrated in Figure 4.1.

Improved Feature Extraction step

```
Input stimuli of size 4x4 data part is presented to S-layer cells, each plane receives the same input connection
time matrix of same dimension of input stimuli is created. Each element of the time matrix corresponds to one input vector
Input is compared with the S-cells content according to threshold (θ =1)
  **If identical then**
    Begin
      For each identical cell
        Strengthening parameters of s-cell, and output of identical cells (Us)

          $v_j = (\sum_{i=1}^{n}(\mu_{ij})^m u_W(t)_i)/(\sum_{i=1}^{n}(\mu_{ij})^m), \ \forall_j = 1, 2, 3 \ldots .c$

          $\theta$ is the threshold of the S-cell and determines the selectivity in extracting features
    **end**
  **Else**
    Generate new plain and store input pattern in new cell in this plain
**end if**
```

Universiti Utara Malaysia

Figure 4.1. Feature extraction step of improved HTM

4.2 Time S-layer Neocognitron Step

This section presents the results obtained by applying the time S-layer Neocognitron algorithm ($_T$S-layer) on five multi-label datasets mentioned in Chapter Three (section

3.2.1). The process which is taking place, is extracting features from input pattern. In this step, when the input pattern is presented to the S-layer (level-1 of HTM), it is compared with each cell in the cell- plain, all cells in cell plain receives the same input, S-layer has 16 cell-planes, each plain extract a specific feature depending on the response of S- cells in the plain. Output of the response of an S-cell takes a maximum value approximately equal to 1 when the input vector is identical to the reference (stored) vector and becomes 0 if the similarity s is less than the threshold (θl) of the cell. In the multi-dimensional feature space, the area that satisfies s < (θl) becomes the tolerance area in feature extraction by the S-cell, and the threshold (θl) determines the radius of the tolerance area. If the threshold is low, the radius of the tolerance area becomes large, and the S-cell responds even to features largely deformed from the stored vector (shift in position). The selectivity of an S-cell to its preferred feature (or the stored vector) can thus be controlled by the threshold (θl). If two or more S-cells becomes winner cells (has the largest output values) appear in a single S-plane, only the one which is yielding the largest output among them is selected as the representative from that S-plane. In a case in which only one winner appears in an S-plane, the cell is unconditionally determined as the representative from that S-plane. If no winner cell appears in an S-plane, no representative is selected from that S-plane.

Since the representative cells are determined in this manner, each S-plane becomes selectively sensitive to one of the features of the input patterns, and there is not a possibility of formation of redundant connections such that two or more S-planes are used for detection of one and the same feature. If the pattern feature is not in the memory (If all S-cells in a competition area are silent), a new cell-plain will be created in S-layer to store the feature extracted from the new input pattern. As mentioned previously,

Universiti Utara Malaysia

each input pattern is a 16-length vector for 4x4 input portion pattern, so it has 16 features. The features that are follow each other in time are in the same pattern.

4.2.1 Dataset 1

In this section, a blood type images are used. 3000 images each of size 64 x 64 pixels, 2500 for training and 500 for testing are used. Three example images of different densities will be demonstrated in experiment illustration to see the difference in performances. Comparison results are obtained between the original and improved Neocognitron neural network and between level-1 original HTM as mentioned in Chapter Three section 3.4.1.

4.2.1.1 $_T$S-layer vs. Neocognitron

This section presents the acomparison of results is obtained between the time S-layer algorithm ($_T$S-layer) in HTM and original Neocognitron. Figure 4.2 shows a part of the portioned image into 4x4 pixel, each as mentioned in step 1 in Figure 4.1. Table 4.1 illustrates the number of features that are extracted from the input pattern and Accuracy for $_T$S-layer HTM and Original S-layer Neocognitron. From this table, it is seen that the number of features extracted from the input pattern using $_T$S-layer is more than the features extracted by the original one, which can significantly increase classification accuracy. Feature extraction technique is a major factor affecting the accuracy of the recognition process (Pechenizkiy, et al.,2005). Figure 4.3 graphically explains the comparison between original Neocognitron and $_T$S-layer HTM, with the original marked with a black line and improved one which is marked with green line. It is obvious that the deference increases between the two lines when the number of input patterns increase, which assigns good extraction when the number of input patterns

Universiti Utara Malaysia

increases. From the tables and figures, it may be seen that original Neocognitron can extract 60% of the features, while $_T$S-layer HTM can extract 90%.

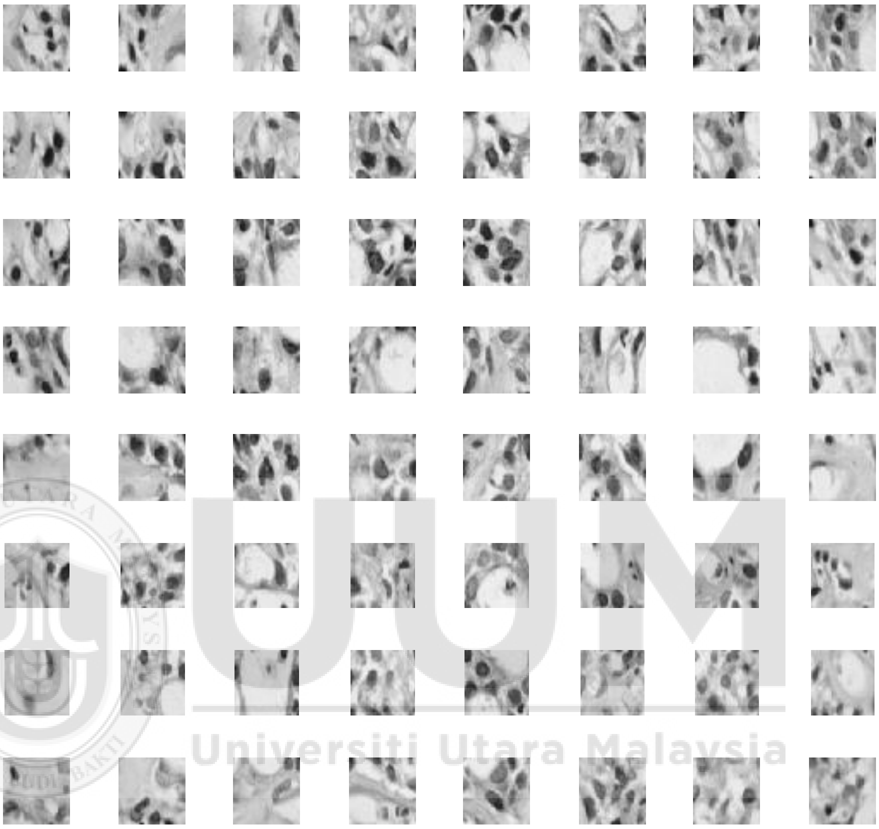


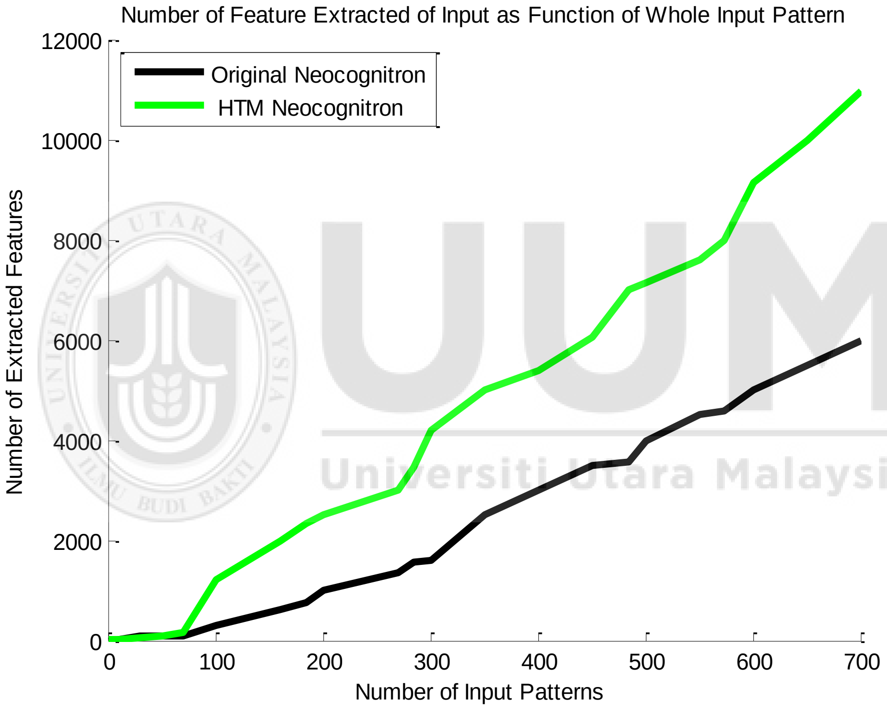
Figure 4.2. Patterns results from dividing image. Each pattern is of length 16 and is shown as a 4x4 pixel array.

Universiti Utara Malaysia

Table 4.1

Feature extracted of dataset 1: ${}_{T}$S-layer HTM vs. original Neocognitron

No. of input pattern	Extracted features in %	
	Original Neocognitron	${}_{T}$S-layer HTM
100	56	90.1
200	60.6	80.7
300	59.8	87.5
400	58.5	94.6
500	61.5	94.75
600	58.7	95.3
700	56.4	95



UUM
Universiti Utara Malaysia

Figure 4.3. Number of features extracted as function of input patterns, dataset 1: ${}_{T}$S-layer HTM vs. original Neocognitron.

4.2.1.2 Accuracy and Feature Numbers for ${}_{T}$S-layer HTM vs. Original HTM

Table 4.2 illustrates the number of feature extracted from the input pattern by improved level-1 HTM and original HTM. From this table, it is seen that the improved HTM is better than the original one in extraction features from input pattern by about 47%, it can extract 90% of input pattern features, while original HTM can extract about 43%.

This affects the recognition accuracy, because more extracting features allow more accurate recognition (Pechenizkiy, et al.,2005). Figure 4.4 graphically explains the comparison between the improved HTM with improved S-layer Neocognitron and original HTM in extracting features. The table also illustrates the difference in accuracy value between original HTM and improved HTM with improved S-layer Neocognitron. The table demonstrates that the improved HTM has higher accuracy than the original one, which enhances performance in important applications.

Table 4.2

Feature extracted and accuracy, dataset 1: $_T$S-layer HTM and original HTM

Number of Input Patterns	Features Extracted %		Accuracy in %	
	Original HTM	$_T$S-layer HTM	Original HTM	$_T$S-layer HTM
100	42.6	90.1	80	95
200	39	80.7	81	95.3
300	41.6	87.5	82	95.9
400	42.1	94.6	84	96.1
500	40.6	94.75	84.3	96.4
600	41.6	95.3	85	96.8
700	51.4	95	85.1	98

UUM
Universiti Utara Malaysia

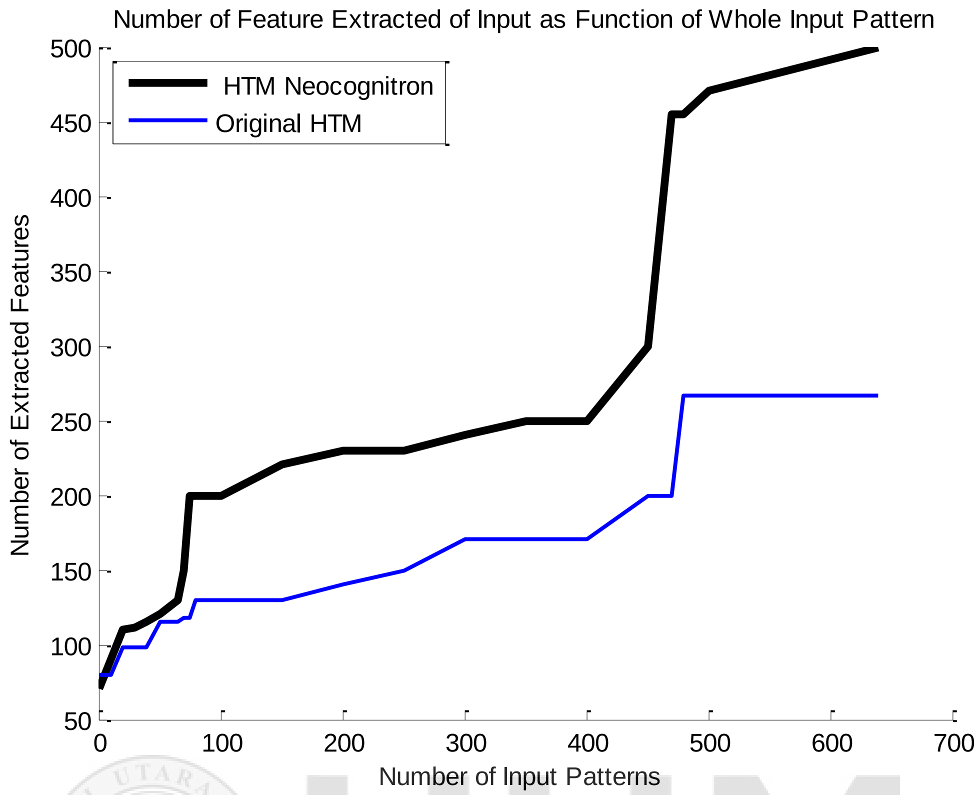


Figure 4.4. Comparison between $_{T}$S-layer HTM and original HTM

Universiti Utara Malaysia

4.2.2 Dataset 2

In this experiment, blood cancer images are used. As mentioned in Chapter one (section 3.2.1.2), 108 peripheral blood smear images are taken of size 32 x 32 pixels for training and testing, 75 for training and 33 for testing. Comparison results are obtained between original, improved Neocognitron neural network and level-1 HTM as mentioned in chapter 3 section 3.4.1.

4.2.2.1 $_{T}$S-layer HTM vs. Neocognitron

This section presents the A comparison of results is obtained between the $_{T}$S-layer HTM and original Neocognitron. Figure 4.5 shows the image portioned into 4x4 pixels for each portion image as mentioned in step 1 in Figure 4.1.

Table 4.3 shows the number of features that are extracted from each input pattern by original Neocognitron and $_T$S-layer. From this table, it is learned that the number of features extracted from the input pattern by using $_T$S-layer is greater than the features extracted by the original one. Figure 4.6, on the other hand, shows the graphical presentation of this table, from figure and table, it was learned that the original Neocognitron which is marked with black line can extract a small number of features (74%) compared with the number of features extracted by the $_T$S-layer which marked with green line (97%), however the difference between the two lines is not large in terms of the size of image.

4.2.2.2 $_T$S-layer HTM vs. Original HTM

Table 4.4 illustrates the number of features extracted from the input pattern by improved level-1 HTM and original HTM and Accuracy value. From this table, it is seen that the HTM with improved level-1 can extract features by percent about 94% from the whole features, while it can extract about 54% by using the original one. Accuracy of HTM with improved level-1 is higher than original one by about 13 %. Figure 4.7, on the other hand explains graphically the comparison between the improved HTM with improve level -1 which is marked with black line and original HTM which is market with blue line in extracting features. From this figure, it is obvious that the difference between the two lines is not so high.

Universiti Utara Malaysia

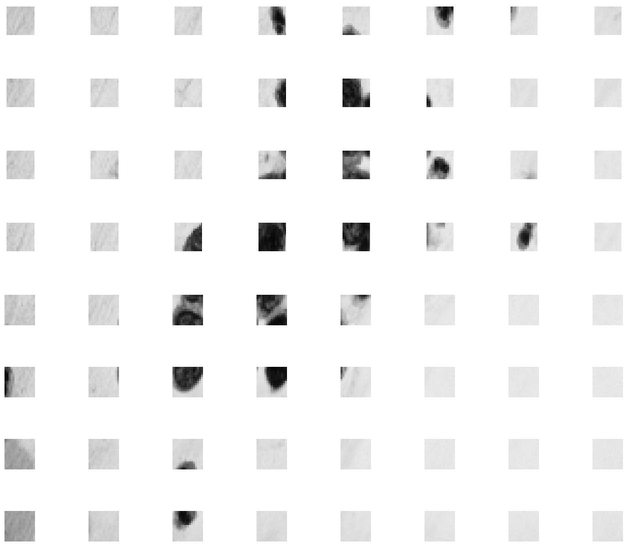


Figure 4.5. Patterns results from dividing the image. Each pattern is of length 16 and is shown as a 4x4 Pixel Array.

Table 4.3

Feature extracted of dataset 2: $_T$S-layer HTM vs. original Neocognitron

Number of input pattern	Feature extracted %	
	Original Neocognitron	$_T$S-layer HTM
10	50	90.1
40	78.1	80.7
70	71.4	87.5
100	77.5	94.6
130	84.6	94.75
160	73.2	95.3
190	84.5	95

Universiti Utara Malaysia

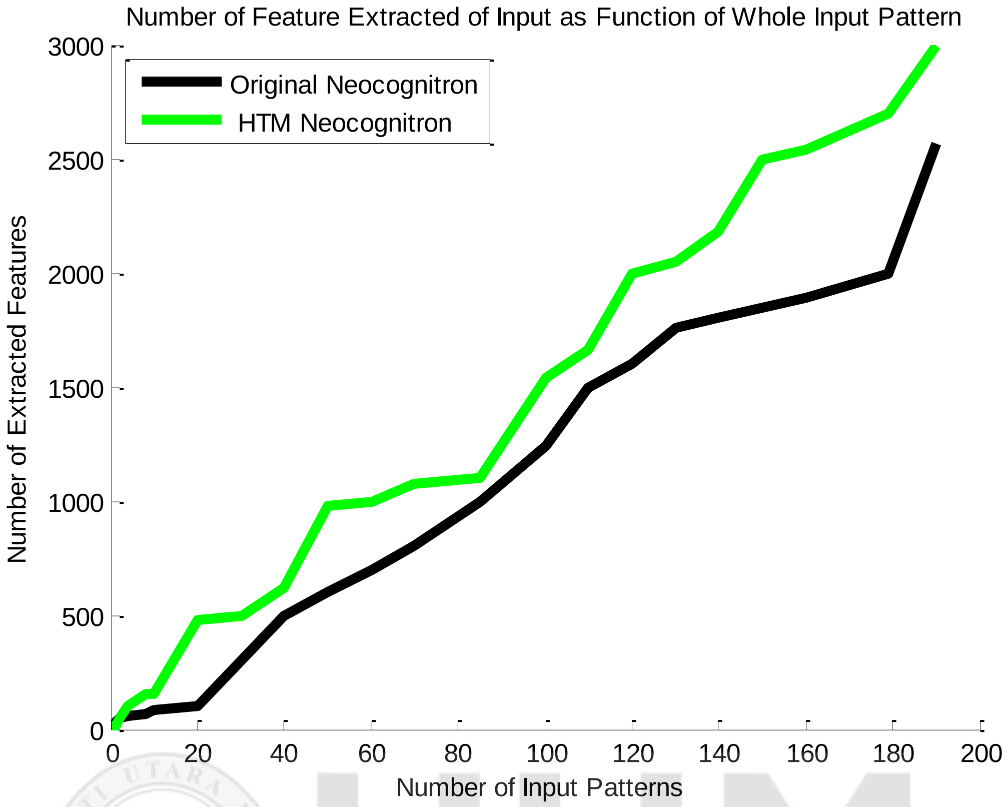


Figure 4.6. Number of features extracted as function of input patterns, dataset 2: $_T$S-layer HTM vs. original Neocognitron

Universiti Utara Malaysia

Table 4.4

Feature extracted of dataset 2: $_T$S-layer and original HTM

Number of Input Patterns	Features extracted %		Accuracy %	
	Original HTM	$_T$S-layer HTM	Original HTM	$_T$S-layer HTM
10	42.5	90.1	81	94.8
40	36.7	80.7	82	95
70	70.4	87.5	82.4	95.1
100	66	94.6	84.8	95.7
130	53	94.75	85	96
160	69	95.3	85.1	96.2
190	60	95	85.4	97.3

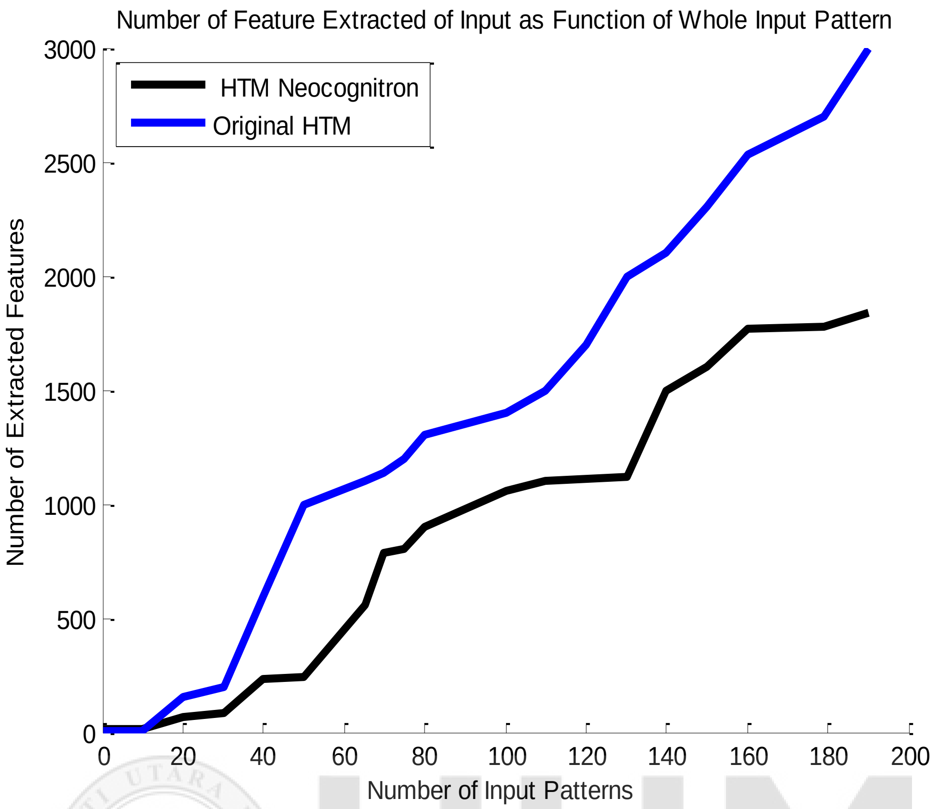


Figure 4.7. Comparison between proposed HTM with $_T$S-layer and original HTM

Universiti Utara Malaysia

4.2.3 Dataset 3

In this experiment, a STARKEY"93 dataset is used. Comparison results are obtained between original and improved Neocognitron neural network and level-1 HTM as mentioned in Chapter Three (section 3.4.1.1). A window of 4x4 data points will be taken from the dataset matrix as input pattern.

4.2.3.1 $_T$S-layer HTM vs. Original Neocognitron

This section presents the A comparison of results is obtained between the $_T$S-layer HTM and original Neocognitron. Table 4.5 shows the number of features that are extracted from each input pattern by original Neocognitron and $_T$S-layer. From this table, it is noticed that the number of features extracted from the input pattern by using $_T$S-layer is more than the features extracted by the original one. Figure 4.8, on the other hand, shows the graphical presentation of this table. From the table and figure, it was learned that the original Neocognitron which is marked with black line in figure can extract features less than which extracted by the $_T$S-layer, which is marked with green line. The features extracted by the original Neocognitron is about 82%, while it is 95% by $_T$S-layer.

Table 4.5

Number of feature extracted dataset 3: $_T$S-layer HTM vs. original S-layer Neocognitron

Number of input pattern	Feature extracted %	
	Original Neocognitron	$_T$S-layer HTM
100	70.8	90.1
200	76	80.7
300	79	87.5
400	81	94.6
500	87	94.75
600	92	95.3
700	88	95

Universiti Utara Malaysia

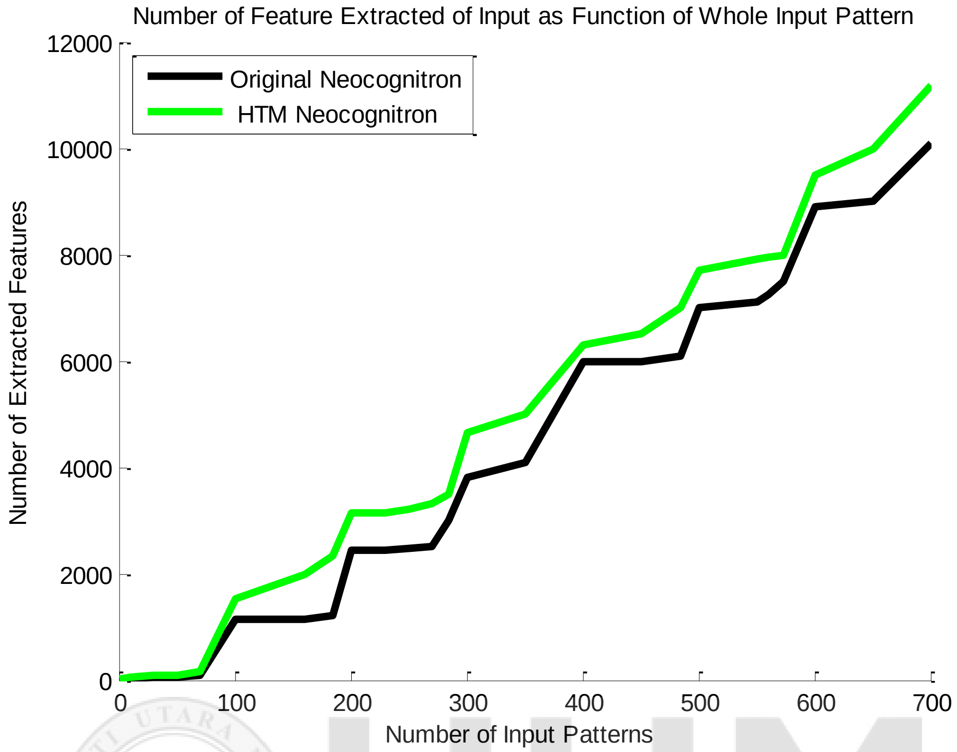


Figure 4.8. Number of features extracted as function of input patterns, dataset 3: proposed S-layer HTM vs. original S-layer Neocognitron

Universiti Utara Malaysia

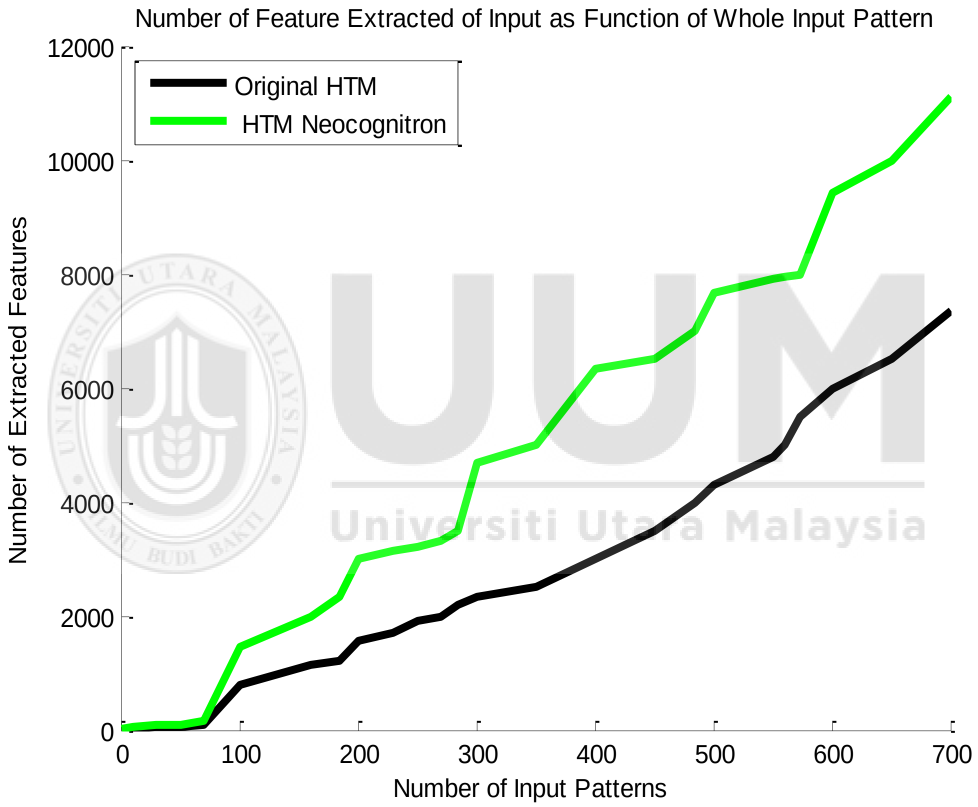
4.2.3.2 $_T$S-layer HTM vs. Original HTM

Table 4.6 illustrates the number of features that are extracted from input pattern and accuracy values by the two algorithms. It is seen that $_T$S-layer HTM has the higher accuracy in recognition patterns than the original one. Figure 4.9 presents the graphical presentation of the table. From the table and figure, it is learned that the $_T$S-layer can extracts features from the input pattern by about 96%, while the original HTM can extract about 57.5%. It is obvious that there is a great difference between the two methods, demonstrating good performance of the proposed method when applied on datasets than on images.

Table 4.6

Number of feature extracted dataset 3: $_{T}$S-layer HTM and original HTM.

Number of Input Patterns	Features Extracted %		Accuracy %	
	Original HTM	$_{T}$S-layer HTM	Original HTM	$_{T}$S-layer HTM
100	53	90.1	80	94.8
200	48.7	80.7	79	95
300	48.7	87.5	80	95.1
400	60	94.6	82.8	95.7
500	53.7	94.75	83.4	96
600	62.5	95.3	84	96.2
700	65.6	95	84.4	97.3



Universiti Utara Malaysia

Figure 4.9. Comparison between $_{T}$S-layer HTM and original HTM

4.2.4 Dataset 4

In this experiment, a Genbase dataset is used. Comparison results are obtained between the original and improved Neocognitron neural network as mentioned in Chapter Three (section 3.4.1.1).

4.2.4.1 $_T$S-layer HTM vs. Neocognitron

This section presents the A comparison of results is obtained between the $_T$S-layer and original Neocognitron. Table 4.7 shows the number of features that are extracted from each input pattern by original Neocognitron and $_T$S-layer. From this table, it is noticed that the number of features extracted from the input pattern by using $_T$S-layer is still more than the features extracted by the original one, which indicates the robustness of the proposed method. Figure 4.10 shows the graphical presentation of this table, from this figure it was learned that the original Neocognitron which is marked with black line can extract features by about 74%, while TS-layer, which is marked with green line is 95%.

UUM Universiti Utara Malaysia

Table 4.7

Number of feature extracted dataset 2: $_T$S-layer HTM vs. original Neocognitron

No. of input pattern	Feature extracted %	
	Original Neocognitron	$_T$S-layer Neocognitron
100	73.6	90.1
200	74	80.7
300	77	87.5
400	93.7	94.6
500	87.6	94.75
600	92.5	95.3
700	90	95

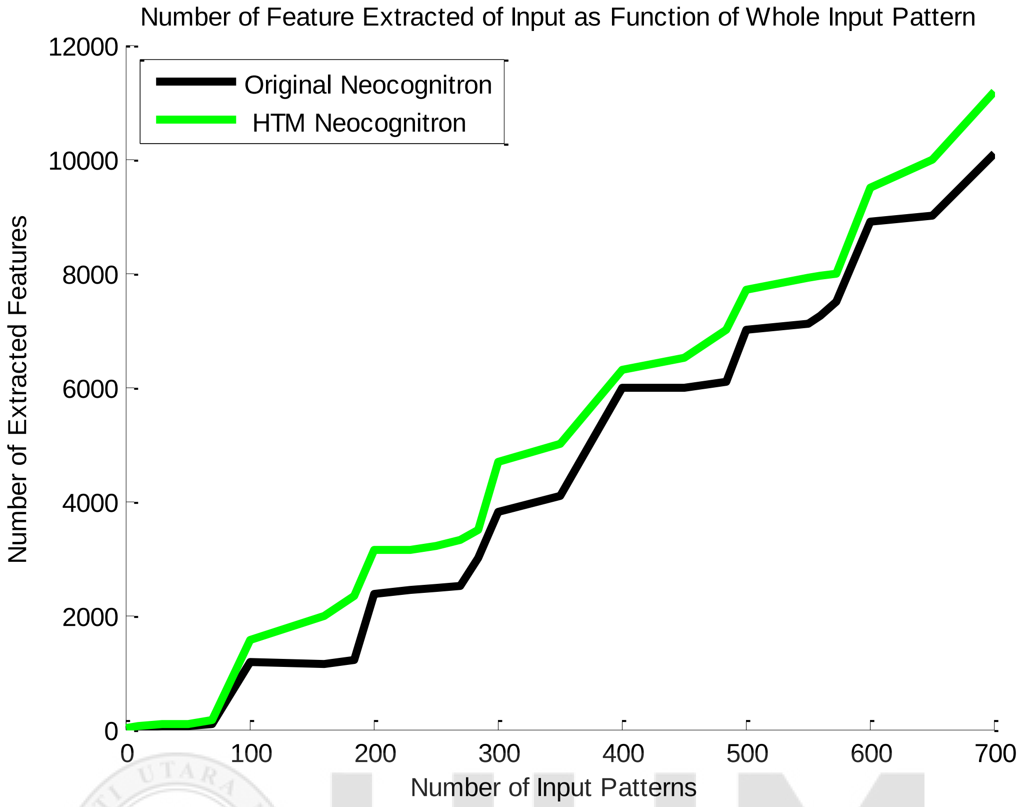


Figure 4.10. Number of features extracted as function of input patterns, dataset 4: $_T$S-layer Neocognitron vs. original Neocognitron.

Universiti Utara Malaysia

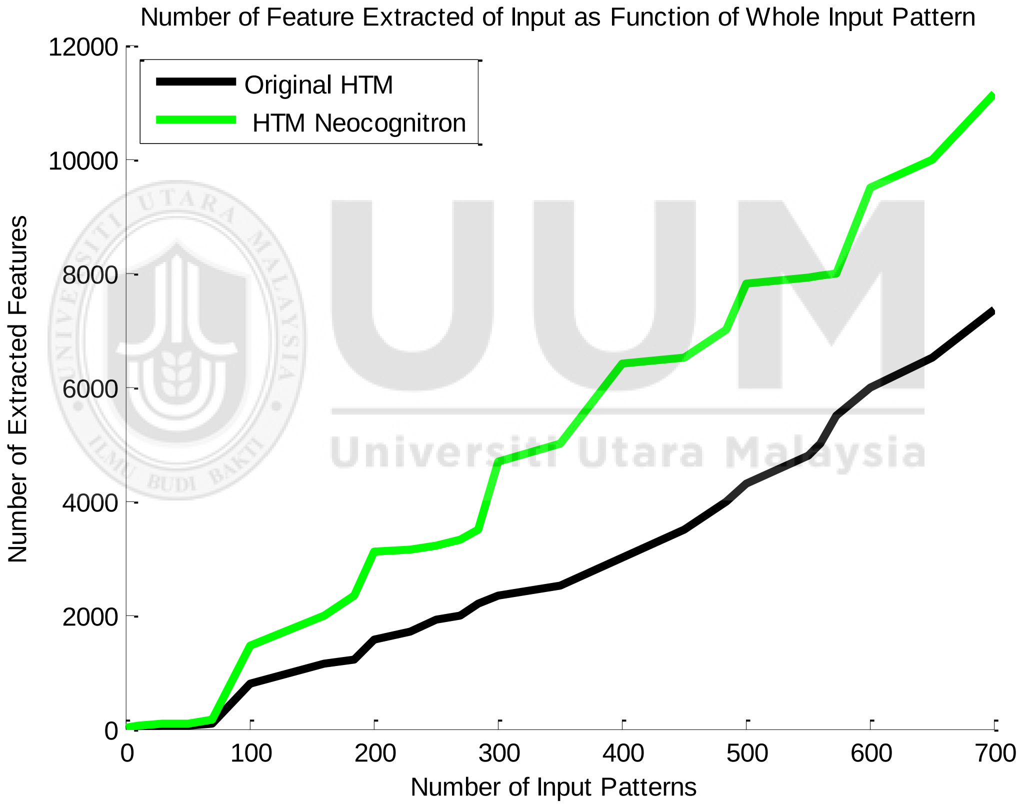
4.2.4.2 $_T$S-layer HTM vs. Original HTM

Table 4.8 illustrates the number of features that are extracted from input pattern and accuracy by the two algorithms. Figure 4.11 presents the graphical presentation of the table. From the table and figure, it is learned that the $_T$S-layer can extracts features from the input pattern by about 96%, while the original HTM can extract about 49% of the input pattern. It is obvious that there is great difference between the two methods which assign good performance of the proposed method.

Table 4.8

Feature extracted of dataset 2: $_{T}$S-layer HTM and original HTM

No. of Input Patterns	Features Extracted %		Accuracy %	
	Original HTM	$_{T}$S-layer HTM	Original HTM	$_{T}$S-layer HTM
100	49	90.1	83	95
200	48.6	80.7	83.5	95.3
300	48.7	87.5	84	95.8
400	46.8	94.6	84.8	95.9
500	53.7	94.75	85	96
600	62.5	95.3	85.2	96.8
700	65.6	95	87	97.8



UUM
Universiti Utara Malaysia

Figure 4.11. Comparison between $_{T}$S-layer HTM and original HTM

4.2.5 Dataset 5

In this experiment Yeast dataset is used. Comparison results are obtained between original and improved Neocognitron neural network as mentioned in chapter 3 section 3.4.1.

4.2.5.1 $_T$S-layer HTM vs. Neocognitron

This section presents the A comparison of results is obtained between the Improved Feature Extraction algorithm ($_T$S-layer) and original Neocognitron. Table 4.9 shows the number of features extracted from each input pattern by original and improved S-layer Neocognitron. From this table, it is seen that the number of features extracted from the input pattern by using $_T$S-layer is more than the features extracted by the original one. Figure 4.12, on the other hand, shows the graphical presentation of this table. From this figure, it was learned that the original Neocognitron which is marked with black line can extract features at about 74%, less than features extracted by the $_T$S-layer, which is marked with green line at about 97%.

Universiti Utara Malaysia

Table 4.9

Feature extracted dataset 3: $_T$S-layer HTM vs. original Neocognitron

No. of input pattern	Feature extracted %	
	Original Neocognitron	$_T$S-layer HTM
100	74	90.1
200	74	80.7
300	77	87.5
400	79.6	94.6
500	91	94.75
600	93	95.3
700	89	95

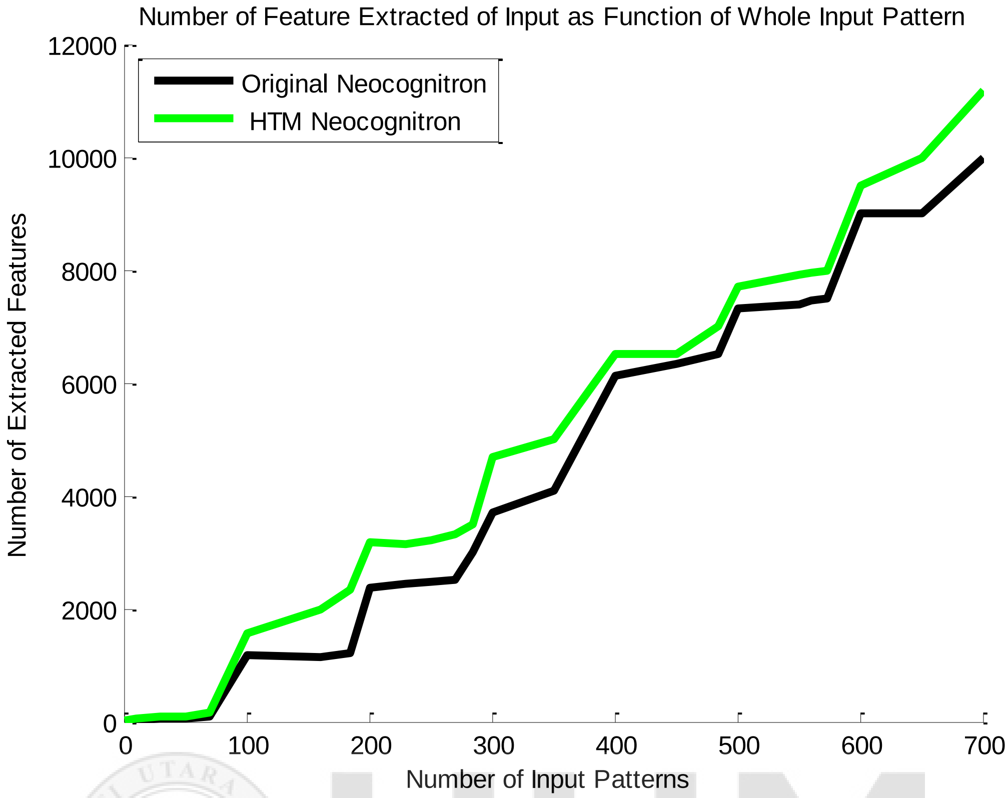


Figure 4.12. Number of features extracted as function of input patterns, dataset 5: ${}_{T}$S-layer Neocognitron vs. original Neocognitron.

Universiti Utara Malaysia

4.2.5.2 ${}_{T}$S-layer HTM vs. Original HTM

Table 4.10 illustrates the number of features that are extracted from input pattern and accuracy values by the two algorithms. Figure 4.13 presents the graphical presentation of the table. From the table and figure, it is learned that the ${}_{T}$S-layer can extracts features from the input pattern by about 98%, while the original HTM can extract about 47% of the input pattern. It is also seen that there is great difference between the two methods which demonstrates the good performance of the proposed algorithm.

Table 4.10

Feature extracted dataset 5: $_T$S-layer HTM and original HTM

No. of Input Patterns	Features extracted %		Accuracy %	
	Original HTM	$_T$S-layer HTM	Original HTM	$_T$S-layer HTM
100	49	90.1	83	96
200	48.5	80.7	84	96.2
300	48.7	87.5	86	96.5
400	46.8	94.6	89	96.9
500	53.7	94.75	89.5	97
600	62.5	95.3	90	98.8
700	65.6	95	90.6	99

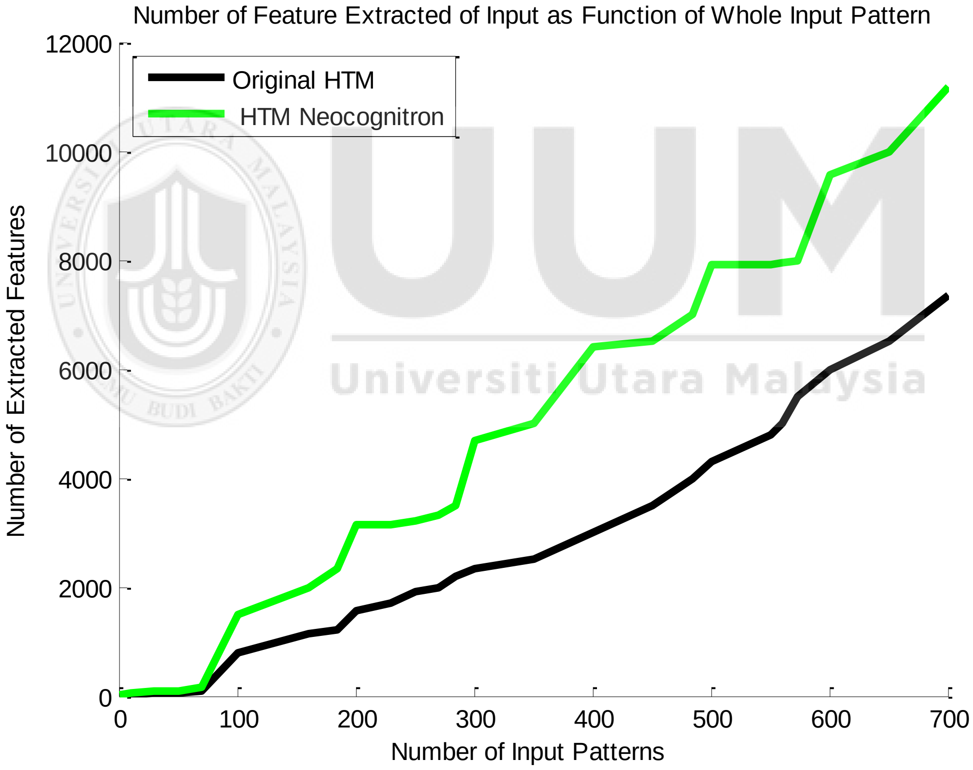


Figure 4.13. Comparison between $_T$S-layer HTM and original HTM

UUM Universiti Utara Malaysia

4.3 Evaluation of Classification Process

In this research, various evaluation measures as suggested by Tsoumakas (Kafrawy et al., 2015b) for classification are used. The evaluation measures of predictive performance are divided into two groups: bipartitions-based and rankings-based. The bipartitions-based evaluation measures are calculated based on the comparison of the predicted relevant labels with the ground truth relevant labels. This group of evaluation measures is further divided into example-based and label-based. The ranking-based evaluation measures compare the predicted ranking of the labels with the ground truth ranking.

Based measures, which are Hamming loss, accuracy, F1 score, Subset accuracy and label-based measures, which contain Macro-F1 and Micro-F. As shown in table 4.14, which demonstrate differences between improved Neocognitron and original one, it is seen that the $_{T}$S-layer generates higher value of accuracy, F1 score, Subset accuracy measures, Macro-F1 and Micro-F and smaller value for hamming loss than the original one.

Table 4.11 illustrates the evaluation results obtained by applying classification metrics.

Universiti Utara Malaysia

Table 4.11

Classification metrics for $_T$S-layer HTM vs. original Neocognitron

		Algorithm											
		$_T$S-layer HTM						**Original Neocognitron**					
	Evaluation Measure	Hamming loss	Accuracy	Subset _accuracy	Macro-F1	Micro-F1	F1 score	Hamming loss	Accuracy	Subset _accuracy	Macro-F1	Micro-F1	F1 score
Datasets	Blood Type	**0.1624**	**0.4233**	**0.6478**	**0.9880**	**0.3086**	**0.5478**	0.2923	0.2476	0.4027	0.4321	0.3011	0.5002
	Blood Cancer	**0.0765**	**0.3990**	**0.4778**	**0.6118**	**0.7578**	**0.6850**	0.0998	0.1830	0.1599	0.4391	0.5693	0.5010
	STARKEY"93	**0.1434**	**0.3893**	**0.5269**	**0.8317**	**0.4598**	**0.9867**	0.2476	0.1934	0.4018	0.5978	0.2943	0.5963
	Genbase	**0.1321**	**0.9499**	**0.2453**	**0.4953**	**0.3320**	**0.6678**	0.1351	0.5161	0.1051	0.2894	0.3001	0.5131
	Yeast	**0.1575**	**0.9311**	**0.1531**	**0.6053**	**0.9482**	**0.8965**	0.2875	0.6519	0.0099	0.5291	0.8028	0.5091

Universiti Utara Malaysia

From table 4.11, it could be seen that $_{T}$S-layer HTM achieves best result in evaluation than the original S-layer Neocognitron, which approve the high performance in processing shift in position. On the other hand, it could be also included that Yeast and STARKEY”93 provide the best results for almost all evaluation metrics. These data sets are large but non-medical, which indicate that it has large number of labels which give a chance to more overlapping. Hence, it can be concluded that the proposed algorithm achieves best performance in extracted features, greatly reducing the effect of shift in position that could be found in medical data more than other data (Wong et al., 2017), also it could be handle the large datasets effectively.

UUM Universiti Utara Malaysia

4.4 Comparison between $_{T}$S-layer and other methods

The previous sections have shown that the results of comparison between $_{T}$S-layer and statistical approaches (Neocognitron) and structural approaches (HTM) indicate that the $_{T}$S-layer is superior. In this section, a comparison is made between $_{T}$S-layer and an example method of the other approaches that were mentioned in Chapter two section 2.2. The sum of Absolute Difference (SAD) method of Template matching (Mahalakshmi et al., 2012), and Principal Component Analysis (PCA) of statistical approach (Huan et al., 2012), will be used in this section, these approached used for comparison as traditional statistic pattern recognition approaches Vs improved pattern recognition approach. This comparison is made to prove the high performance of $_{T}$S-layer algorithm with respect to the number of feature extracted to other approaches. Table 4.12 represents a portion of the digital representation of original image, the SAD and PCA result image. Table 4.13 includes the number of features that are extracted by the Improved HTM, SAD and PCA.

From the table and figure, it is seen that the percentage of the features extracted by SAD is 25% and 35% by PCA for blood type images 1. It is 15%, by SAD, while it is 35 % in PCA for blood cancer images. For dataset 1 it is 45% for PCA and for 40% for SAD. For dataset 2, the SAD method can extract 39% for PCA and 27% for SAD. For dataset 3, PCA can extract 42% of features, while SAD can extract 40%. It is also noticed that PCA method has better extraction percentage than SAD method, which indicates it is most robustness than SAD, as mentioned before of the simplicity of template matching approaches and robustness of PCA (Chapter Two, section 2.2.1, and section 2.2.2.1). Figure 4.14 represents the graphical representation of this comparison.

Table 4.12

Digital representation of the example image of blood type: (a) original image. (b) SAD result image, (c) PCA result image

255	255	255
255	255	255
255	255	255
255	255	255
254	254	254
254	254	254
254	254	254
254	254	254
254	254	254

a

255	255	255
248	248	248
252	254	253
231	198	218
228	192	212
225	191	212
229	197	214
222	180	208
255	255	255

b

204	228	228
228	255	255
228	255	255
228	255	255
228	255	255
228	255	255
228	255	255
228	255	255
228	255	255

c

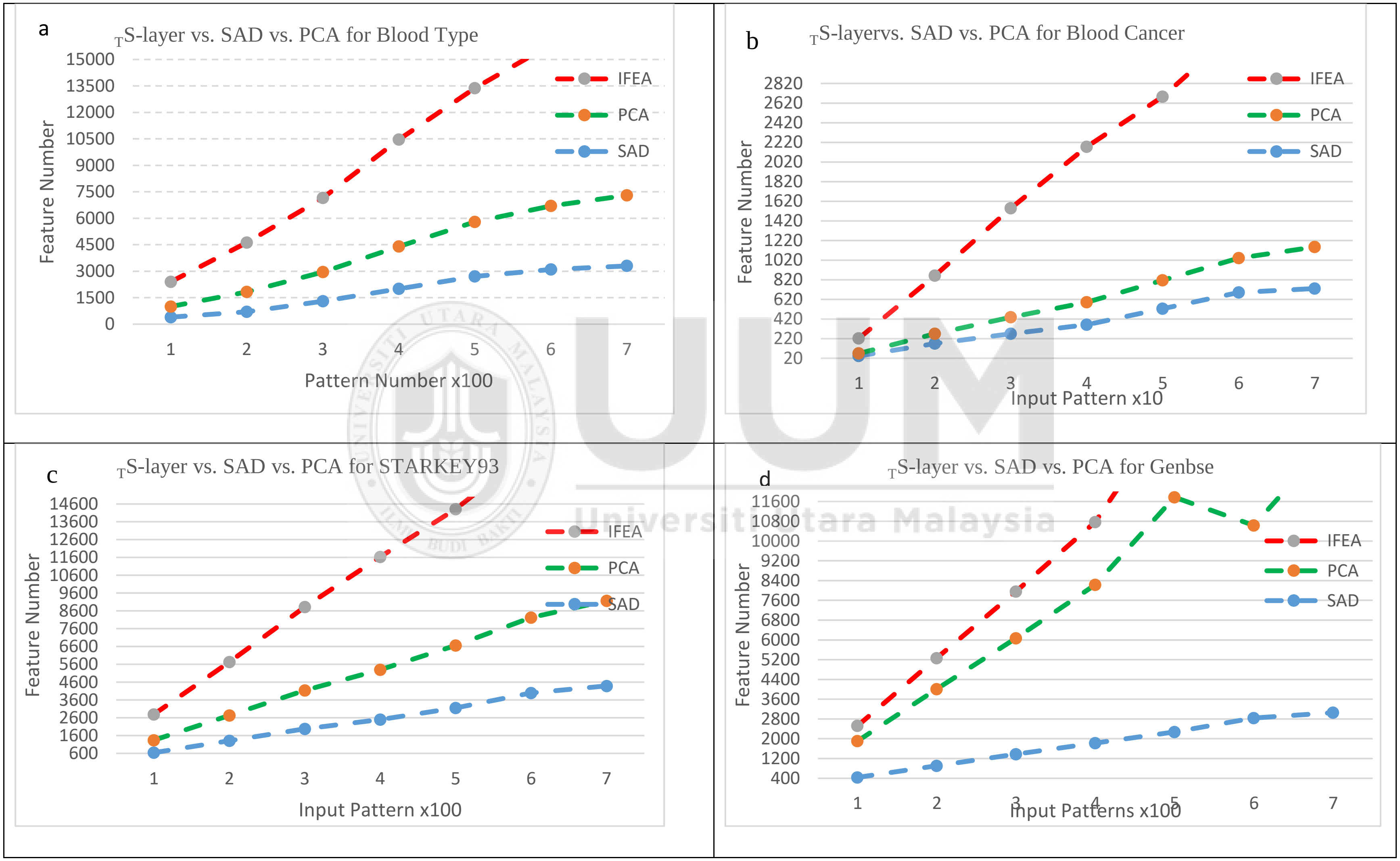
UUM Universiti Utara Malaysia

Table 4.13

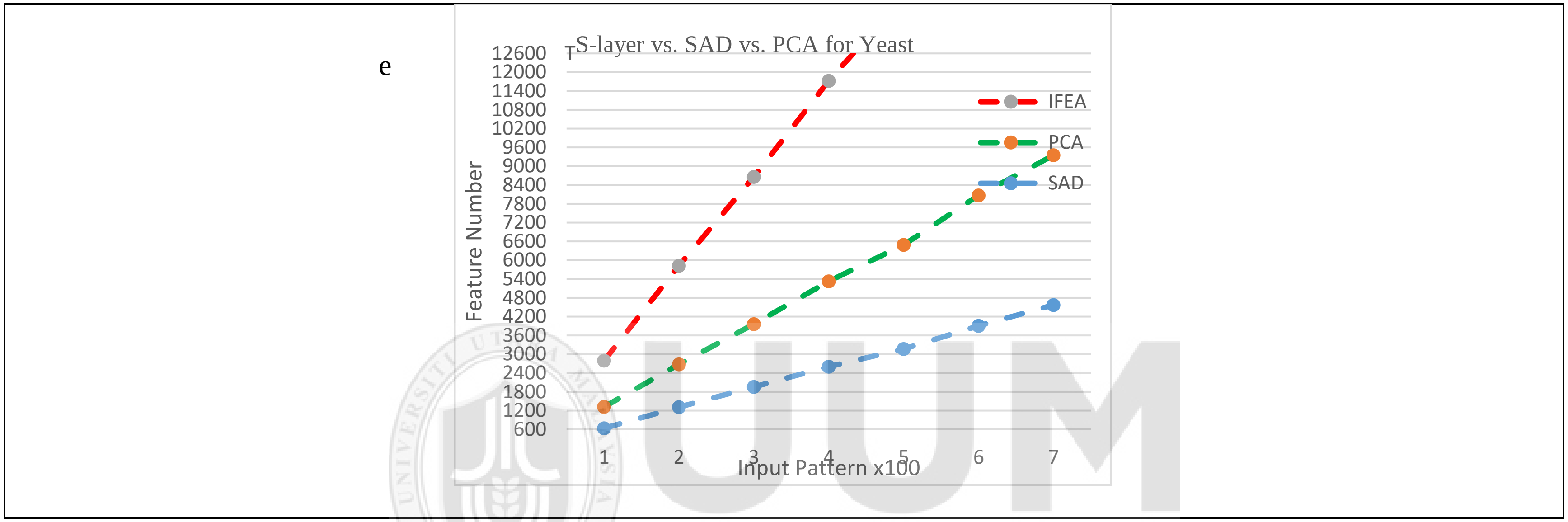
Number of features extracted: $_T$S-layer, SAD and PCA

No. of input pattern	Dataset	Number of Features		
		$_T$S-layer	**PCA**	**SAD**
100	Blood Type	**1400**	600	400
200		**2800**	1120	700
300		**4200**	1660	1300
400		**6060**	2400	2000
500		**7580**	3100	2700
600		**9150**	3600	3100
700		**10950**	4000	3300
10	Blood Cancer	**154**	43	25
40		**590**	170	100
70		**1110**	269	169
100		**1584**	360	230
130		**1870**	524	290
160		**2334**	689	350
190		**2725**	730	423
100	STARKEY'93	**1450**	700	640
200		**3000**	1420	1312
300		**4689**	2150	1980
400		**6340**	2800	2495
500		**7650**	3500	3157
600		**9430**	4234	3986
700		**11100**	4780	4389
100	Genbase	**1470**	615	438
200		**3100**	1261	908
300		**4698**	1890	1375
400		**6400**	2534	1832
500		**7800**	3120	2839
600		**9500**	3725	2278
700		**11150**	4376	3056
100	Yeast	**1477**	680	635
200		**3150**	1367	1305
300		**4698**	2010	1951
400		**6400**	2722	2604
500		**7900**	3330	3160
600		**9550**	4170	3900
700		**11170**	4780	4569

UUM Universiti Utara Malaysia



UUM
Universiti Utara Malaysia



Universiti Utara Malaysia

Figure 4.14. Graphical representation of feature number comparison of IFE vs. PCA vs. SAD for datasets: (a) Blood type. (b) Blood cancer, (c) STARKEY”93, (d) Genbase, (e) Yeast

Table 4.14illustrates the comparison of accuracy of recognition between original HTM, as traditional structural pattern recognition approach, and improved HTM. From this table, the improved HTM has higher accuracy than the original one which make the recognition process more accurate. It can be seen also in this table that the difference of accuracy values between improved and original HTM is high, which indicates a good performance of the improved HTM. The highest accuracy value is registered with blood cancer images.

Table 4.14

Summary of accuracy values between original HTM and imHTM

	Accuracy in %	
Dataset	Original HTM	imHTM
Blood type	84	96.2
Blood cancer	84.4	96.9
STARKEY"93	87	98
Genbase	83.3	97
Yeast	90.8	99

Universiti Utara Malaysia

4.5 Result analysis for $_{T}$S-layer

In comparison between $_{T}$S-layer and other approaches, which are original HTM (Skrynnik, Petrov, & Panov, 2016), Neocognitron neural network (Kunihiko, 2016), Principal Component Analysis (PCA) (Reddy & Raju, 2016) and Sum of Absolute Difference (SAD) (Dawoud, Samir, & Janier, 2011) with five experiments on five datasets, it is obvious that the $_{T}$S-layer is better than the others in terms of general performance.

From the previous discussion it is found that:

In comparison with the other pattern recognition approaches, the $_{T}$S-layer generates better results in terms of number of features extracted from the input pattern as shown in bold. From table 4.15 it is noted that Neocognitron is better than original HTM in extracting features for all datasets.

From the previous discussion it can be concluded that the relationship among classes may help in classification process, more close classes to each other, more chance of they are return to the same pattern. More close classes, more close distance which may leads to belongs them to the same pattern. It is also obvious that the improved method improves the performance of the original HTM especially when there is an overlapping in data points, as seen in table 4.14 where the difference between the accuracy of medical datasets (blood type, blood cancer, Genbase) of original and improved HTM is greater than it is between non-medical datasets (STARKEY”93 and Yeast). Table 4.15 includes a summary of result analysis of the four algorithms with the five datasets under consideration.

UUM Universiti Utara Malaysia

The following conclusion are also defined:

Data size: big data size may result in more overlapped in pixel points. As mentioned in section 4.2.1.1 the percentage of the features extracted by the original Neocognitron is 60%, while it is 90% by $_{T}$S-layer, with an image size of 64x64. The percentage of features extracted by original HTM is 43% and 90% by $_{T}$S-layer. While the percentages are 74%, 97% and about 50%, 89 respectively as in sections 4.2.2.1 and 4.2.2.2, when the size is

32x32. It can be seen that the imHTM achieves the best results for the number of extracted features than the other methods.

Type of data: The medical datasets contain more overlapping in data points (Ali, Ahma, George, Der, & Aziz, 2013), which leads to difficulties in extracting features and matching process. Dataset 2 is a medical dataset; the proposed method can achieve about 21% of extracted features more than the original Neocognitron achieve. Against the original HTM, the proposed method can achieve 47% more features and. As more datasets become available, more classes are added, and there is a very high imbalance between the classes.

For non-medical datasets 1 and 3, the proposed method can extract 18% more than original Neocognitron, and 45% more than the original HTM.

In multi-label classification, relationships between classes are:

The number of labels grows to the hundreds or even thousands, existing multi-label classification methods often become computationally inefficient.

Dependencies and correlations between labels for the large multi-label datasets depend upon number of transaction between labels. More transactions mean less correlation.

Because of the previous two points, some classes are become close to each other, which make the classification process more complex. Correlations of data points makes it difficult to relate them into their real classes, which complicate the classification process

Universiti Utara Malaysia

then affect classification accuracy. Hence, the ${}_{T}$S-layer algorithm provides an instrument for handling the correlation and number of labels by providing temporal matching.

Table 4.15

Summary of feature extraction comparison results of ${}_{T}$S-layer vs. Neocognitron vs. HTM vs. PCA vs. SAD

Datasets	Algorithms				
	${}_{T}$S-layer	HTM	Neocognitron	PCA	SAD
Blood Type	**90**	43	60	35	25
Blood Cancer	**89**	54	74	35	15
STARKEY”93	**95**	57	82	45	40
Genbase	**95**	49	74	39	27
Yeast	**97**	47	74	42	40

From the table, it can be concluded that ${}_{T}$S-layer extract a maximum number of features compared to other methods.

4.6 Summary

This chapter presents feature extraction for image and dataset by using improved S-layer Neocognitron algorithm with time factor. Experiments were conducted on 500 blood type images, 33 blood cancer images and three multi- label datasets. Evaluation metrics of classification accuracy which are Based measures (which are Hamming loss, accuracy, F1 score, Subset accuracy) and label-based measures (Macro-F1 and Micro-F) were applied to verify the performance of ${}_{T}$S-layer.

UUM Universiti Utara Malaysia

Comparison was made against original HTM algorithm based on number of features extracted by the two algorithms and to find the difference in accuracy. Another comparison is made against original Neocognitron for the purpose of evaluate the classification accuracy. Last, comparison made against PCA and SAD as pattern recognition method based on number of feature extraction.

It is learned that the proposed $_T$S-layer generates better results than the original one in terms of number of features extracted. A comparison of accuracy indicates that proposed $_T$S-layer provides better recognition of patterns.

UUM
Universiti Utara Malaysia

CHAPTER FIVE
FUZZY C-MEANS WITH TEMPORAL INFORMATION FOR HIERARCHICAL TEMPORAL MEMORY

5.1 Introduction

As mentioned previously (see Chapter One, section 1.2), existing HTM employs the Agglomerative Hierarchical Clustering (AHC), which is sensitive to noise and outliers. It is replaced with improved FCM clustering algorithm that incorporates temporal information into the membership function used for clustering. The algorithm is shown in Figure 5.1.

UUM
Universiti Utara Malaysia

Algorithm of $_{T}$FCM

```
1:  Initialize $\mu = [\mu_{ij}]^0 \; matrix$ of membership
2:  Choose parameter $l$ >0 to stop the iteration. Set the iteration counting parameter $l$ equal to 1
3: Set time matrix $t[i] = 0$
4:  At k-step calculate the centers vectors $v^k$=$[v_j]$ by
        $\boldsymbol{v_j} = (\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} (\mu_{\boldsymbol{ij}})^{\boldsymbol{m}} \boldsymbol{x_i}) / (\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} (\mu_{\boldsymbol{ij}})^{\boldsymbol{m}}), \; \forall_{\boldsymbol{j}} = \mathbf{1, 2, 3} \ldots . \boldsymbol{c}$
5:  Update the membership matrix $[\mu_{ij}]^k$ , $[\mu_{ij}]^{k+1}$ by
        $\mu_{\boldsymbol{ij}} = \mathbf{1} / \sum_{\boldsymbol{k=1}}^{\boldsymbol{c}} (\boldsymbol{d_{ij}} / \boldsymbol{d_{ik}})^{(\frac{\mathbf{2}}{\boldsymbol{m-1}})}$
6:  if  $[\mu_{ij}]^{k+1}$- $[\mu_{ij}]^k > \beta$    where $\beta$ is the termination condition between [0, 1] such that
        $\boldsymbol{\beta} = \Delta_{\boldsymbol{j}}^{\boldsymbol{N}} \left| [\mu_{ij}]^{k+1} - [\mu_{ij}]^k \right|$
7:          Goto step 4
8:      else
9:          stop at some iteration $l$
10:         set $t[i]$
11:   end if
12:  Calculate the temporal function and map it to the new membership function by
        $\boldsymbol{p_{ij}} = \sum_{\boldsymbol{t=1}}^{\boldsymbol{m}} \boldsymbol{\mu_{it}}$
        $\boldsymbol{\mu}_{\boldsymbol{new_{ij}}} = \frac{\boldsymbol{\mu_{ij} p_{ij}}}{\sum_{\boldsymbol{t=1}}^{\boldsymbol{c}} \boldsymbol{\mu_{tj} p_{tj}}}$
13:  if $d_{ij} == 0.02$
14 go to step 4
15:  else
16:    Goto step 17
17:  end if
```

Figure 5.1. Algorithm of $_{T}$FCM

In this step, the node partitions the set of vertices of the Markov graph into a set of temporal groups. Each temporal group is a subset of the set of vertices of the Markov graph. Partitioning is done such that the vertices of the same temporal group are highly

Universiti Utara Malaysia

likely to follow one another. The node uses improved FCM clustering method ($_T$FCM) with improved distance metric for separate Markov graph obtained from the previous step. Clustering based on this metric puts patterns that are likely to follow one another in time into the same cluster. Temporal information indicates that the neighboring pixels, in time, are highly correlated, and thus possesses the same feature value. This means that the probability that they belong to the same cluster is high. Hence, $_T$FCM will exploit temporal information and process it. Patterns that are unlikely to follow each other fall into different clusters. Since each cluster is a set of patterns that are very likely to follow each other in time, each cluster is termed a temporal group.

$_T$FCM clustering algorithm starts with all stored features in one clusters, as initial state, and then recursively separate and merge clusters depending on highest membership. As mentioned in Chapter 3 $_T$FCM has two phases, first the standard FCM will be applied to calculate membership function for the feature clustering (return features to its pattern). This phase will do clustering which assigns each feature extracted from the first step of HTM to its pattern portion. After a clustering process comparison is made against the threshold, which is the distance, if it is more than specific value, which determined previously, a Markov process will be carried out again.

Universiti Utara Malaysia

Second phase temporal FCM will be applied for final clustering of patterns which return a portion of patterns resulted from first phase to its real pattern then recognition of input image.

5.2 Results

This section presents the results obtained by applying the improved temporal FCM ($_{T}$FCM) algorithm in HTM on the datasets mentioned in Chapter Three (section 3.2.1).

5.2.1 Dataset 1

In this section, blood type images are used for experiments of applying $_{T}$FCM. As mentioned in Chapter Four, 3000 images each were used of size 64 x 64 pixels, with 2500 for training and 500 for testing. Three example images of different densities will be demonstrated in experiment illustration to observe the difference in performance on different densities. Comparisons of results were obtained between $_{T}$FCM and conventional one.

Universiti Utara Malaysia

5.2.1.1 $_{T}$FCM vs. Conventional FCM

This section presents the A comparison of results is obtained with the improved temporal FCM ($_{T}$FCM) and conventional FCM. For demonstration of example of experiment, three test images with different cluster densities are reviewed, in order to demonstrate the performance of the proposed algorithm.

From the results tabulated in table 5.1, which illustrate the values of iteration count (IC) and minimum objective function (MOF), it is noted that $_{T}$FCM has the smallest (IC) that gives (MOF) which leads to high performance of clustering. The value of objective function affects the accuracy of the clustering. This is because when data point is close to the cluster center, the value of objective function becomes smaller, which causes a high membership function.

As in table 5.1, the MOF value is (29.732) at IC (7) in $_{T}$FCM, while its value is (7127.155) at IC 20 in FCM for image (a). For image (b), MOF is (7.028649) at IC (Bova & G, et) for $_{T}$FCM and (1029.81) at IC (Bova & G, et) for FCM, for image (c), MOF is (6.877901) at IC (Bova & G, et) for $_{T}$FCM and (2768.0708) at IC 18 for FCM. The high deference between the values of objective function for the two algorithms is because of the high overlap in medical data. Such results indicate that the conventional FCM is sensitive to overlapping, especially in medical data, as mentioned in Chapter One (section 1.2). Table 5.2 includes a summary of the values of IC and MOF for the three images and for whole dataset. Figure 5.3 (a) shows the graphical presentation for the progress of the objective function for $_{T}$FCM. From this figure, it is seen that (MOF) of the $_{T}$FCM is gained with less value of (IC).

It is also seen from the table that images (b) and (c) have the best clustering performance with respect to the values of IC and MOF, this conclusion supports the cluster density assumption, that the clustering performance is highly affected by the data structure and cluster density (Bonchi, Gionis, & Ukkonen, 2011). It shows poor performance when cluster densities are highly different. This is clear in figure 5.2, which shows the cluster densities for this data. From this figure, the differences in cluster density for image (a) is high, hence the performance is poor, while in (b) and (c) the cluster density of the two clusters are approximately the same, which leads to high clustering performance. This is made clear in table 5.2. Figure 5.3 (b) shows the graphical presentation for the progress of the objective function for FCM. Results also show that $_{T}$FCM is better than FCM in terms of average distance between cluster canter and data points, it is (1.045) for FCM,

Universiti Utara Malaysia

while it is (1.030) for $_{T}$FCM. as mentioned in table 5.2. Hence, temporal information is useful to facilitate the mapping of a data point to its cluster.

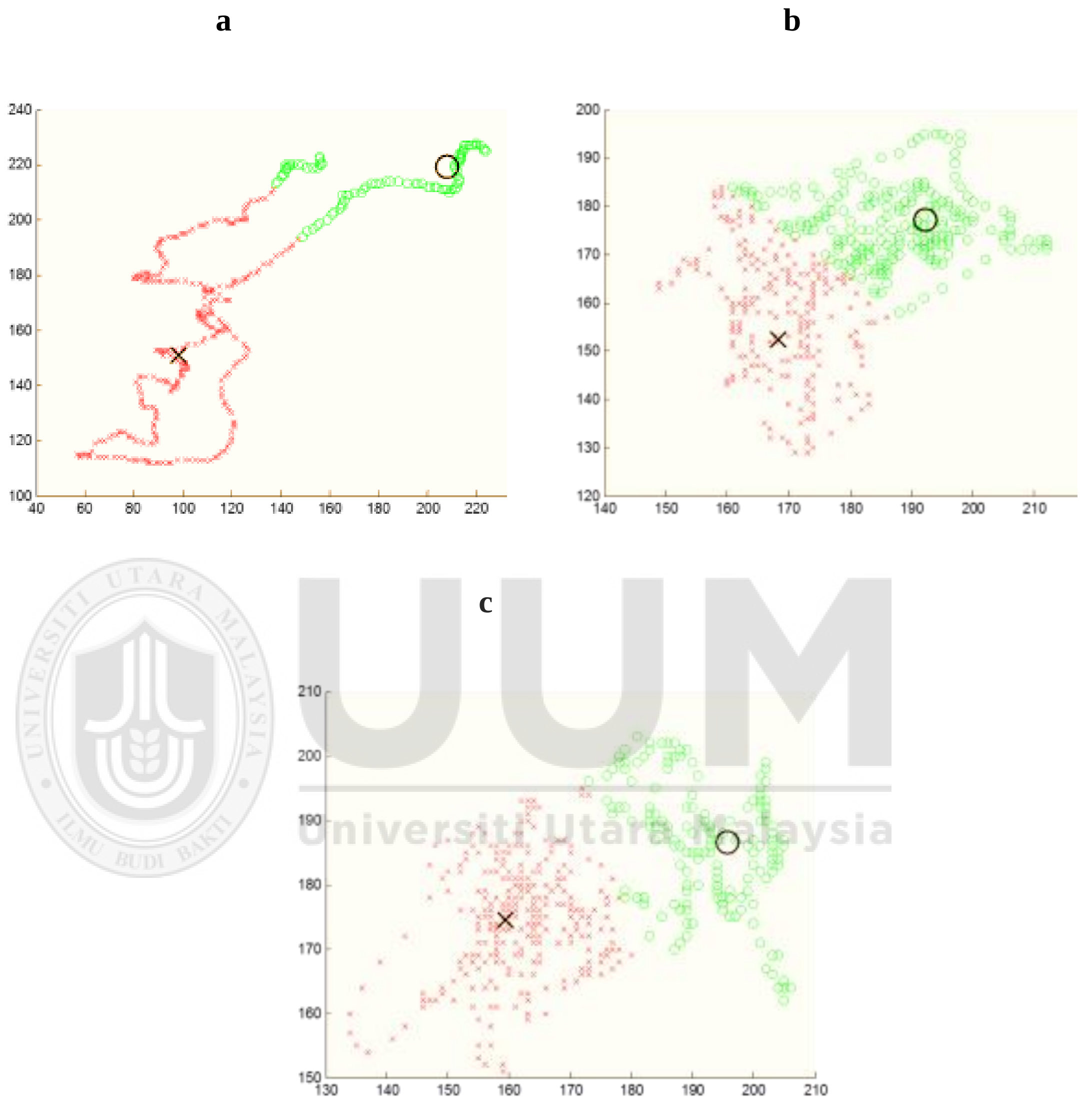


Figure 5.2. Cluster density for the three sample images

UUM
Universiti Utara Malaysia

Table 5.1

Value of IC and MOF for image (a), (b), (c), FCM vs. $_T$FCM

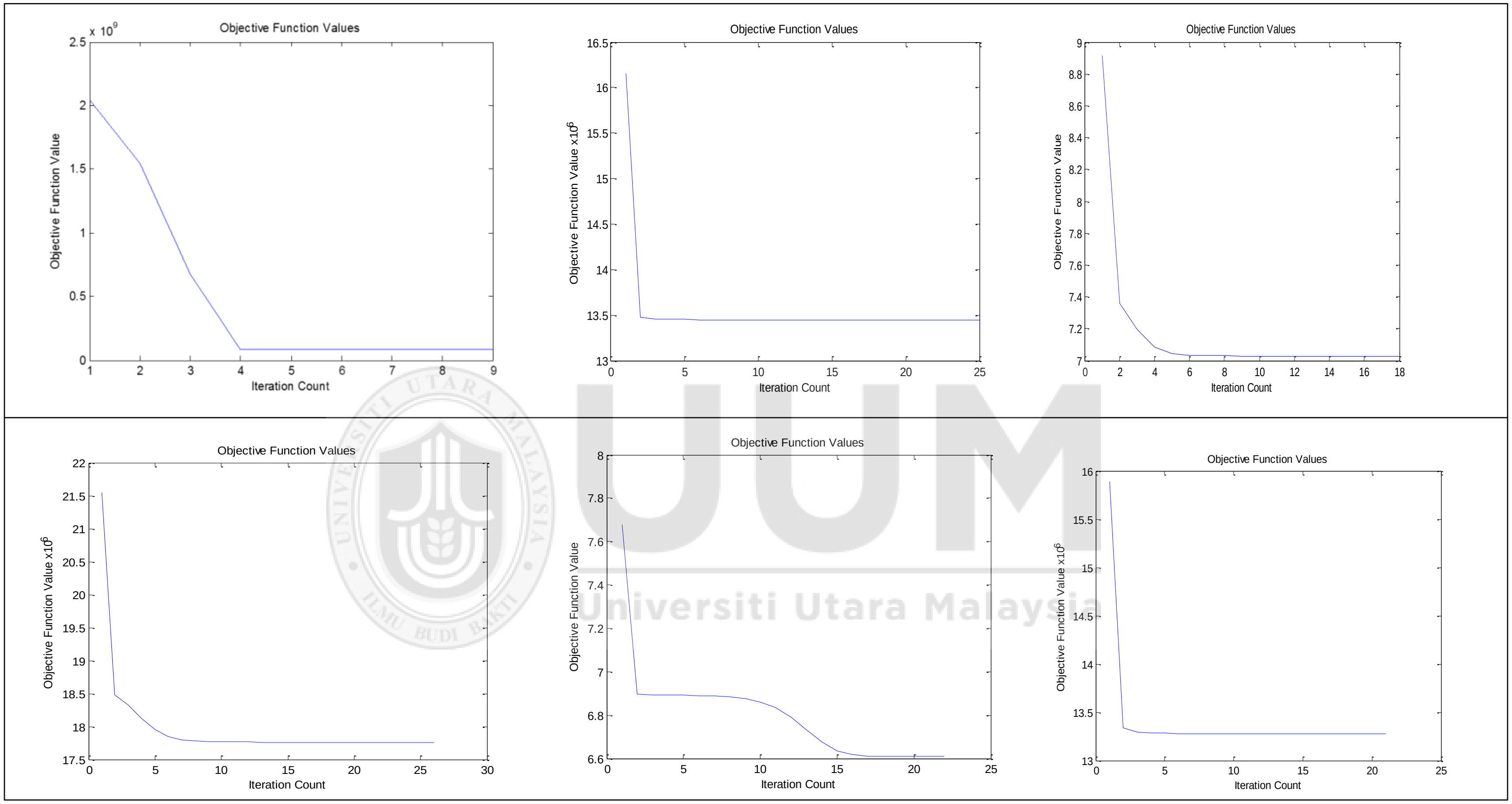
IC	MOF					
	Image (a)		Image (b)		Image (c)	
	FCM	$_T$FCM	FCM	$_T$FCM	FCM	$_T$FCM
1	8968.421	72.418	1044.23	8.468218	9241.8166	8.235676
2	7228.506	70.301	1044.05	7.152326	7030.1382	6.893992
3	7227.483	63.126	1042.48	7.061755	7312.5940	6.889045
4	7227.321	61.632	1041.23	7.036605	6163.1725	6.885044
5	7227.261	59.071	**1029.81**	**7.028649**	5907.0850	**6.877901**
6	7227.226	35.207	7338.7		3520.7451	
7	7227.204	**29.732**	6334.1		2972.1806	
8	7227.189		6260.0		2824.4825	
10	7227.177		6256.9		2783.3919	
12	7227.161		6156.9		2771.8378	
18	7227.155				**2768.0708**	
20	**7127.155**					

UUM Universiti Utara Malaysia

Table 5.2

Summary of iteration counts that leads to the minimum objective functions for dtaset1

Images	**Iteration count**		**Objective Function**		**Average Distance**	
	FCM	**$_T$FCM**	**FCM**	**$_T$FCM**	**FCM**	**$_T$FCM**
Image (a)	20	7	7127.155	**29.732**	1.045	1.030
Image (b)	5	5	6156.9	**7.02869**	1.012	0.731
Image (c)	18	5	2768.0708	**6.87790**	1.021	1.002
500 images	15	5	4236.55	**10.4432**	1.243	**0.675**



UUM
Universiti Utara Malaysia

Figure 5.3. Progress of objective function for a: ${}_{T}$FCM, b: FCM

5.2.2 Dataset 2

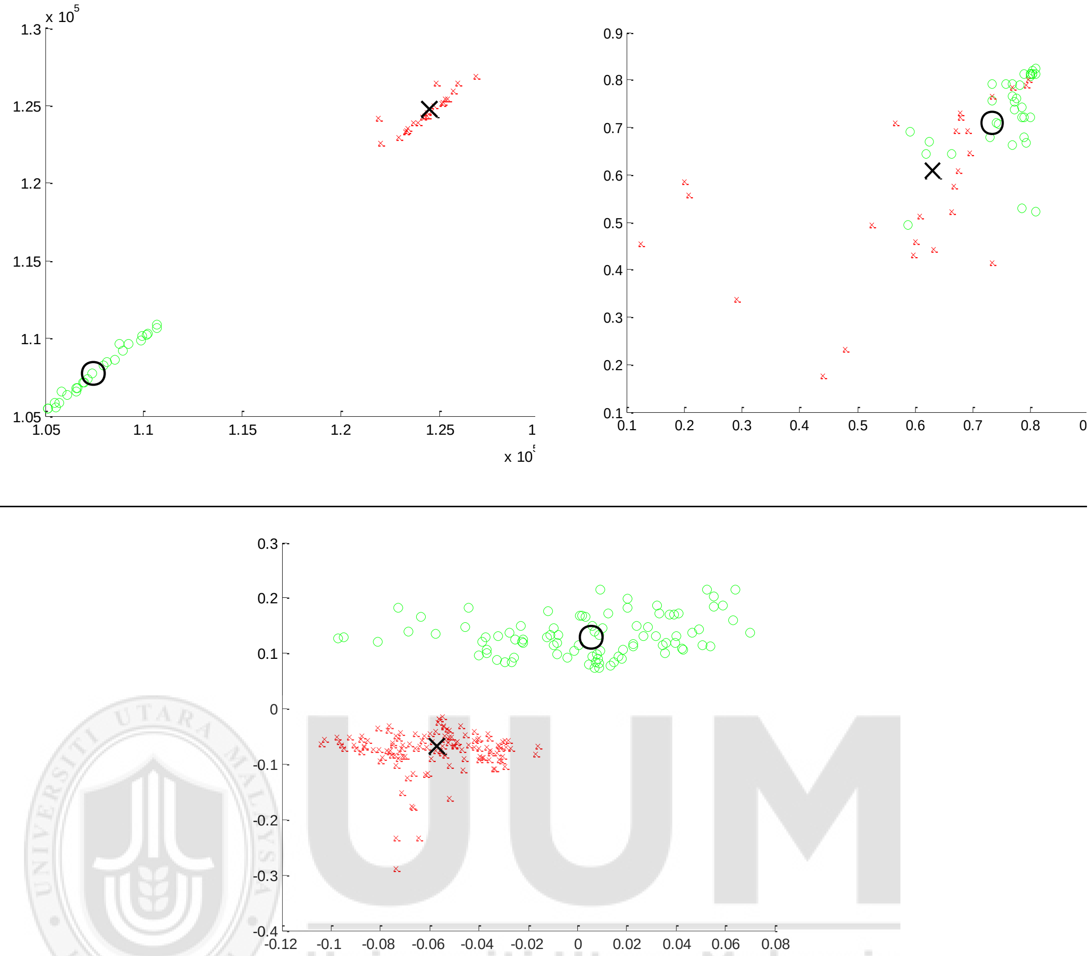
In this section, a blood cancer images are used. The images are as mentioned in Chapter Four for blood cancer images. A comparison of results is obtained between original FCM and $_{T}$FCM.

5.2.2.1 $_{T}$FCM vs. Conventional FCM

Table 5.3 illustrate the IC and MOF for the three images. From the results tabulated in table 5.3, it is noted that the $_{T}$FCM has the least iteration count that gives minimum objective function which leads to high performance of clustering.

$_{T}$FCM generates the smallest values for IC and MOF, which are (8) and (0.710689) for image (a). (12) and (8.609809) for image (b), (9) and (4.763007) for image (c) respectively. It is also seen that image (a) has the minimum value for IC and MOF, because the image has a small difference in cluster density which leads to minimum objective function (Lou et al., 2012), as seen in Figure 5.4, which represents the cluster densities of the three images. Table 5.4 demonstrates a summary of the values of minimum IC and MOF for the three images. The values of IC and MOF for FCM are: (13) and (1.3127) for image (a), (17) and (7227.139) for image (b), (18) and (17.160) for image (c), it is 1.45631 at IC 4 for whole test data with $_{T}$FCM, while it is (234.433) at IC (13) for FCM. For average distance, it is (1.011), which is better than (1.098) for FCM. Figure 5.5 show a graphical representation of the progress of the objective function for $_{T}$FCM and FCM. It is clear from this figure that $_{T}$FCM generates the minimum objective function in terms of least iteration count. The results also show that $_{T}$FCM is better than FCM in terms of average distance between cluster canter and data points as mentioned in table 5.4.

UUM Universiti Utara Malaysia



UNIVERSITI UTARA MALAYSIA ILMU BUDI BAKTI
UUM
Universiti Utara Malaysia

Figure 5.4. Cluster density for the three sample images

Table 5.3

Value of IC and MOF for image (a), (b), (c), FCM vs. $_{T}$FCM

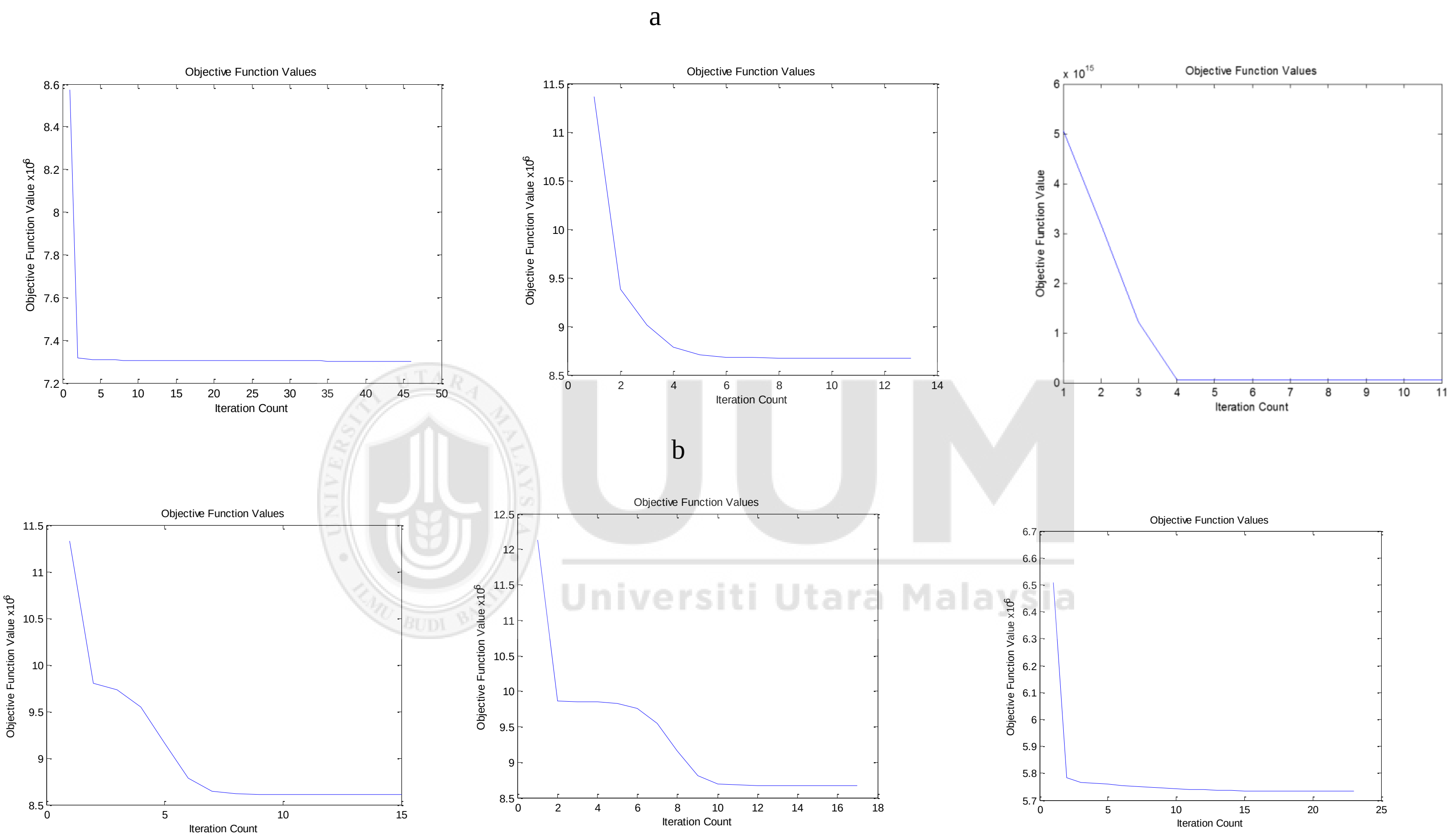
IC	**MOF**					
	Image (a)		**Image (b)**		**Image (c)**	
	FCM	**$_{T}$FCM**	**FCM**	**$_{T}$FCM**	**FCM**	**$_{T}$FCM**
1	1.5192	2.967552	8762.484	11.51296	21.042	5.83759
2	1.3384	1.713002	7228.703	9.838618	18.650	4.77601
3	1.3362	1.711142	7227.476	9.828054	18.643	4.764985
4	1.3349	0.710827	7227.296	9.822346	18.641	4.763582
5	1.3339	0.710734	7227.235	9.805333	18.639	4.763207
6	1.3335	0.710703	7227.206	9.74963	18.634	4.763079
7	1.3334	0.710692	7227.188	9.584734	18.276	4.763032
8	1.3334	**0.710689**	7227.176	9.223209	18.123	4.763014
9	1.3339		7227.167	8.817234	17.988	**4.763007**
10	1.3349		7227.166	8.648225	17.768	
11	1.3212		7227.155	8.617109	17.765	
12	1.3213		7227.151	**8.609809**	17.764	
13	1.3127		7227.148	9.001232	17.762	
14			7227.145	9.012345	17.762	
15			7227.143	9.108484	17.761	
16			7227.142		17.761	
17			7227.139		17.760	
18					17.160	

Universiti Utara Malaysia

Table 5.4

Summary of iteration counts that leads to the minimum objective functions for dataset 2.

Images	Iteration count		Objective Function		Average Distance	
	FCM	$_{T}$FCM	FCM	$_{T}$FCM	FCM	$_{T}$FCM
Image (a)	13	8	1.3127	**0.71068**	1.063	1.07
Image (b)	17	12	7227.139	**8.60980**	1.084	1.231
Image (c)	18	9	17.160	**4.76300**	1.041	1.03
33 images	13	4	234.433	**1.45631**	1.098	**1.011**



UUM
Universiti Utara Malaysia

Figure 5.5. Progress of objective function or a: $_{T}$FCM, b: FCM

5.2.3 Dataset 3

In this section, a STARKEY”93 dataset is used. A comparison of results is obtained between original and improved FCM and conventional one.

5.2.3.1 ${}_{T}$FCM vs. Conventional FCM

Table 5.5 shows the value of IC and MOF that are obtained from applying original FCM and ${}_{T}$FCM on this dataset. From this table, it is seen that ${}_{T}$FCM generates MOF with less IC than FCM which leads to high performance of clustering.

For the STAEKEY 93 dataset, (as in Table 5.5), the values of IC and MOF are 10 and 1.109 for ${}_{T}$FCM, and 20 and 4.12621 for FCM. ${}_{T}$FCM has minimum objective function with a small number of iterations, which leads to high performance. The difference in iteration count is high between the FCM and ${}_{T}$FCM. This is because that the size of the dataset affects the number of iteration count. Big dataset results more iteration counts that gives minimum objective function. Figure 5.6 shows the progress of the objective function for the two methods. Figure 5.7 represents the cluster density of the dataset, showing that the density is quite poor, which affects the performance of this dataset with respect to the other dataset. This also affects the average distance between cluster center and data point. With these results, it is learned that ${}_{T}$FCM is better than FCM in terms of average distance with (0.5932) for FCM, while it is (0.231) for ${}_{T}$FCM, as illustrated in table 5.5

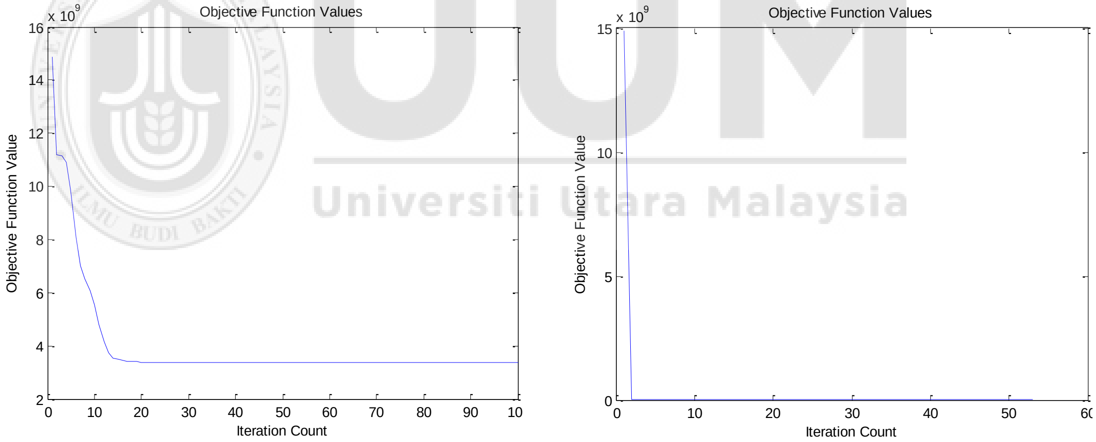
Universiti Utara Malaysia

Table 5.5

Value of average distance, IC, MOF of FCM vs. $_T$FCM

	IC	MOF	
		FCM	$_T$FCM
	1	8.42134	3.9209
	2	7.50610	2.9409
	3	6.48311	2.9309
	4	6.32112	2.8809
	5	6.26113	2.709
	6	5.22613	2.609
	7	5.22425	2.309
	8	5.22525	2.209
	9	5.21622	1.980
	10	4.22622	**1.109**
	20	4.12621	
Average Distance		0.5932	0.231

Universiti Utara Malaysia



a

b

Figure 5.6. Progress of objective function for: (a) FCM, (b) $_T$FCM

Figure 5.7 shows the cluster density for this dataset. From this figure, the differences in cluster density is high, hence the so the performance is quite poor. This is clear in the small deference of entropy and purity values between the two methods

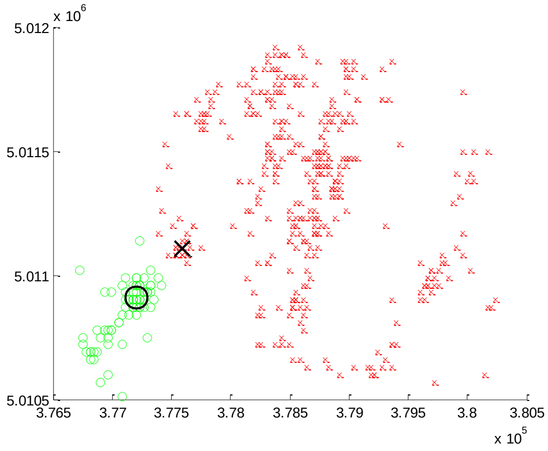


Figure 5.7. Cluster densities of STARKEY”93 dataset

5.2.4 Dataset 4

In this section, a Genbase dataset is used. Comparisons of results are made between original and $_T$FCM.

Universiti Utara Malaysia

5.2.4.1 $_T$ FCM vs. Conventional FCM

Table 5.6 shows the value of IC and MOF obtained from applying original FCM and $_T$FCM on this dataset. From this table, it is seen that the values of IC and MOF that is generated by applying $_T$FCM, are (7) and (1.10), while they are (30) and (1.455) for FCM, which proof that $_T$FCM will leads to high performance than use original FCM, hence the clustering process becomes to be more accurate, then accurate recognition of pattern. Also, there is a high difference in iteration count between the FCM and $_T$FCM, which it is an evidence of the powerful robustness of $_T$FCM on medical dataset which has overlapping in data points. Figure 5.8 shows the progress of the objective function for the two methods.

From table 5.6, it is also seen that ${}_{T}$FCM is better than FCM in terms of average distance between cluster canter and data points, it is (4.1774) for FCM, while it is (2.0804) for ${}_{T}$FCM, which affects the clustering process, the smallest the average, the more accurate clustering. Figure 5.9 represents the cluster density of this dataset. From this figure, the differences in cluster density are small, and good performance is achieved.

Table 5.6

Value of average distance, IC, and MOF of ${}_{T}$FCM vs. FCM

	IC	**MOF**	
		FCM	**${}_{T}$FCM**
	1	7.14	2.96
	2	7.139	2.94
	3	7.138	2.93
	4	7.137	2.88
	5	7.129	2.70
	6	6.126	1.60
	7	6.125	**1.10**
	8	6.130	
	9	6.129	
	10	5.110	
	20	4.056	
	21	3.042	
	24	3.300	
	25	2.220	
	28	1.125	
	30	1.455	
Average Distance		4.1774	3.0804

Universiti Utara Malaysia

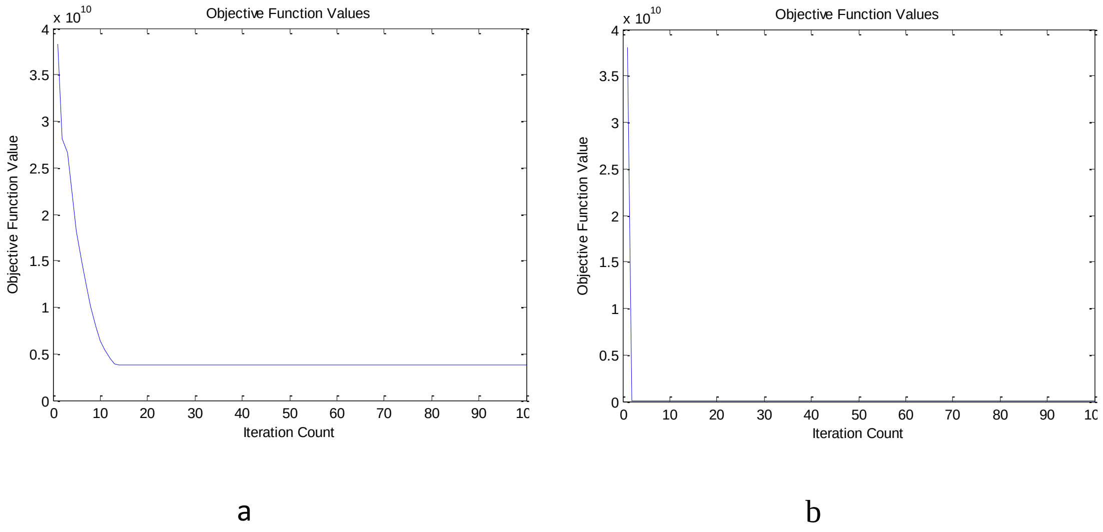
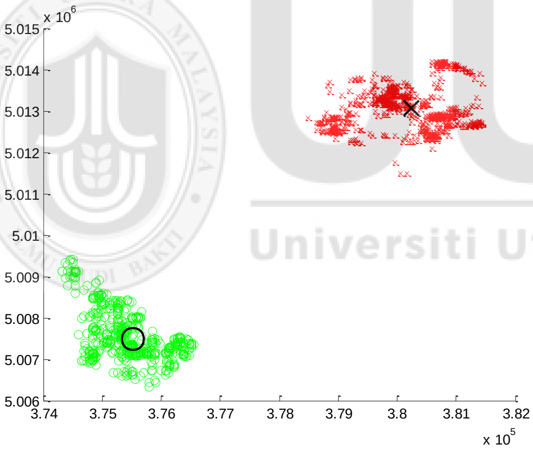


Figure 5.8. Progress of objective function for: (a) FCM, (b) $_T$FCM



Universiti Utara Malaysia

Figure 5.9. Cluster densities of Genbase dataset.

5.2.5 Dataset 5

In this section, the Yeast dataset is used. A comparison of results is obtained between FCM and $_T$FCM.

5.2.5.1 ${}_{T}$FCM vs. Conventional FCM

Table 5.7 shows the values of IC and MOF obtained from applying original FCM and ${}_{T}$FCM on this dataset. From the results tabulated in table 5.7, it is noted that ${}_{T}$FCM has the least iteration count that gives minimum objective function which leads to high performance of clustering.

The value of IC is (8) for MOF (1.07E+10) in ${}_{T}$FCM, while the value of IC is (90) for MOF (3.55E+09) in FCM. ${}_{T}$FCM has minimum objective function with small number of iterations, which leads to high performance. There is high difference in iteration count between FCM and ${}_{T}$FCM, (as seen in table 5.7), because of the big size of the dataset. Figure 5.10 shows the progress of the objective function for the two methods. Also results show that ${}_{T}$FCM is better than FCM in terms of average distance between cluster canter and data points, it is (4.7864) for FCM, while it is (0.1911) for ${}_{T}$FCM, as mentioned in table 5.7. This is clear in figure 5.11, which shows the cluster density for this dataset.

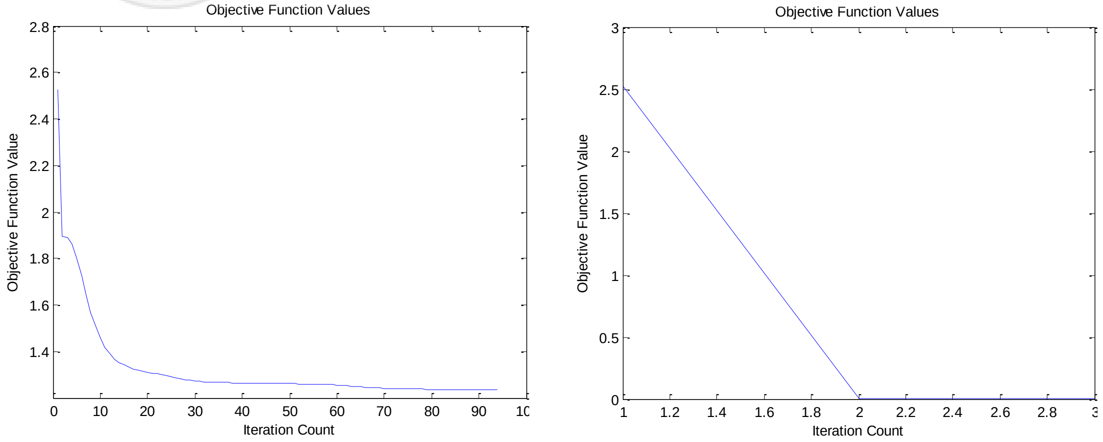
UUM Universiti Utara Malaysia

Table 5.7

Value of IC and MOF of ${}_{T}$FCM vs. FCM

	IC	MOF	
		FCM	**${}_{T}$FCM**
	1	3.78E+10	3.78E+10
	2	3.94E+09	2.8E+10
	3	3.79E+09	2.64E+10
	4	3.59E+09	2.38E+10
	5	3.56E+09	2.02E+10
	6	3.55E+09	1.64E+10
	7	3.55E+09	1.33E+10
	8	3.55E+09	1.07E+10
	50	3.55E+09	
	55	3.55E+09	
	60	3.55E+09	
	65	3.55E+09	
	70	3.55E+09	
	75	3.55E+09	
	80	3.55E+09	
	85	3.55E+09	
	90	3.55E+09	
Average Distance		4.7864	0.1911

Universiti Utara Malaysia



a

b

Figure 5.10. Objective function for a: FCM, b: improved ${}_{T}$FCM for dataset 3.

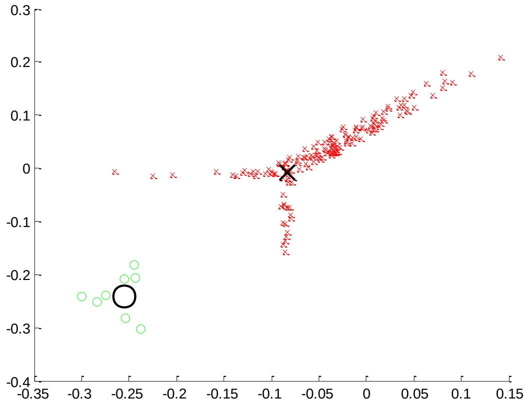


Figure 5.11. Cluster densities of dataset 5.

5.3 Evaluation

The proposed temporal FCM algorithm is later evaluated in order to achieve the optimum accuracy of clustering process for overlapping in data points contained in medical and multi- label datasets. Table 5.8 illustrates the results of the Purity and Entropy evaluation matrices that are used to evaluate the proposed algorithm. From table 5.8 it is noted that the value of purity for $_T$FCM is higher than it is in FCM, which leads to high performance of clustering, as mentioned in Chapter 3 (section 3.4.2). The best performance is assigned to the value closest to zero. Also, it is obvious from this table that the difference between the purity values of FCM and $_T$FCM, for blood type dataset is high (0.0370 and 0.9753). This is medical data has high overlapping which complicate the clustering process then affect the performance (Yinghua Lu et al., 2013). This is support the effectiveness of proposed $_T$FCM with overlapping in data points. Also, the highest purity value is recorded for Genbase dataset (0.9970). From previous discussion it can be concluded that the proposed algorithm indicates more robustness than original one.

It is also noted from table 5.8 that the difference in purity for blood cancer data is less than it is in blood type, which gives evidence of the robustness of proposed FCM algorithm with large datasets which have more overlapping.

For entropy, the difference between FCM and $_T$FCM, for blood cancer is high. The difference between the values of entropy of the two algorithms for Genbase (0.9526 and 2.1230) is also high as a result of overlapping, which also refer to the robustness of the proposed algorithm, while the difference between the purity values for Genbase of FCM and $_T$FCM, is higher than it is in STARKEY"93, which is non -medical. It may also be noted from table 5.11 that the smallest value of entropy could be seen in blood cancer data, while the higher values are in the Yeast dataset (1.232). Figure 5.9 and 5.10 illustrates the graphical representation of the result.

Universiti Utara Malaysia

Table 5.8

Purity and Entropy external evaluation metric for clustering of FCM vs. $_T$FCM.

Evaluation metric	Dataset	$_T$FCM	FCM
	Blood Type	**0.9753**	0.0370
	Blood cancer	**0.9345**	0.1231
	STARKEY"93	**0.7001**	0.0596
	Genbase	**(0.9970)**	0.0016
	Yeast	**0.5865**	0.3234
Entropy	Blood type	**0.9823**	3.7781
	Blood Cancer	**(0.9120)**	1.2461
	STARKEY"93	**1.1834**	2.2341
	Genbase	**0.9526**	2.1230
	Yeast	**1.2321**	1.0923

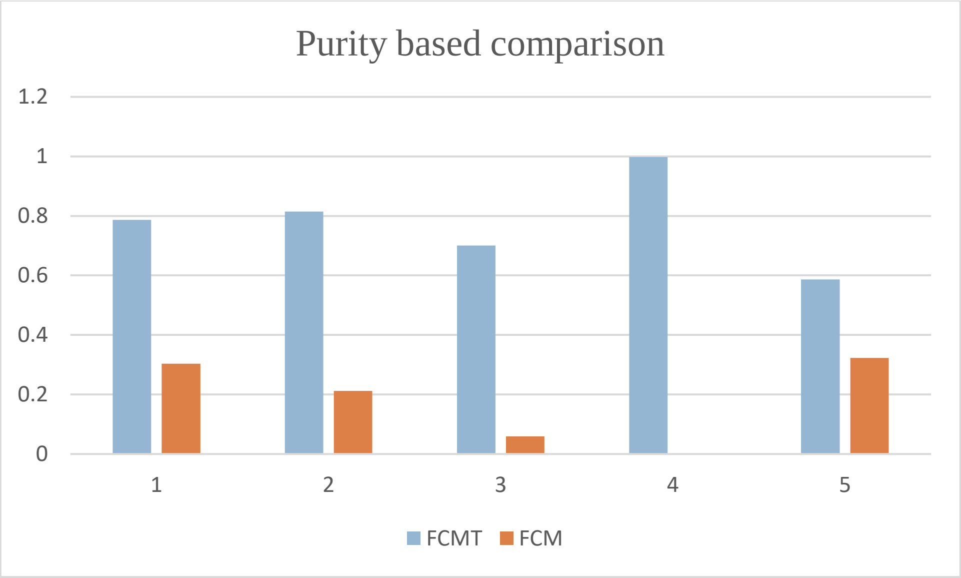


Figure 5.12. Purity result: FCM vs. $_{\mathrm{T}}$FCM for the five datasets

UUM
Universiti Utara Malaysia

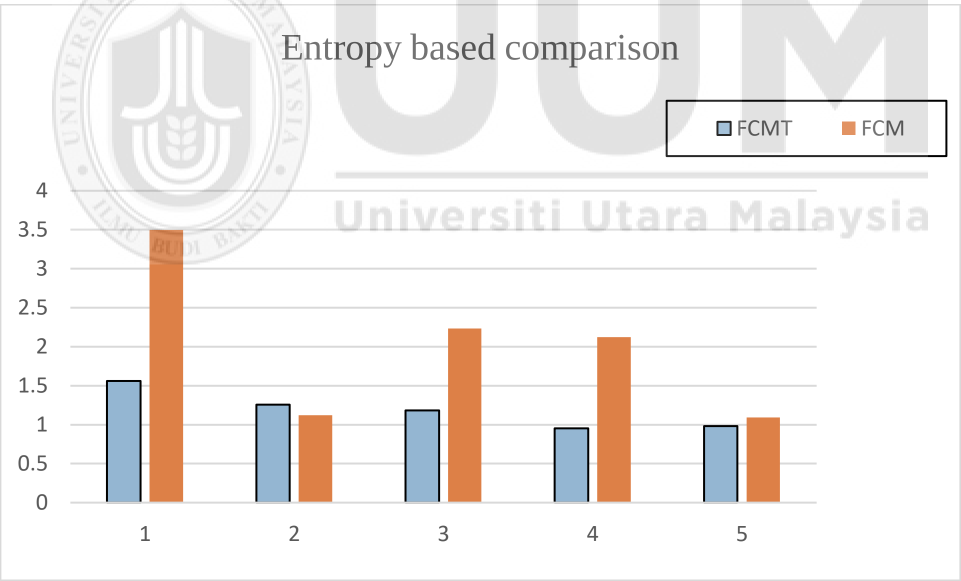


Figure 5.13. Entropy result: FCM vs. $_{\mathrm{T}}$FCM for the five datasets

5.4 Result analysis for $_{T}$FCM vs. FCM

In comparison between $_{T}$FCM and FCM with five experiments on datasets under consideration, it is obvious that $_{T}$FCM is better than the original in terms of general performance. Five criteria that affect the clustering performance are considered.

In terms of iteration count that gives the minimum objective function, it is noted from the previous discussion that $_{T}$FCM has the least iteration count that gives minimum objective function, which leads to high performance of clustering. For dataset 1 and 2, the difference in iteration count value gives the minimum objective function between $_{T}$FCM and FCM due to overlapping. For datasets 3, 4, and 5, the difference is high due to the big size of dataset, which require more iterations to reach the minimum objective function.

In terms of average distance between cluster center and data points, the results show that the $_{T}$FCM is better than the FCM. For dataset 1, 2 and 4, the difference in values of average distance between clusters is quite small between FCM and $_{T}$FCM. This is because of the high overlap in medical data. Such a result indicates that FCM is sensitive to overlapping data which found especially in medical data. For dataset 3 the difference is high because of the big size of data. In dataset 5 difference in average distance is quite high, they are big datasets.

In summary, overlap is a more important factor than size in affecting the clustering performance. The proposed method doing well when applying with overlapped data, which is has difficult clustering when applying FCM.

Universiti Utara Malaysia

The validation of performance indicates that for the purity measure, the difference between the value of $_{T}$FCM and FCM for blood type, blood cancer and Genbase is high, while in the rest datasets the differences are quite high. This is due to the high overlapping found in medical data, which complicates the clustering process and affects performance.

5.5 Accuracy Comparison

The proposed HTM with temporal information is evaluated to study the effectiveness of recognition process. A comparison is made against original Hierarchical Temporal Memory (HTM), $_{T}$S-layer as in Chapter Four and ($_{T}$S-layer + $_{T}$FCM). Table 5.12 illustrates the comparison results; this result shows the robustness of propose temporal HTM in recognizing overlapping patterns. Hence, ($_{T}$S-layer + $_{T}$FCM) improves the performance of recognition process of overlapped data contained in multi- label datasets. Table 5.9 illustrates the accuracy values of original HTM, TS-layer, ($_{T}$S-layer + $_{T}$FCM) and $_{T}$FCM. From this table, it is obvious that HTM with improved Neocognitron and temporal information achieves higher accuracy on overlapping data points. Also, it has the higher accuracy than other algorithms for all datasets under consideration.

Universiti Utara Malaysia

The obtained results of accuracy indicate that the proposed imHTM with ($_{T}$S-layer + $_{T}$FCM) has better recognition. Meanwhile, imHTM is the best for generating accurate clusters based on clustering evaluation metrics purity and entropy. Proposed imHTM also achieves minimum objective function with least iteration count. With the results obtained, the second objective of this research has been achieved.

Table 5.9

Comparison of accuracy values for original HTM vs. $_{T}$S-layer vs. $_{TN}$FCM vs. $_{T}$FCM.

Accuracy in %					
	Blood type	Blood Cancer	STaRKEY"93	Genbase	Yeast
Original HTM	84	84.4	87	83.3	90.8
$_{T}$S-layer	96.2	96.9	98.8	97	98.2
$_{T}$HTM	96	98	96	96.7	97
$_{T}$S-layer + $_{T}$FCM	**98**	**99.4**	**99**	**98.9**	**99.3**

The obtained results of accuracy indicate that the proposed imHTM with ($_{T}$S-layer + $_{T}$FCM) have better recognition. Meanwhile imHTM is the best for generating accurate clusters based on clustering evaluation metrics purity and entropy. Proposed imHTM also achieves minimum objective function with least iteration count. With the results obtained, the second objective of this research has been achieved.

Universiti Utara Malaysia

5.6 Summary

This chapter presents dataset clustering using improved fuzzy c-means clustering algorithm with temporal information. Experiments were conducted on 500 blood type images, 33 blood cancer images and three multi- label datasets. Comparisons were made against HTM algorithm presented in Chapter Four and original HTM and combinations of THTM and $_{T}$S-layer algorithms. Purity and entropy evaluation metrics were applied to verify the performance of THTM.

It is learned that the proposed THTM generates better results than the original in terms of minimum objective function and least iteration counts. Comparison of accuracy indicates that proposed THTM provide better recognition of patterns.

Universiti Utara Malaysia

CHAPTER SIX
LIMIT- CHEBYSHEV DISTANCE MEASURE FOR FUZZY C-MEANS WITH TEMPORAL INFORMATION

6.1 Introduction

As mentioned previously (see Chapter one, section 2.3), the distance metric employed in original FCM is the Euclidean distance (E.D), which has a weakness to data correlation, which means that it is difficult to establish when there is a correlation in data points (Cai, Chen, & Zhang, et al.,2010). In addition, in FCM the employment of Euclidean distance in determining the cluster centers and then data points relate to the class of specific cluster center. By using the mentioned distance, this may cause a specific cluster may be out of the data set (cluster center value may be far from data points of datasets), this will cause that the distance value is high leads to empty cluster (Chattopadhyay et al., et al.,2011). Also, the parameter estimate resulting from an objective function based on Euclidean metric may not be robust with strong correlation between neighboring pixels. E.D is also sensitive to outliers, and high skew will throw off the mean and alter covariance (Lai & Garibaldi, et al.,2013). The nonlinear relationships of data will not be accurately measured, resulting in a low correlation. Upon the completion of clustering using FCM, each data point will be associated with a membership value for each class (Das et al., et al.,2014). Each data point is also assumed to be independent of every other data point and spatial interaction between data points is not considered. However, for medical data, there is strong correlation between neighboring pixels.

Universiti Utara Malaysia

In this chapter, an improved Chebyshev distance metric is proposed to replace the existing Euclidean distance employed in FCM algorithm in order to produce a better clustering as mentioned in Chapter Three (section 3.2.4). To achieve the goals of this improvement, the algorithm in Figure 6.1 is applied. This algorithm shows the steps of implementing the iFCM with limit Chebyshev distance metrics.

```
Improved distance metric
1:  Initialize $\mu = [\mu_{ij}]^0\ matrix$ of membership
2:  Choose parameter $l$>0 to stop the iteration. Set the iteration counting parameter $l$ equal to 0
3:  for $i$= 1 to $x$
4:      At k-step calculate the centers vectors $v^k$=[$v_j$] by
            $\boldsymbol{v_j} = (\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} (\mu_{\boldsymbol{ij}})^{\boldsymbol{m}} \boldsymbol{x_i}) / (\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} (\mu_{\boldsymbol{ij}})^{\boldsymbol{m}}),\ \forall_{\boldsymbol{j}} = \mathbf{1, 2, 3} \ldots . \boldsymbol{c}$
5: apply the proposed distance
    $\boldsymbol{dxy} = \lim_{\boldsymbol{p} \to \infty} (\boldsymbol{max_k}(|\boldsymbol{x_{ik}} - \boldsymbol{x_{jk}}|)^{\mathbf{1}/\boldsymbol{p}})$   $\infty < \boldsymbol{p} > \mathbf{2}$ within the membership
    function in the next step
6:  Update the membership matrix $[\mu_{ij}]^k$ , $[\mu_{ij}]^{k+1}$ by
            $\mu_{\boldsymbol{ij}} = \mathbf{1} / \sum_{\boldsymbol{k=1}}^{\boldsymbol{c}} \boldsymbol{dxy}$
7:  if  $[\mu_{ij}]^{k+1}$- $[\mu_{ij}]^k > l$
8:          Goto step 5
9:      else
10:         stop at some iteration $l$
11: end if
```

Figure 6.1. iFCM with limit Chebyshev

Universiti Utara Malaysia

6.2 Results

This section presents the results obtained by applying improved Chebyshev distance metric in FCM algorithm on the datasets mentioned in Chapter Three (section 3.1). The minimum objective function value will be referred to as (MOF), and the iteration count will be referred to as (IC).

6.2.1 Dataset 1

In this section, blood type images are used. The data is as mentioned in Chapter Four. A comparison of results is obtained between original and FCM with improved Chebyshev ($_C$FCM) and conventional one.

6.2.1.1 $_C$FCM vs. Conventional FCM

In this section, the proposed Chebyshev distance metric is applied under the FCM clustering algorithm, and a comparison is also made between the FCM clustering algorithm factors, IC and MOF, when FCM is applied with different distance metrics - Euclidean, Manhattan, Minkowski, standard Chebyshev – and the proposed Chebyshev (Limit-Chebyshev). From the results tabulated in table 6.1, 6.2 and 6.3 which illustrate the IC and MOF that associated with the three-test image (a), (b) and (c), it is seen that the accuracy of clustering can be seen in the form of minimizing the objective function with the progress of clustering when the data point is close to the cluster center, leading to a more robust and accurate clustering. It is noted that the improved FCM minimizes the objective function with least iteration count than it is in original FCM. In table 6.1 it can be seen the (MOF), which is signed in bold style for each distance metric and its (IC) for Limit-Chebyshev has the MOF value (6.97192) with less IC (6) with comparison of the other four distance metrics. Figure 6.2 illustrates the graphical presentation of the progress

Universiti Utara Malaysia

of the objective function of each of the five-distance metrics. From this figure, it can be seen that the value of the (MOF) of the three images for Euclidean distance in S1, the value is multiplied by (10^4), is between 0.7 and 0.8 at IC 20, between 6 and 6.5 at IC 12, between 2 and 3 at IC 18 for three images respectively. In S2 (multiplied by 10^5), between 2 and 3 at IC 35, between 6 and 6.5 at IC 40, and close to 0.7 for Manhattan distance at iteration count 20, respectively. In S3 (multiplied by 10^4) for Minkowski distance it is 0.007 at IC 29, 6 at IC 37, 25 at iteration count 18 for three images respectively. In S4 (multiplied by 10^4) between 0.0033 and 0.0004 at IC 20, between 11 and 12 at IC 25, between 2 and 2.5 for Chebyshev distance at iteration count 12. For the improved Chebyshev distance in S5 between 6 and 7 at IC 6, between 0.3 and 0.4 (multiplied by 10^4) at IC 8, close to 2 (multiplied by 10^4) at iteration count 5.

UUM
Universiti Utara Malaysia

It is obvious from the previous discussion that the improved Chebyshev has the minimum objective function with a smaller iteration count, for the three test images. Also, it can be concluded that Manhattan distance metric has the biggest OF and IC among all distance metrics used in this research. Table 6.4 illustrate a summary of values MOF and IC

Table 6.1

Value of iteration count and objective function of (a) image with the five-distance metrics

IC	Objective Function				
	Euclidean	Manhattan	Minkowski	Chebyshef	Limit-Chebyshev
1	8968.421	330854.0	95.81077	42.89534	8.99534
2	7228.506	270770.6	77.71941	34.99988	8.99322
3	7227.483	270713.2	77.68835	34.97578	7.97578
4	7227.321	270706.3	77.68366	34.97377	7.97677
5	7227.261	270705.0	77.68271	34.97219	6.97219
6	7227.226	270704.7	77.68252	34.97192	**6.97192**
7	7227.204	270704.6	77.68249	34.97377	6.97192
8	7227.189	270704.5	77.68248	34.97219	6.97192
10	7227.177	270704.5	77.68248	34.97180	
15	7227.161	270704.5	77.58248	34.97219	
20	**7227.155**	270704.5	77.57248	**33.97192**	
29		**270704.4**	**76.58248**	33.97192	
30		270704.4		33.97192	

Universiti Utara Malaysia

Table 6.2

Value of iteration count and objective function of (b) image with the five-distance metrics

IC	Objective Function				
	Euclidean	Manhattan	Minkowski	Chebyshef	Chebyshev
1	104423.553445	184838.967828	761058.172910	630060.444920	7734.463654
2	104405.000030	47835.463296	719408.508527	502703.208518	7634.463654
3	104248.744218	47246.851449	71062.851710	442090.768177	7534.463654
4	102981.124027	40306.646272	69893.894533	493893.330832	6734.463654
5	94861.998395	03519.005554	68048.378145	310819.862315	5539.707970
6	73387.812974	65975.592830	65854.465869	315614.860098	5020.144195
7	63341.387360	61078.140748	61762.714464	217725.401015	5744.427414
8	62600.366580	60880.389627	61742.852147	216352.399441	**3236.062344**

12	62569.991671	60871.341392	61561.950379	215528.522331
20		60870.538299	61431.405031	214577.462909
22		60870.435059	61350.994678	214541.183091
23		60870.420585	61258.722205	213595.768855
24		60870.418525	61161.798160	212596.179643
25		60870.418231	51963.031607	118596.063955
30		60870.418189	61962.528561	118596.063955
35		60870.418183	61863.729365	**118596.063955**
37		60670.418183	**61763.810648**	
40		**60370.418183**		

UUM Universiti Utara Malaysia

Table 6.3

Value of iteration count and objective function of (c) image with the five-distance metrics

IC	Objective Function				
	Euclidean	**Manhattan**	**Minkowski**	**Chebyshef**	**Limit-Chebyshev**
1	92419.816678	761959.063467	660812.410366	56695.740164	68378.206317
2	70302.138279	762558.117394	349313.855648	45556.901317	37052.487240
3	73125.594008	756328.463967	349278.166098	45512.641877	34565.654993
4	61636.172515	861202.589528	390815.570083	35284.511849	26342.398115
5	59074.085003	753567.506692	287648.025341	35135.490344	**21266.035000**
6	35206.745129	752770.104481	255275.755644	34775.163324	21266.035000
7	29735.180647	758032.409381	507331.393406	34170.862271	
8	28242.482542	751343.532574	**250170.192635**	33488.148557	
10	27830.391981	752932.933481	501741.348147	32965.518693	
11	27718.837808	753633.980008	502535.423206	**22666.272495**	
12	27689.070816	753933.361223	250288.018488	42527.367096	
13	27681.191665	744059.005636	503025.471052		
14	427679.114971	744112.836033	250308.543955		
15	27678.568829	**734135.568192**	250310.489836		
16	27678.425365	764144.295824	250311.160854		
17	**27678.387701**	764148.630854	250312.060636		
18	27678.568829	764147.630854	250312.690573		
19		764150.756548			
20		753567.506692			

Universiti Utara Malaysia

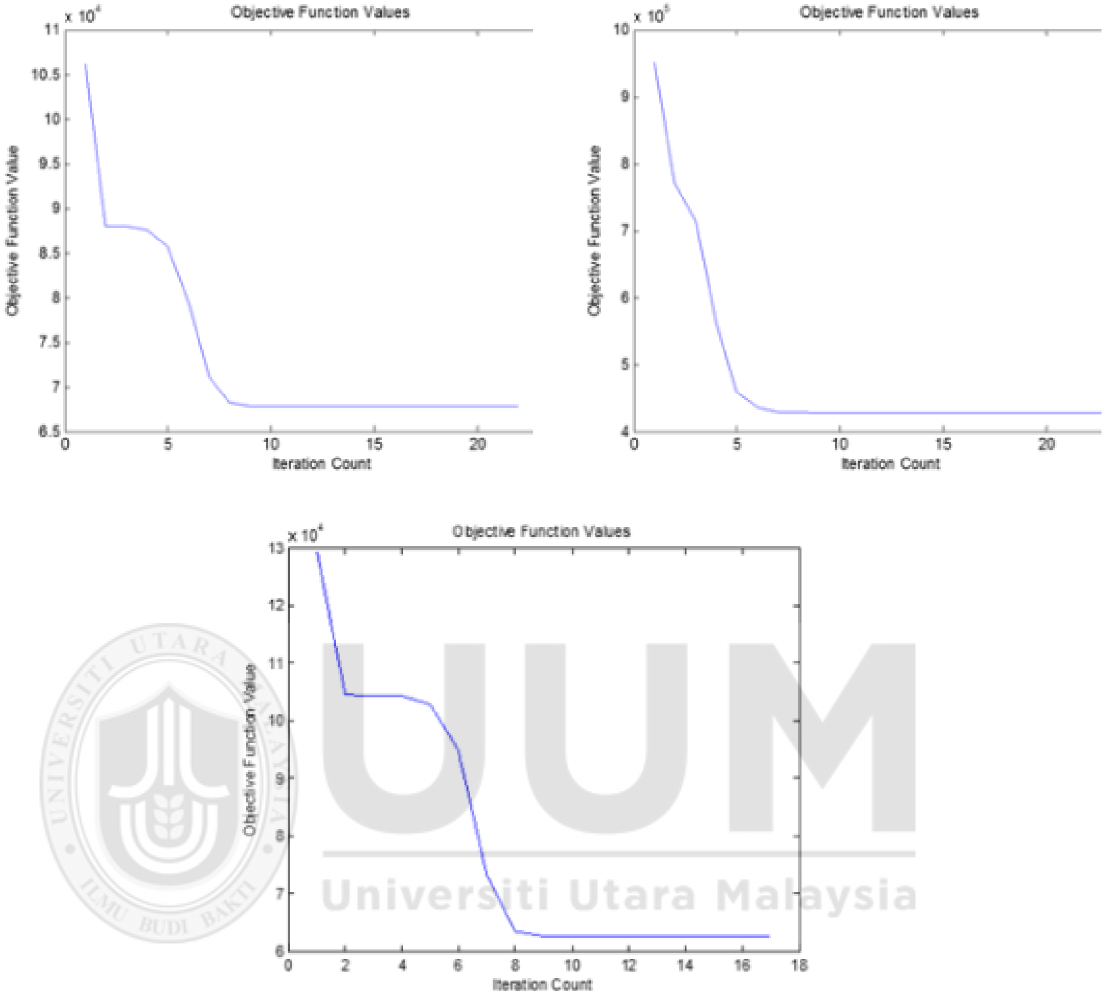
From table 6.4 it can be seen that image (a) has the smaller value of (OF) and IC. This is due to the high image cluster density which is illustrated in figure 6.4. From this figure, it is obvious that image (a) has the higher density, which results in minimum OF.

Table 6.4

Summary of iteration counts that leads to the minimum objective functions for dataset 1

Distance Metrics		**Images**			
		a	**b**	**c**	**All dataset**
Euclidean	**IC**	20	12	17	16
	MOF	7227.15	62569.991671	27678.387701	76453.444
Manhattan	**IC**	29	40	15	27
	MOF	270704.4	60370.418183	734135.568192	865432.4456
Minkowski	**IC**	29	24	8	20
	MOF	76.58248	61161.798160	250170.192635	6454.22334
Chebyshev	**IC**	29	25	11	23
	MOF	33.97192	118596.063955	22666.272495	4567.34177
Limit-Chebyshev	**IC**	6	8	5	**6**
	MOF	6.97192	3236.062344	21266.035000	**34.2234**

Universiti Utara Malaysia

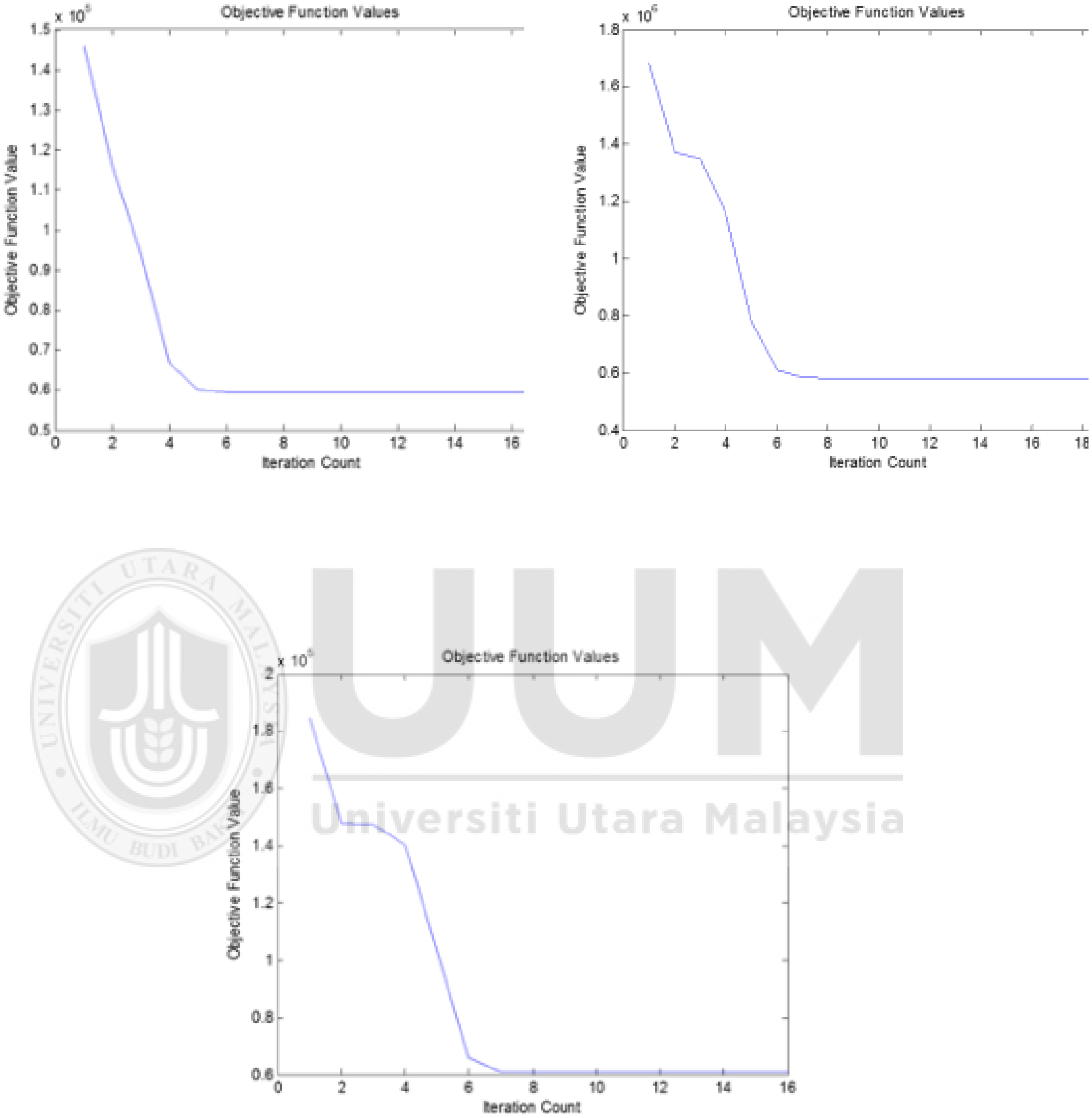


UUM
Universiti Utara Malaysia

S1

Figure 6.2. Progress of the objective function for the three test images for the sample images with (S1): Euclidean distance, (S2): Manhattan distance, (S3): Minkowski distance, (S4): original Chebyshev distance and (S5) improved Chebyshev distance.

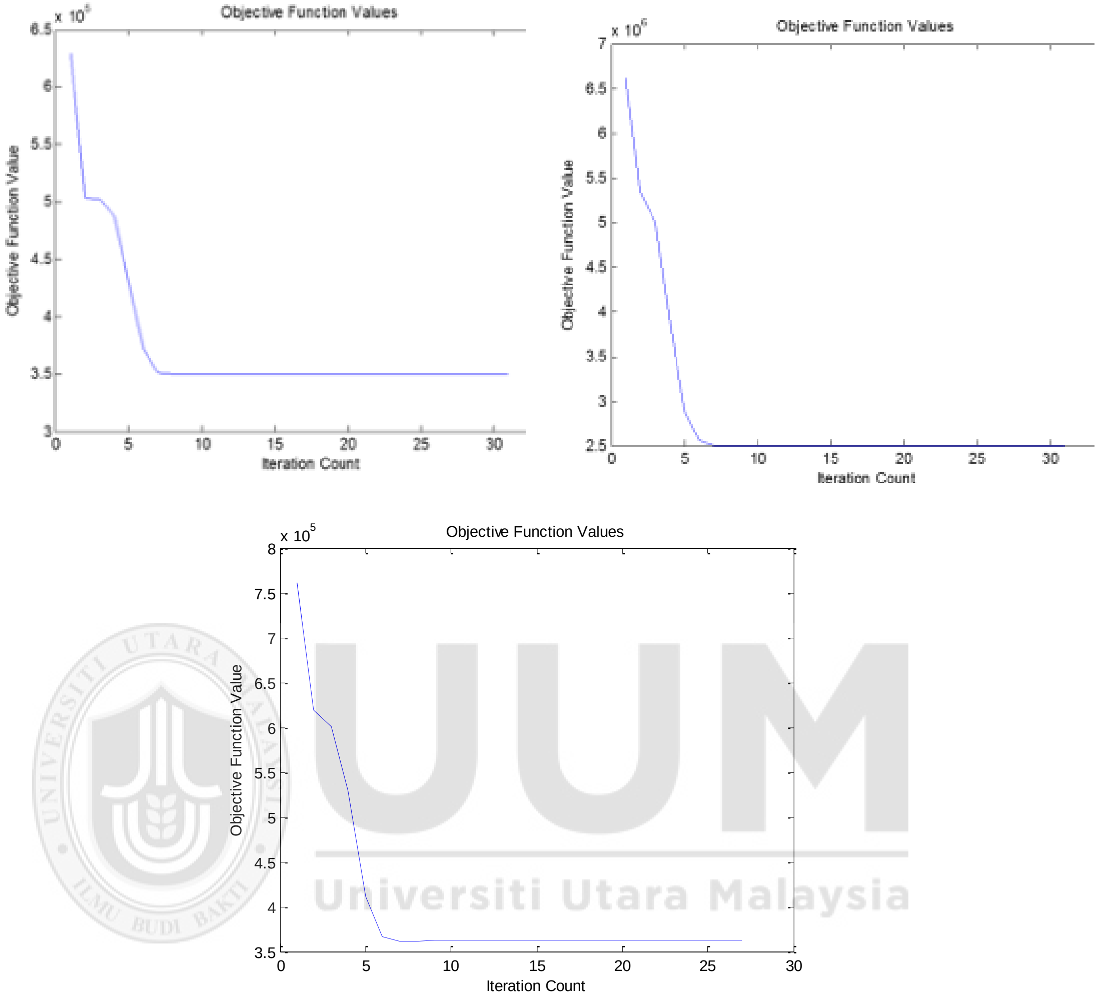
Figure 6.2 continued



UUM
Universiti Utara Malaysia

S2

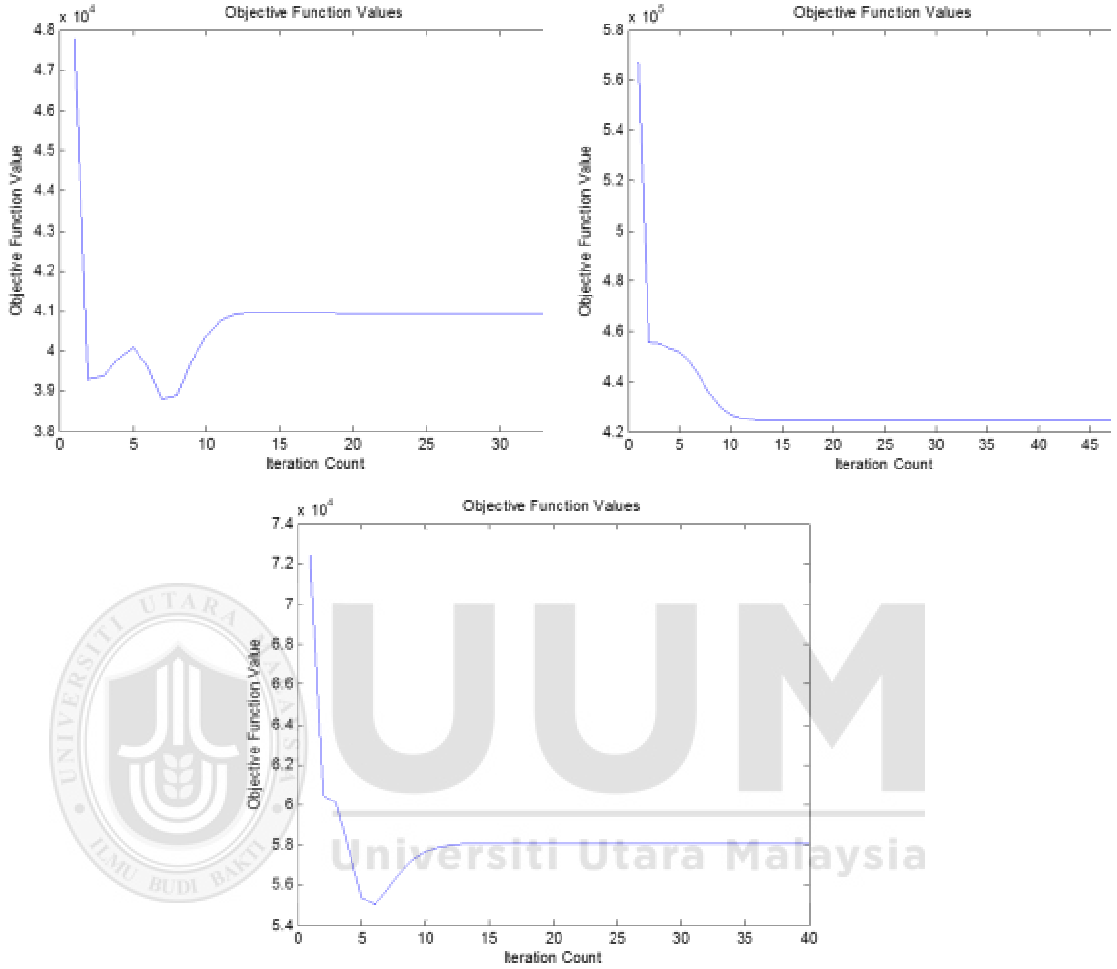
Figure 6.2 continued



S3

UUM
Universiti Utara Malaysia

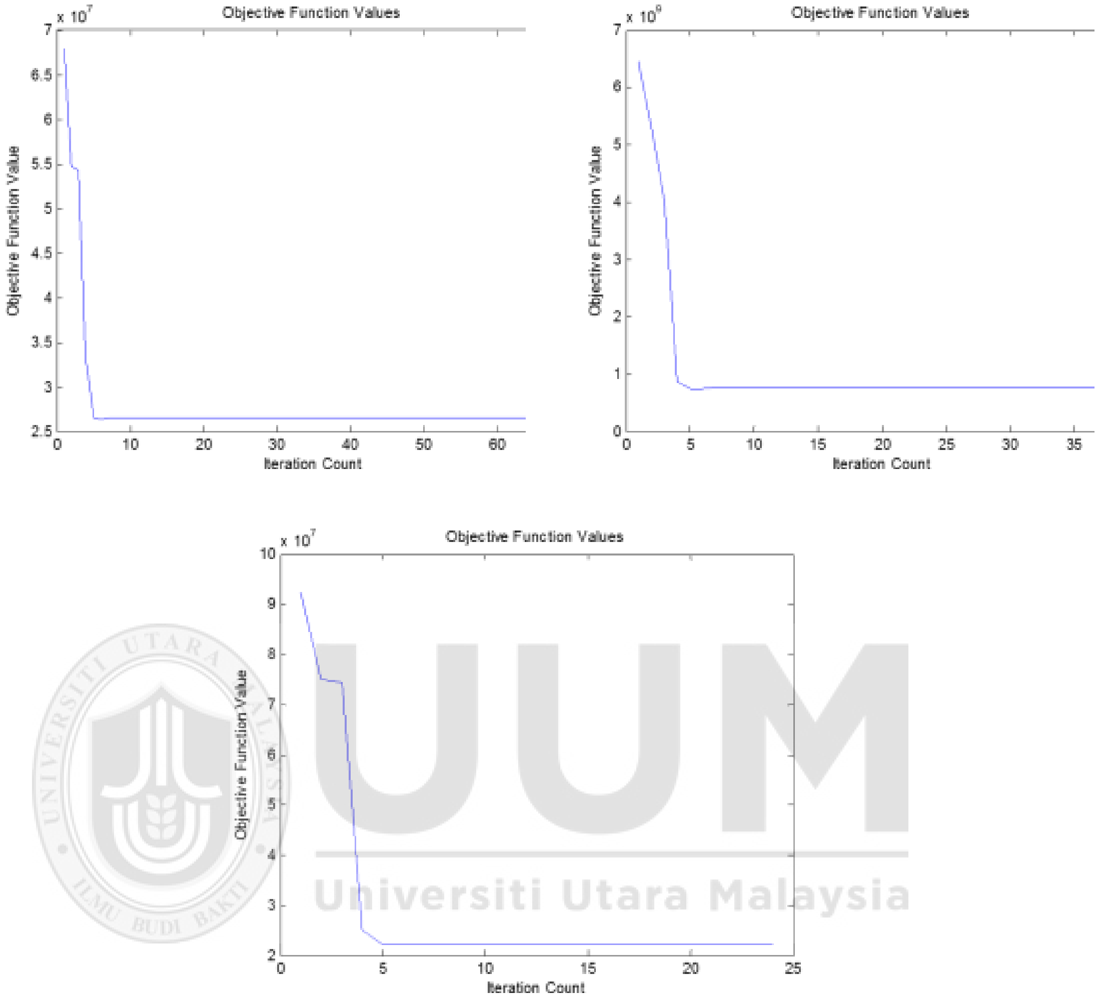
Figure 6.2 continued



S4

UNIVERSITI UTARA MALAYSIA
ILMU BUDI BAKTI
UUM
Universiti Utara Malaysia

Figure 6.2 continued



UUM Universiti Utara Malaysia

S5

In Figure 6.3 the density of clusters of image (a) is higher than the other images, and its objective function is the minimum among them. This is because the cluster density affects the value of objective function, then the accuracy of the clustering (Xiaojun Lou et al., 2012).

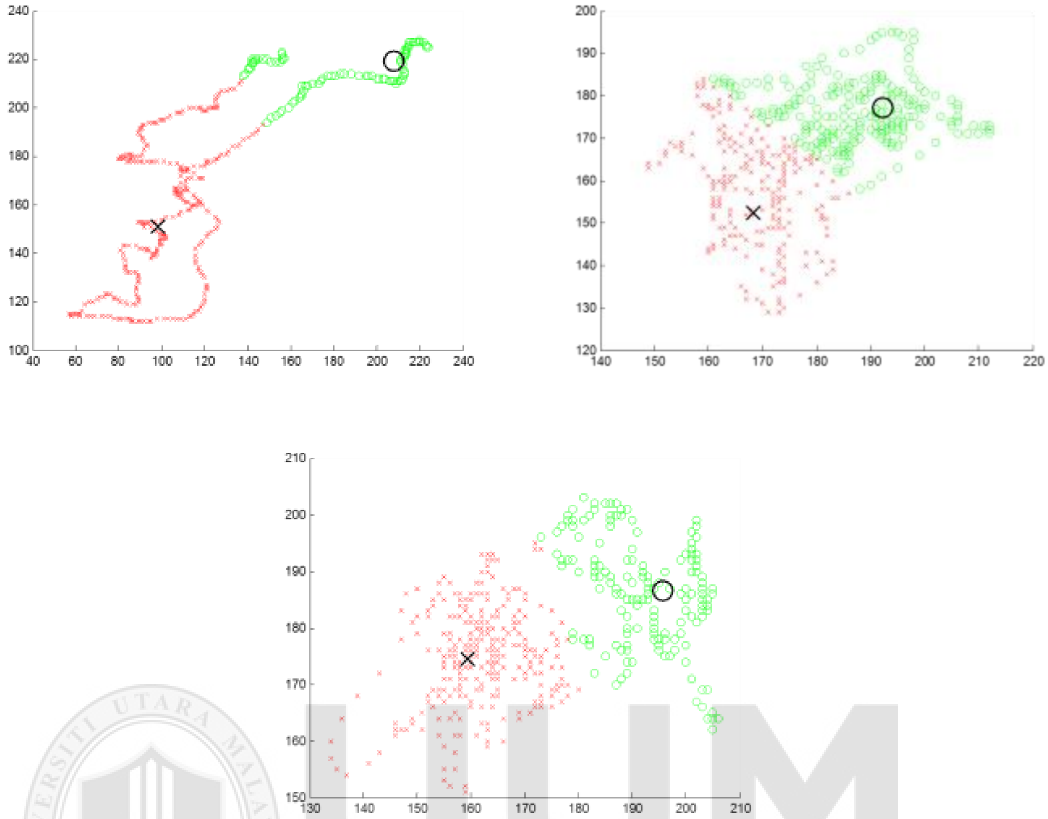


Figure 6.3. Cluster density for the three blood type sample images

UUM
Universiti Utara Malaysia

6.2.2 Dataset 2

In this section, blood cancer is examined as mentioned in Chapter Four. A comparison of results is obtained between original and improved FCM.

6.2.2.1 CFCM vs. Conventional FCM

From the results tabulated in tables 6.5, 6.6 and 6.7, which illustrate the IC and MOF that associated with the three-test image (a), (b) and (c), it is noted that the proposed FCM with Limit-Chebyshev metric distance minimize the objective function with least iteration count than it is in FCM with the other four distances used in this research. In table 6.5, which represents image (a), the value of MOF for Limit-Chebyshev has a value of

(0.50033) with IC (9), which is the lowest value in comparison with the other four distance metrics, which are (1.96844) for original Chebyshev at IC (16), (1.16510) at IC (20) for Miknowski, (120.5628) at IC (16) for Manhattan and (1.1216) at IC (15) for Euclidean distance. Table 6.6 illustrates the MOF and IC for the second image (b). From this table, it can be seen that the Limit-Chebyshev also has MOF at least value of IC, which are (7.87188) at IC (7). However, Chebyshev has MOF (31.77176) at IC (17), Miknowski has MOF (75.28248) at IC (Bova & G, et), Manhattan is (2707045.409) at IC (26) and (7227.139) at IC (17) for Euclidean distance. Table 6.7 shows the result of applying FCM with the five distances for image (c). In this table, the values of MOF are (1.3408) at IC (12), (3.7880) at IC (18), (1.2300) at IC (Bova & G, et), (124.56) at IC (Bova & G, et) and (17.160) at IC (20) for Limit-Chebyshev, original Chebyshev, Minkowski, Manhattan and Euclidean distance respectively. Table 6.8 summarizes the MOF and IC for this dataset. Figure 6.4 illustrates the graphical presentation of the progress of the objective function of each of the five distances metric for the three images.

Universiti Utara Malaysia

Table 6.5

Value of iteration count and objective function of (a) image

IC	Minimum objective function				
	Euclidean	**Manhattan**	**Minkowski**	**Chebyshef**	**Limit-Chebyshev**
1.	1.5192	168.7255	9.95080	12.95080	0.92031
2.	1.3384	144.7819	8.96374	11.96374	0.90906
3.	1.3362	139.4440	7.96451	11.96451	0.87297
4.	1.3349	131.4547	6.96517	10.96517	0.77222
5.	1.3339	126.2362	5.96572	10.96572	0.77188
6.	1.3335	124.8115	4.36618	9.96618	0.67188
7.	1.3334	124.5767	3.99948	8.96658	0.57188
8.	1.3334	124.5480	3.96691	8.96641	0.50884
9.	1.3339	124.5486	2.76726	8.96426	**0.50033**
10.	1.3349	124.5532	2.86752	7.96752	
11.	1.3212	124.5571	2.96771	7.96721	
12.	1.3213	124.5595	1.16787	6.96787	
13.	1.3127	124.5627	1.26800	6.96600	
14.	1.3127	124.5627	1.36810	6.96210	
15.	1.3127	124.5628	1.26819	5.96819	
16.	1.3127	**120.5628**	1.36810	**1.96844**	
17.	1.3127		1.26819		
18.	1.3126		1.26720		
19.	1.3116		1.26619		
20.	1.3103		**1.16510**		
21.	1.3103				
22.	1.2123				
23.	1.2100				
25.	**1.1216**				

Universiti Utara Malaysia

Table 6.6

Value of iteration count and objective function of (b) image

IC	Objective Function				
	Euclidean	**Manhattan**	**Minkowski**	**Chebyshef**	**Limit-Chebyshev**
1.	8762.484	3281060.641	93.85491	42.92031	22.92031
2.	7228.703	2707561.110	77.70487	35.00906	20.00906
3.	7227.476	2707161.496	77.68508	34.97297	18.97297
4.	7227.296	2707111.218	77.68328	34.97222	15.97222
5.	7227.235	2707093.653	77.68272	34.97188	10.97188
6.	7227.206	2707083.604	77.68255	34.97188	7.97188
7.	7227.188	2707076.291	77.68255	34.97177	**7.87188**
8.	7227.176	2707070.536	77.68249	34.97176	
9.	7227.167	2707066.204	77.68248	34.87176	
10.	7227.16	2707062.717	76.68248	33.97176	
11.	7227.155	2707060.010	75.68248	32.67176	
12.	7227.151	2707057.856	75.58248	33.97176	
13.	7227.148	2707056.006	75.48248	32.97176	
14.	7227.145	2707054.074	75.28248	32.87176	
15.	7227.143	2707052.180	74.68248	32.77176	
16.	7227.142	2707050.591	75.58248	31.97176	
17.	**7227.139**	2707049.311	75.38248	**31.77176**	
25.		2707048.379	**75.28248**		
26.		**2707045.409**			

Table 6.7

Value of iteration count and objective function of (c) image

IC	Objective Function				
	Euclidean	**Manhattan**	**Minkowski**	**Chebyshef**	**Limit-Chebyshev**
1.	21.042	168.72	1.5192	4.6555	1.5192
2.	18.650	144.78	1.4192	3.8052	1.3384
3.	18.643	139.44	1.3384	3.8047	1.3334
4.	18.641	131.45	1.3334	3.8055	1.3334
5.	18.639	126.23	1.3334	3.8061	1.3339
6.	18.634	124.81	1.3339	3.8091	1.3349
7.	18.276	124.57	1.3349	3.8099	1.3425
8.	18.123	124.54	1.3425	3.8108	1.3388
9.	17.988	124.54	1.3388	3.8116	1.3331
10.	17.768	124.55	1.3331	3.8120	1.3432
11.	17.765	124.55	1.3432	3.8116	1.3432
12.	17.764	124.55	1.3408	3.8099	**1.3408**
13.	17.762	124.56	1.3410	3.8070	
14.	17.762	124.56	1.3411	3.8039	
15.	17.761	124.56	1.3411	3.8004	
16.	17.761	124.56	1.3411	3.7960	
17.	17.760	124.56	1.3412	3.7912	
18.	17.760	124.56	1.3392	**3.7880**	
19.	17.760	124.56	1.3384	3.7894	
20.	**17.160**	124.56	1.3334	3.7943	
21.		124.56	1.3234	3.7995	
22.		**124.56**	1.3134	3.8143	
23.			1.2334	3.8145	
24.			**1.2300**	3.8146	

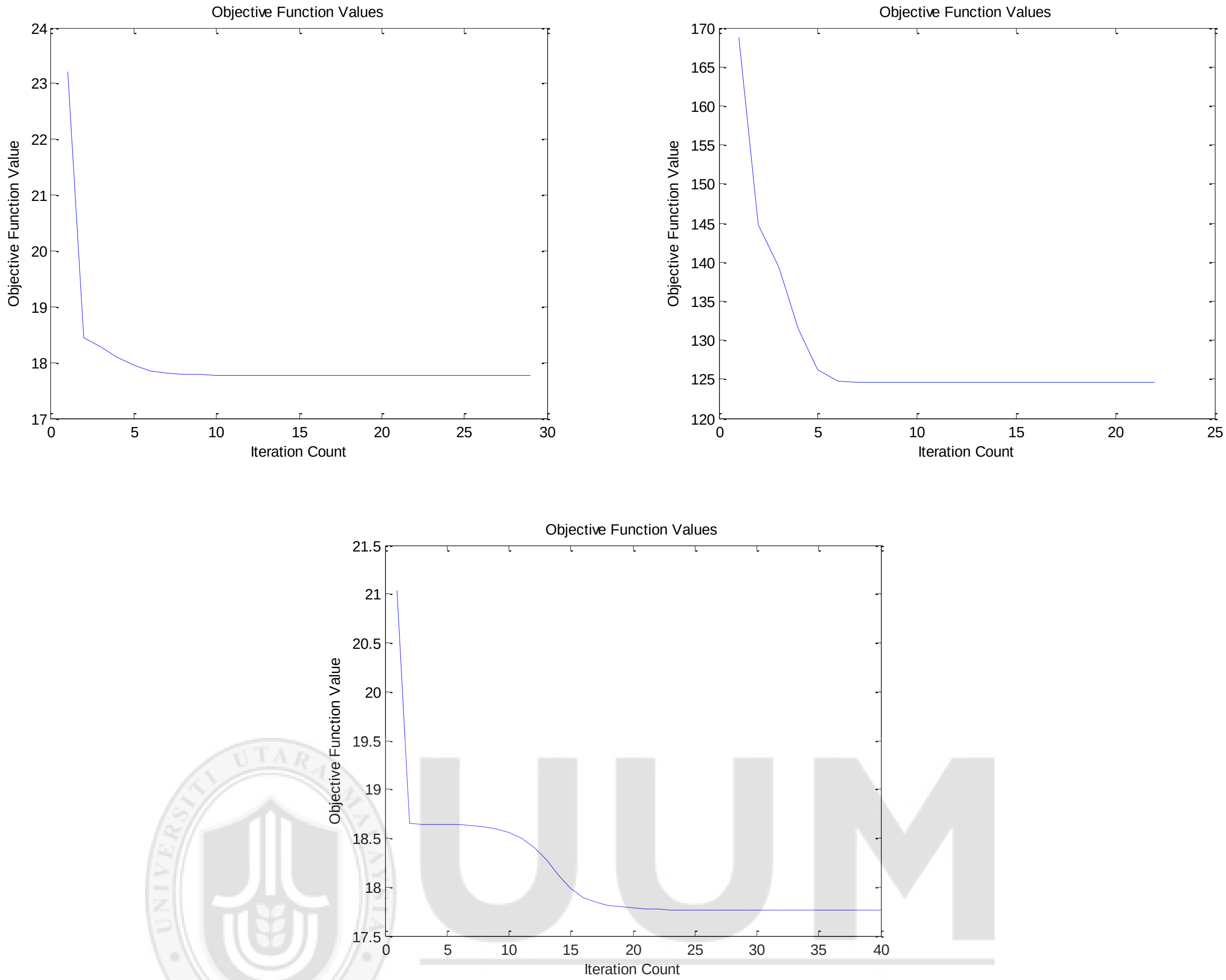
UUM Universiti Utara Malaysia

Table 6.8

Summary of iteration count that leads to the minimum objective function for dataset 2

Distance Metrics		**Images**			
		a	**b**	**c**	**All dataset**
Euclidean	**IC**	25	17	20	23
	MOF	1.1216	7227.139	17.160	231.221
Manhattan	**IC**	16	26	22	25
	MOF	120.5628	2707045.409	124.56	3451.299
Minkowski	**IC**	20	25	24	30
	MOF	1.16510	75.28248	1.2300	30.3331
Chebyshev	**IC**	16	17	18	20
	MOF	1.96844	31.77176	3.7880	10.435
Limit-Chebyshev	**IC**	9	7	12	**8**
	MOF	0.50033	7.87188	1.3408	**1.234**

Universiti Utara Malaysia

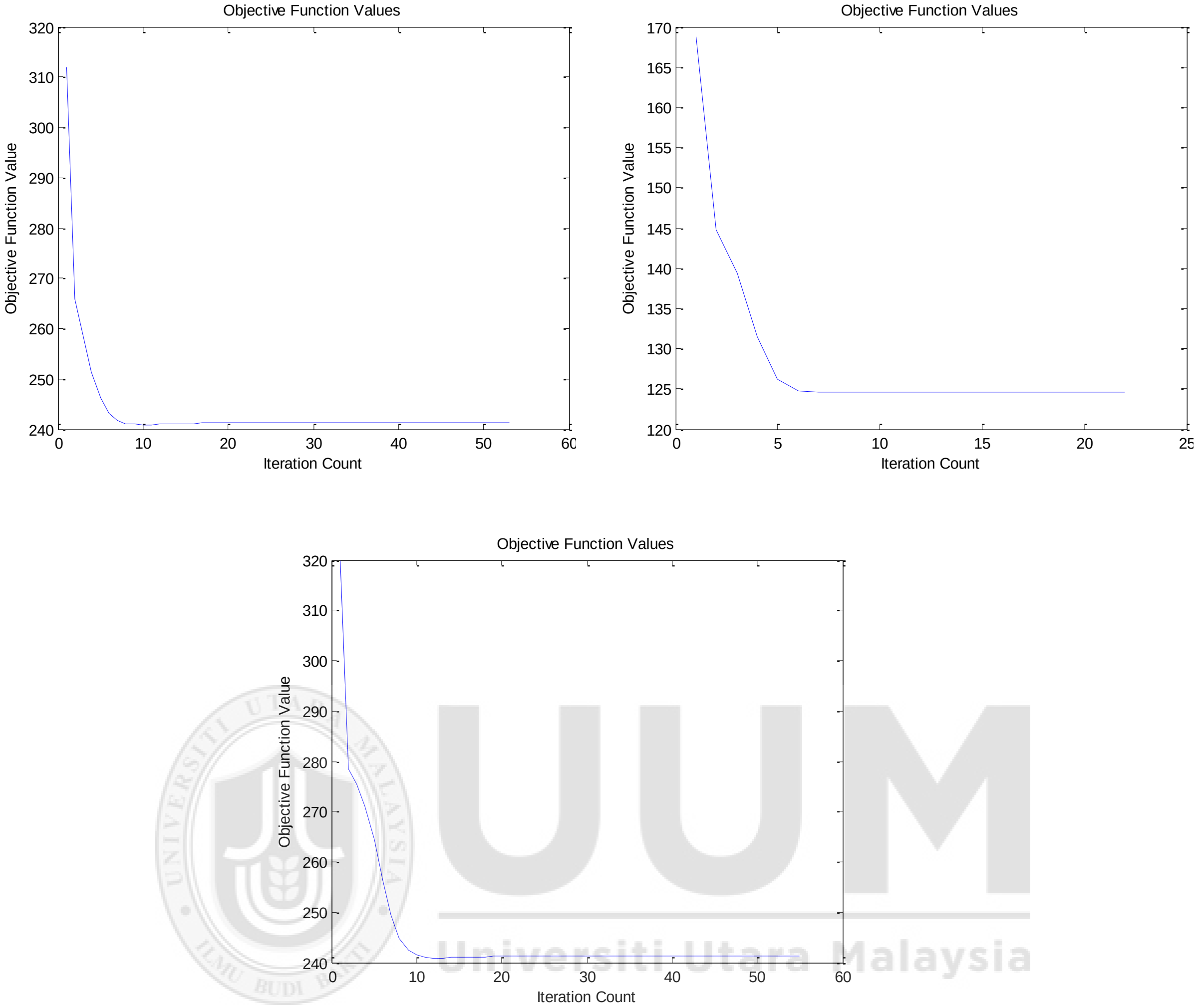


Universiti Utara Malaysia

S1

Figure 6.4. Progress of the objective function for the three test images for the samples images with (S1): Euclidean distance, (S2): Manhattan distance, (S3): Minkowski distance, (S4): Chebyshev distance and (S5) limit-Chebyshev distance.

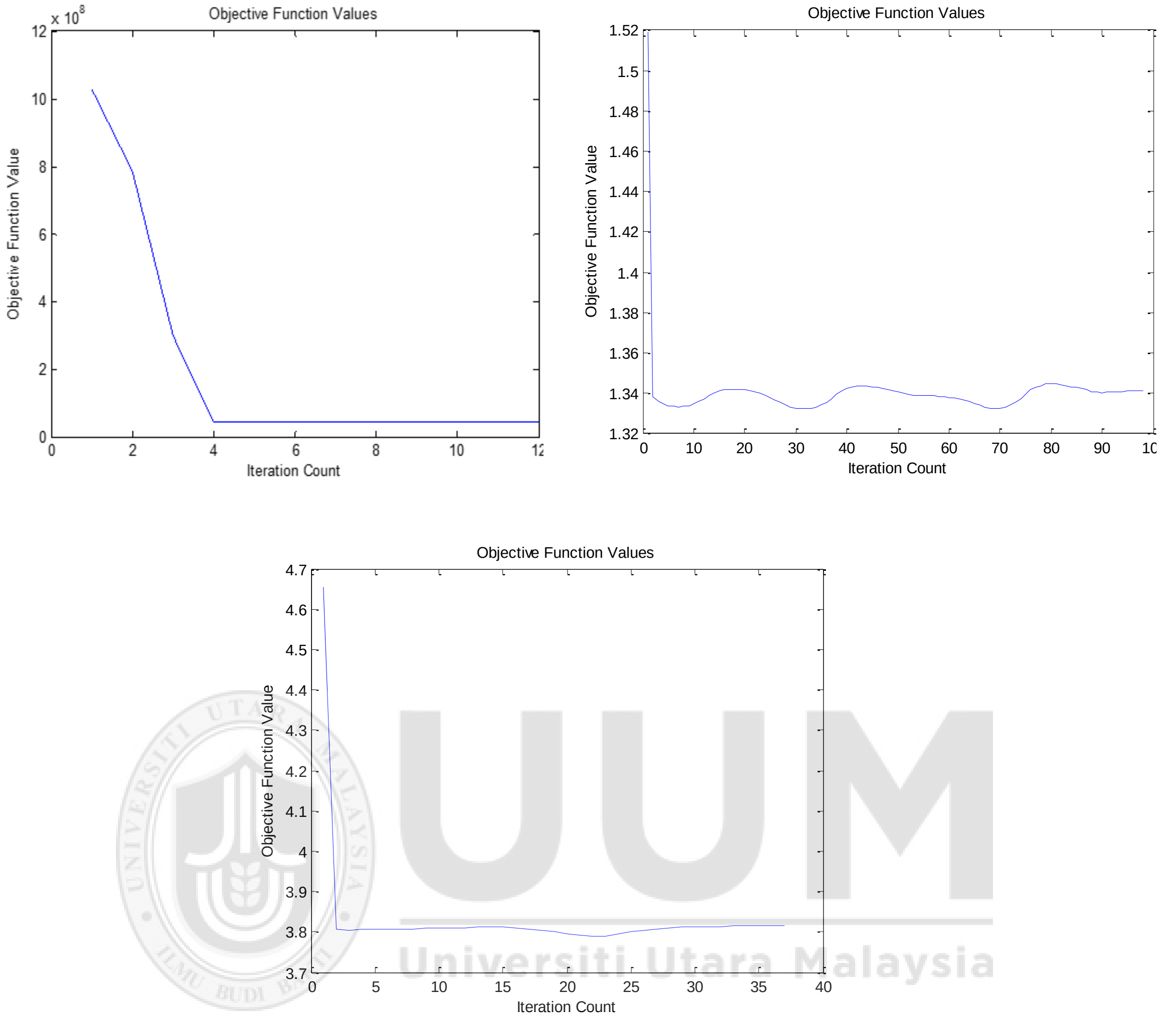
Figure 6.4 continued



UUM Universiti Utara Malaysia

S2

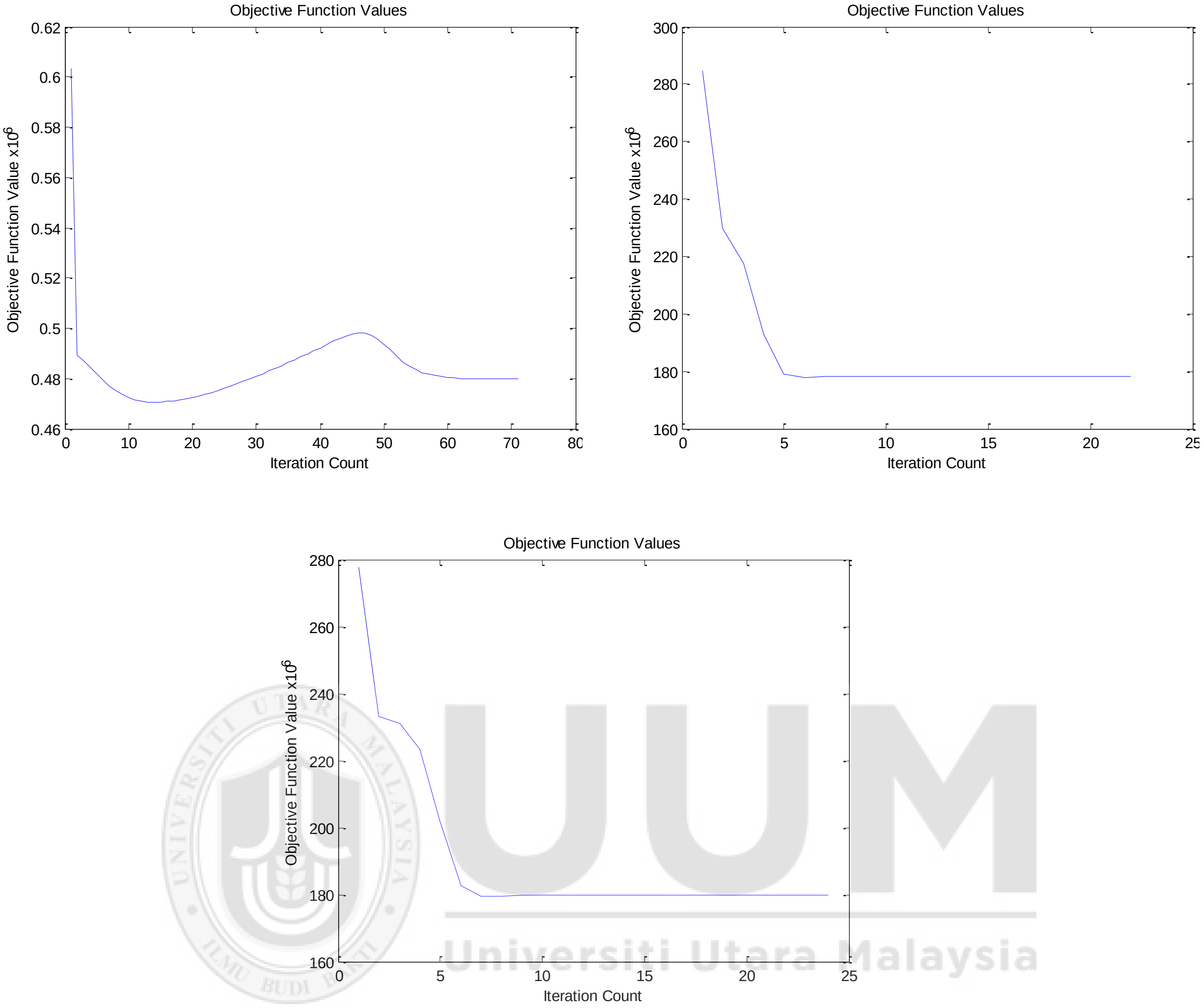
Figure 6.4 continued



S3

UUM Universiti Utara Malaysia

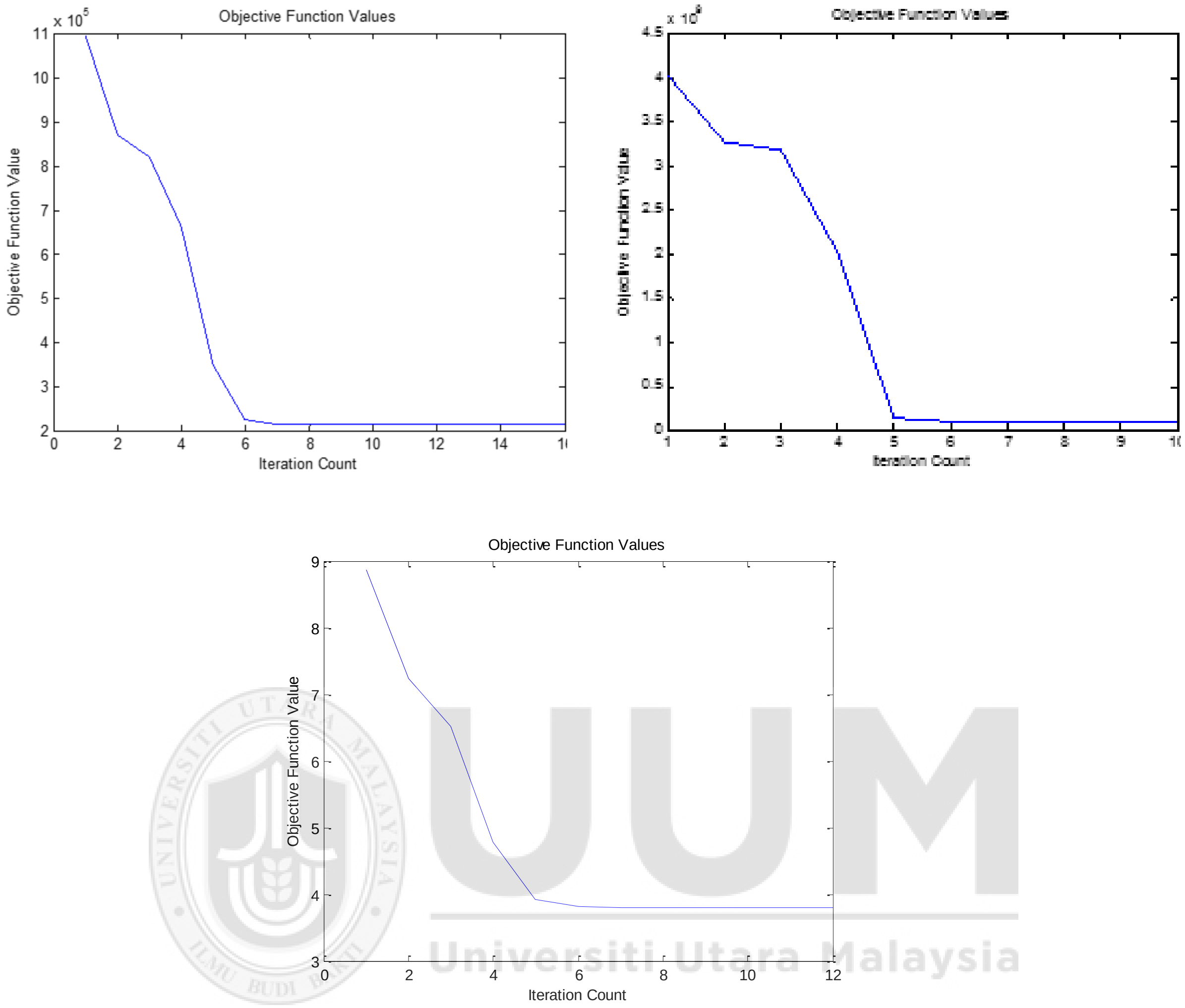
Figure 6.4 continued



S4

UUM Universiti Utara Malaysia

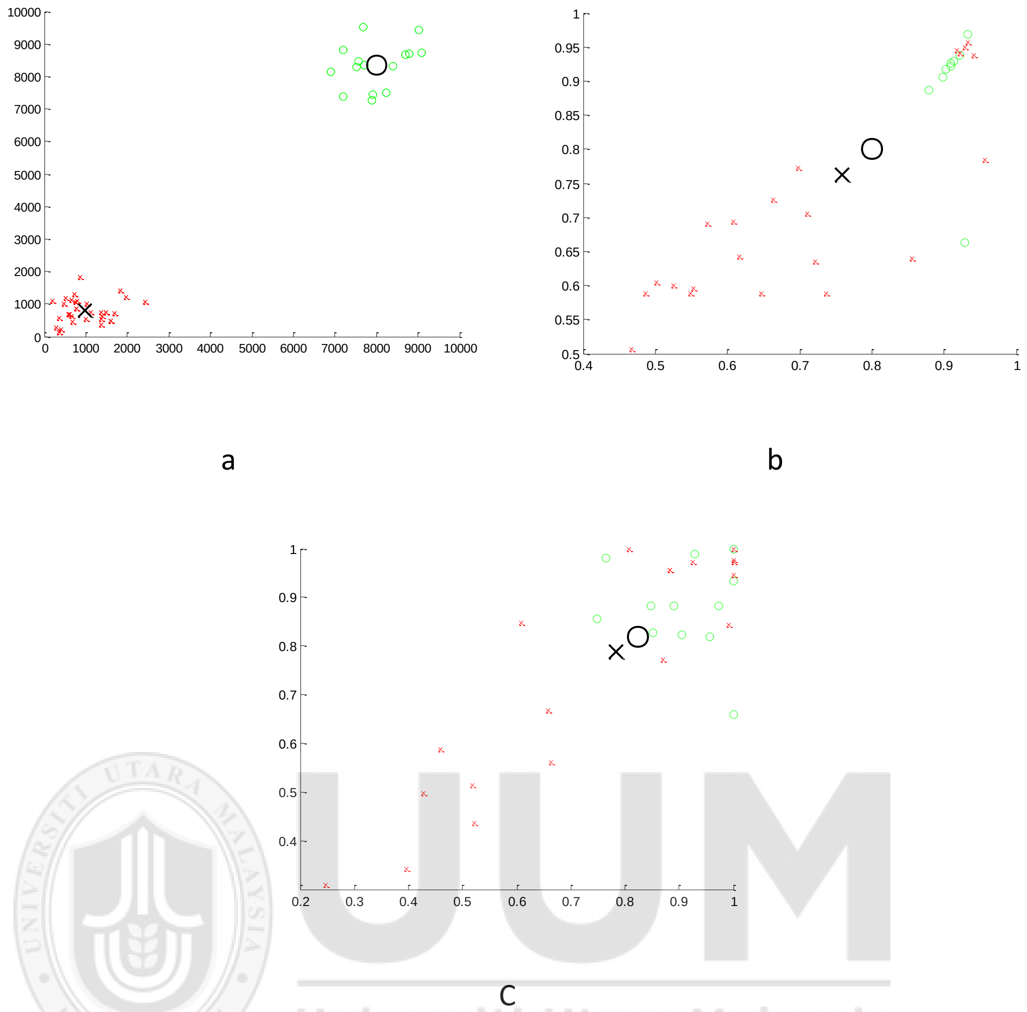
Figure 6.4 continued



S5

Universiti Utara Malaysia

Figure 6.5 shows the cluster density of this dataset. From this figure it is obvious that the difference in cluster density of image (a) is small, which results in having this image the minimum objective function than the other images as mentioned before that the small difference in cluster densities, the minimum objective function. This is clear in the results tabulated in table 6.8, which summarizes the resulting objective functions and iteration counts for the dataset in the state of arts. It is clear that image (a) has the smallest value of MOF.



UUM
Universiti Utara Malaysia

Figure 6.5 Cluster density for the three sample images

6.2.3 Dataset 3

In this experiment: STARKEY”93 dataset is used. A comparison of results is obtained between original and improved FCM.

6.2.3.1 cFCM vs. conventional FCM

From the results tabulated in table 6.9, which shows a comparison of the objective functions values and iteration counts that leads to minimum objective function, it is seen that MOF and IC at their least values are verified with Limit-Chebyshev, it is (0.25398) with IC (10), as it is clear in the table, with comparison by applying FCM with the other four distance metrics used in this research. The values of the MOF and IC for the other four distances are (3.25E+15) at IC (15), (1.853560) at IC (Bova & G, et), (5.60E+13) at IC (23) and (4.12621) at IC (20) for Chebyshev, Minkowski, Manhattan, Euclidean distances respectively. Table 6.10 summarizes the values of MOF and IC.

Figure 6.6 shows the progress of the objective function of each of the five-distance metrics. The figure illustrates that the objective function minimizes with the progress of clustering by using the Limit-Chebyshev distance (refer to Figure 6.6 (e)). In figure 6.6 (a) the value of the objective function is about 4 when used with Euclidean distance at iteration count 20; while in (b) between 5 and 6 by using Manhattan distance at iteration count 23, in (c) close to 2 by using Minkowski at iteration count 22, while in (d) between 3 and 3.5 by using original Chebyshev at iteration count 15 and (e) close to 0.3 by using Limit-Chebyshev metric at iteration count 10.

Universiti Utara Malaysia

Table 6.9

Value of iteration count that leads to the minimum objective function for dataset3

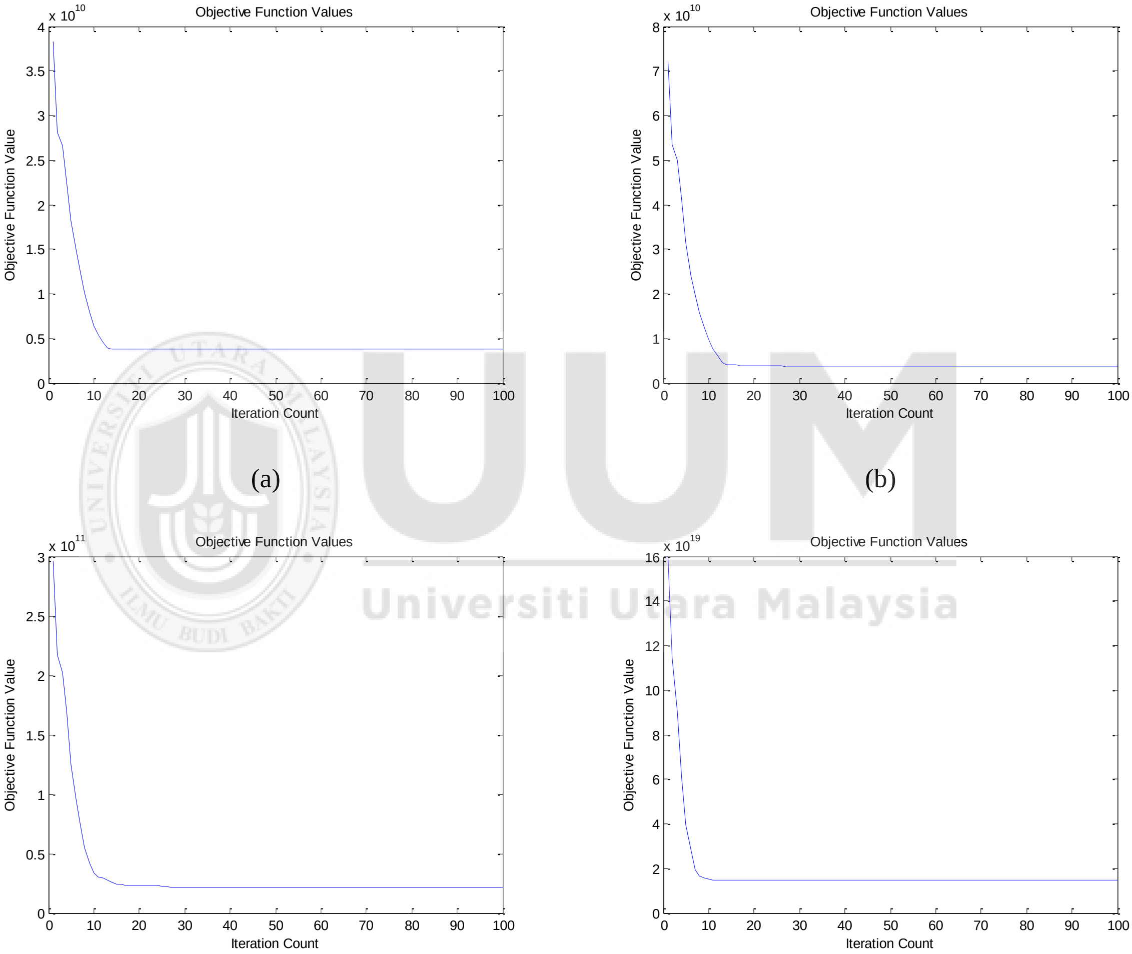
IC	Objective Function				
	Euclidean	**Manhattan**	**Minkowski**	**Chebyshef**	**Limit-Chebyshev**
1	8.42134	3.73E+15	2.953427	4.75E+15	4.75543
2	7.50610	3.35E+15	2.853427	3.46E+15	3.45542
3	6.48311	1.53E+15	2.753427	3.46E+15	3.45542
4	6.32112	5.71E+13	2.653427	3.45E+15	2.45542
5	6.26113	5.71E+13	2.553427	3.36E+15	1.45530
6	5.22613	5.68E+13	2.453427	3.35E+15	1.45520
7	5.22425	5.67E+13	2.353427	3.34E+15	0.56435
10	5.21622	5.67E+13	2.253427	3.33E+15	(0.25398)
15	4.22622	5.65E+13	2.153427	3.25E+15	
20	4.12621	5.63E+13	1.953427		
22		5.61E+13	1.853560		
23		5.60E+13			
24		5.71E+13			
25		5.71E+13			
30		5.71E+13			
35		5.71E+13			
37		5.71E+13			

Universiti Utara Malaysia

Table 6.10

Value of iteration count that leads to the minimum objective function for dataset3

	Euclidean	Manhattan	Minkowski	Chebyshev	Limit-Chebyshev
IC	20	23	22	15	10
MOF	4.12621	5.60E+13	1.853560	3.25E+15	0.25398



(a)

(b)

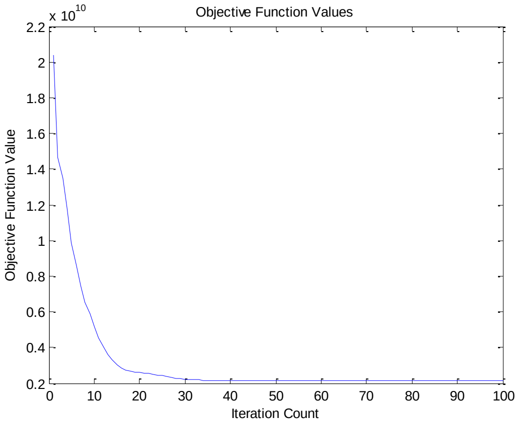
(c)

(d)

Figure 6.6. Progress of the objective function for STARKEY"93 dataset for FCM with (a): Euclidean distance (b): Manhattan Distance, (c): Minkowski distance (d): Chebysheve distance (e): Limit-Chebyshev distance.

UUM
Universiti Utara Malaysia

Figure 6.6 continued



(e)

Figure 6.7 shows the cluster density of this dataset. From this figure, it is obvious that the difference in cluster density is small, which affects the clustering accuracy then the value of minimum objective function.

Universiti Utara Malaysia

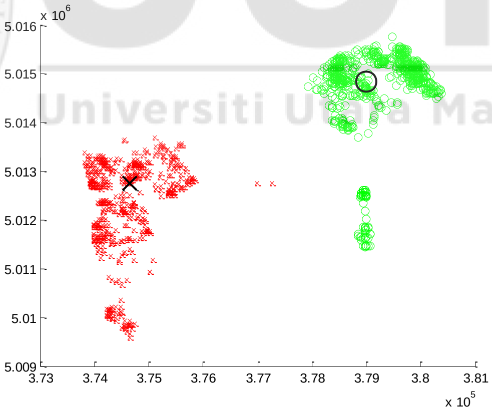


Figure 6.7. Cluster density for dataset 3

6.2.4 Dataset 4

In this section Genbase datasets are used. A comparison of results is obtained between original and improved FCM.

6.2.4.1 $_C$FCM vs. conventional FCM

From the results tabulated in table 6.11, which shoes a comparison of the objective functions values and iteration counts that leads to minimum objective function. In this table, it is seen that the MOF and least IC are verified with Limit-Chebyshev, it is (0.230833) with IC (20), as it is clear in the table, with comparison with applying FCM with the other four distance metrics under consideration. For Chebyshev distance, the MOF is (1.859295) at IC (35), while it is (1.02824743) at IC (37) for Minkowski, for Manhattan it is (1.994467) at IC (80) and last distance is Euclidean which has MOF (1.125) at IC (29). Table 6.12 summarize the values of MOF and IC for each distance metrics. As shown in this table, Manhattan has higher values of MOF and IC. This is because this distance is incompatible with FCM (Grabusts, et al.,2011) and the value of Manhattan distance is the biggest among the other distance (as mentioned in the example demonstrated in Chapter Two (section 2.3). Figure 6.8 shows the progress of the objective function of each of the five-distance metrics. The figure illustrates that the objective function minimizes with the progress of clustering using Limit-Chebyshev distance (refer to Fig. 6.8).

Universiti Utara Malaysia

Table 6.11

Value of iteration count and objective function dataset 4

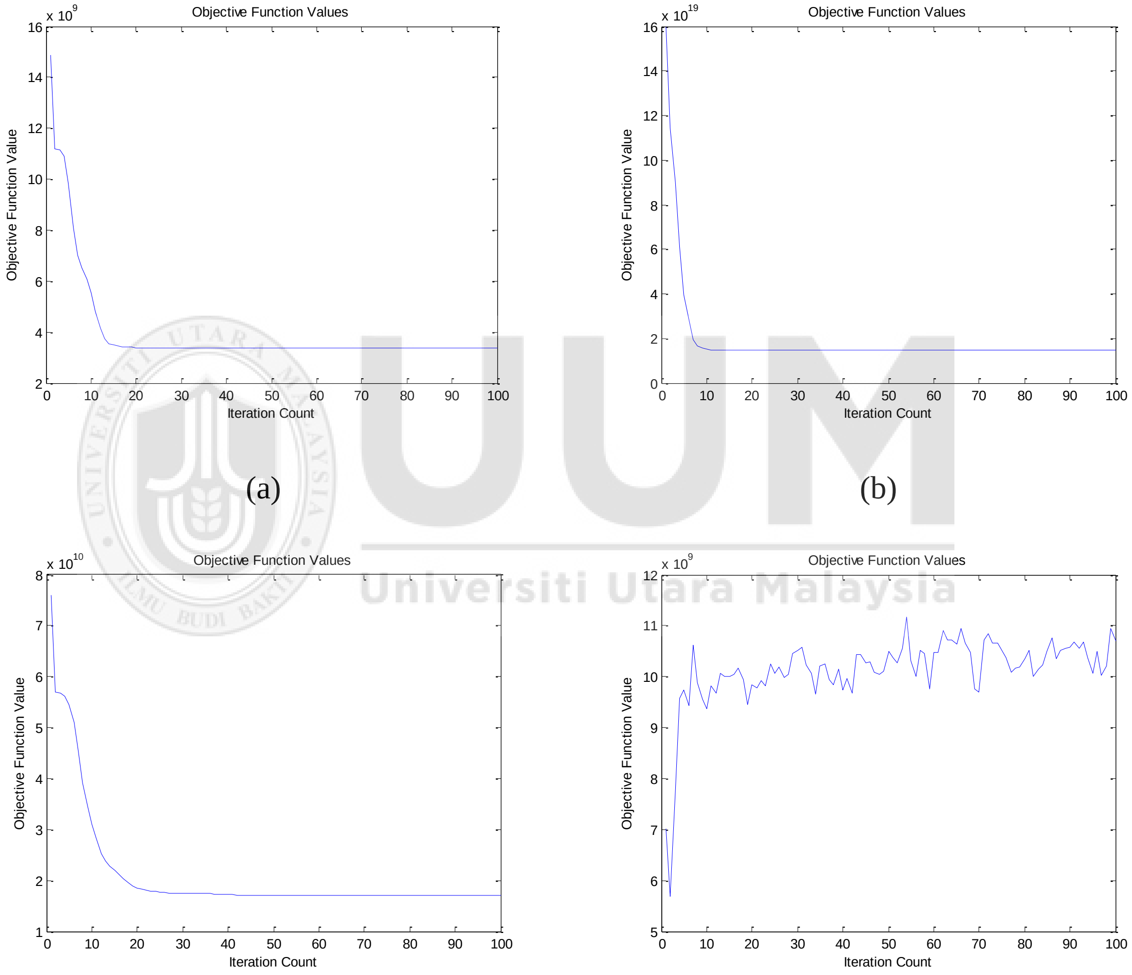
			Objective Function		
IC	**Euclidean**	**Manhattan**	**Minkowski**	**Chebyshef**	**Limit-Chebyshev**
1	7.14	2.062621	1.067716822	1.90452	1.40452
2	7.139	2.056409	1.06449368	1.930667	1.430667
3	7.138	2.051245	1.060970646	1.922361	1.362361
4	7.137	2.047075	1.057071247	1.86202	1.32202
5	7.129	2.043678	1.05299538	1.89609	0.89609
6	6.126	2.040547	1.049466338	1.830843	0.230843
7	6.125	2.037688	1.046122282	1.881059	0.230842
8	6.130	2.035324	1.042685003	1.917807	0.230841
9	6.129	2.032804	1.039468937	1.948442	0.230840
10	5.110	2.029883	1.03696179	1.88447	0.230833
15	4.056	2.026913	1.035212552	1.883791	0.230833
20	3.042	2.02406	1.033861189	1.884945	**0.230833**
25	3.300	2.020768	1.032773852	1.900423	
27	2.220	2.017347	1.031808493	1.935505	
29	**1.125**	2.013915	1.03092372	1.895413	
30	1.455	2.010393	1.03011209	1.937899	
33		2.006689	1.029426978	1.934352	
35		2.003059	1.028790337	**1.859295**	
37		1.998906	**1.02824743**	1.897755	
70		1.994468	1.027818098		
80		**1.994467**			

Universiti Utara Malaysia

Table 6.12

Summary of iteration count that leads to the minimum objective function for dataset4

	Euclidean	Manhattan	Minkowski	Chebyshev	Improved Chebyshev
IC	30	80	70	37	20
OF	1.455	1.994467	1.027818098	1.842837	0.230833



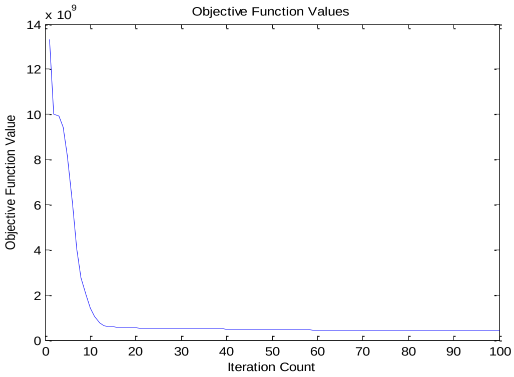
(a)

(b)

(c)

(d)

UUM Universiti Utara Malaysia



(e)

Figure 6.8. Progress of the objective function for Genbase dataset for FCM with (a): Euclidean distance (b): Manhattan Distance, (c): Minkowski distance (d): Chebysheve distance (e): Limit-Chebyshev.

Figure 6.9 shows the cluster density of this dataset. From this figure, it is obvious that the difference in cluster density is smaller than it is in the previous dataset, which affects the clustering accuracy. This is clear in the values of minimum objective function of dataset 3 and 4, while dataset 4 has smaller values of objective function than dataset 3.

Universiti Utara Malaysia

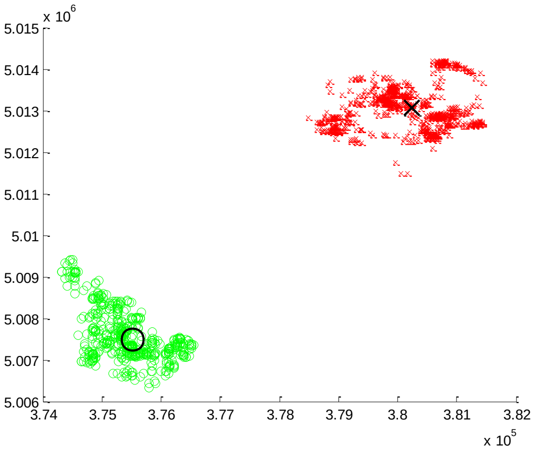


Figure 6.9. Cluster density for dataset 4

6.2.5 Dataset 5

In this experiment, the Yeast dataset is used. A comparison of results is obtained between improved FCM and conventional one.

6.2.5.1 $_C$FCM vs. conventional FCM

The results are tabulated in table 6.13, which gives the values of iteration counts and objective function generated from applying the FCM algorithm with the four-distance metrics and the proposed distance metric. From this table, it is seen that the FCM with Limit-Chebyshev has the MOF with less iteration count, which is 0.40 at IC equal 55. For the other four distances, the value of MOF and IC are: 2.06E+09 at 85, 2.91E+09 at 95, 6.2E+09 at 96 and 3.55E+09 at 90 for Chebyshev, Minkowski, Manhattan and Euclidean distances respectively. Table 6.14 summarize the IC, s that leads to MOF for the five-distance metrics in use. From this table, it can be seen that the higher OF and IC belongs to applying FCM algorithm with Manhattan distance (6.2E+09) at IC (96). Figure 6.10 shows the graphical presentation of the progress of the objective function of each of the five-distance metrics. The figure illustrates that the objective function minimizes with the progress of clustering by using the Limit-Chebyshev distance (refer to Figure 6.10 (e)). In figure 6.10 (a) the value of the objective function is close to 4 when used with Euclidean distance at iteration count 90; while in (b) close to 6.5 by using Manhattan distance at iteration count 96, in (c) close to 3 by using Minkowski at iteration count 95, while in (d) it is 2 by using original Chebyshev at iteration count 85 and (e) close to 0.5 by using the Limit-Chebyshev metric at iteration count 55.

Universiti Utara Malaysia

Table 6.13

Value of objective function and iteration counts for dataset3

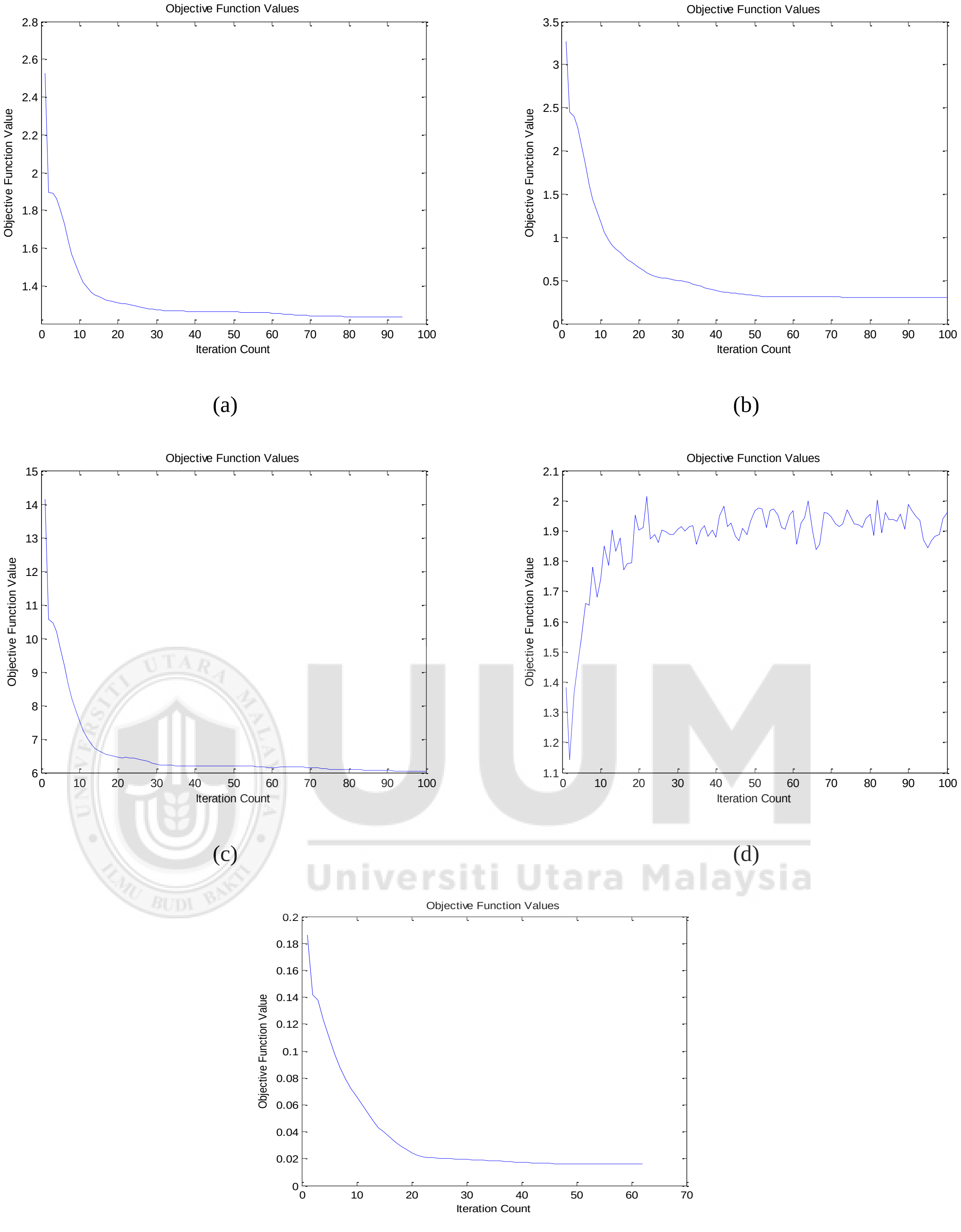
IC	Objective Function				
	Euclidean	**Manhattan**	**Minkowski**	**Chebyshef**	**Limit-Chebyshev**
1	3.78E+10	7.8E+10	2.13E+10	9.64E+10	1.94 E+9
15	3.94E+09	7.8E+10	1.54E+10	9.47E+10	0.82 E+9
20	3.79E+09	7.7E+10	1.42E+10	9.35E+10	0.77 E+9
25	3.59E+09	7.7E+10	1.17E+10	9.18E+10	0.51 E+9
30	3.56E+09	7.7E+10	9.33E+09	9.14E+09	0.51 E+9
35	3.55E+09	7.6E+10	7.75E+09	8.57E+09	0.50 E+9
40	3.55E+09	7.5E+10	6.73E+09	7.45E+09	0. 50 E+9
45	3.55E+09	7.4E+10	5.66E+09	6.54E+09	0.41 E+9
50	3.55E+09	7.3E+10	4.69E+09	5.9E+09	0.41 E+9
55	3.55E+09	6.8E+09	4.15E+09	5.18E+09	**0.40 E+9**
60	3.55E+09	6.2E+09	3.79E+09	4.55E+09	
65	3.55E+09	6.9E+09	3.51E+09	4.04E+09	
70	3.55E+09	6.5E+09	3.33E+09	3.61E+09	
75	3.55E+09	6.9E+09	3.2E+09	3.79E+09	
80	3.55E+09	6.5E+09	3.07E+09	3.14E+09	
85	3.55E+09	6.4E+09	2.99E+09	**2.06E+09**	
90	**3.55E+09**	6.3E+09	2.94E+09		
95		6.3E+09	**2.91E+09**		
96		**6.2E+09**			

UUM Universiti Utara Malaysia

Table 6.14

Value of iteration count that leads to the minimum objective function for dataset 5

	Euclidean	**Manhattan**	**Minkowski**	**Chebyshev**	**Limit-Chebyshev**
IC	90	96	95	85	55
MOF	3.55E+09	6.2E+09	2.91E+09	2.06E+09	0.40 E+9



(a)

(b)

(c)

(d)

(e)

Universiti Utara Malaysia

Figure 6.10. Progress of the objective function for Yeast dataset for FCM with (a): Euclidean distance (b): Manhattan Distance, (c): Minkowski distance (d): original Chebyshev distance (e): Limit-Chebyshev distance.

Figure 6.11 shows the cluster density of this dataset. From this figure, it is obvious that the deference in cluster density is high, which thereby results in poor performance of clustering, then affect the value of minimum objective function and iteration count. This is clear from table 6.14.

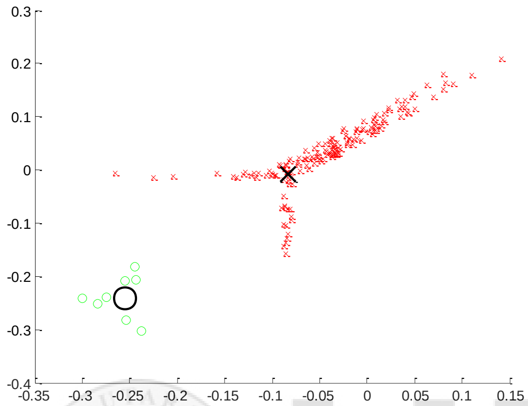


Figure 6.11. Cluster density for dataset 5

Universiti Utara Malaysia

6.3 Evaluation

The proposed metric distance of FCM is later evaluated to study the performance of the FCM algorithm in improving the accuracy of clustering. A comparison is made against the FCM algorithm with metric distances under consideration, which are Euclidean distance (Grabusts, et al.,2011), Manhattan distance ((Jun & Shi-Tong, et al.,2010), Minkowski distance ((Kouser1 & Sunita, et al.,2013), and Chebyshev distance ((Lai & Garibaldi, et al.,2013). Another comparison is made of $_T$FCM algorithm with metric distances under consideration. Table 6.17 illustrates the results of applying original FCM with the five distances. From table 6.17, it is seen that the Limit-Chebyshev distance has the highest purity value for all datasets, then the Chebyshev distance which achieves the better purity values after Limit-Chebyshev and the highest value is associated with STARKEY”93 (0.9125), while Manhattan register the smallest purity value (0.1001) with Genbase dataset. Figure 6.12 illustrates the graphical representation of the result.

From previous discussion, it is concluded that:

1- The proposed distance metric achieves high purity values among other distances under consideration, which later improves the performance of the clustering algorithm

2- Chebyshev distance has higher values with respect to the rest distances (Euclidean, Manhattan, Minkowski), which support the selection of Chebyshev distance in this research.

3- Limit-Chebyshev works good with big size dataset which have the chance of more overlapping than small size datasets.

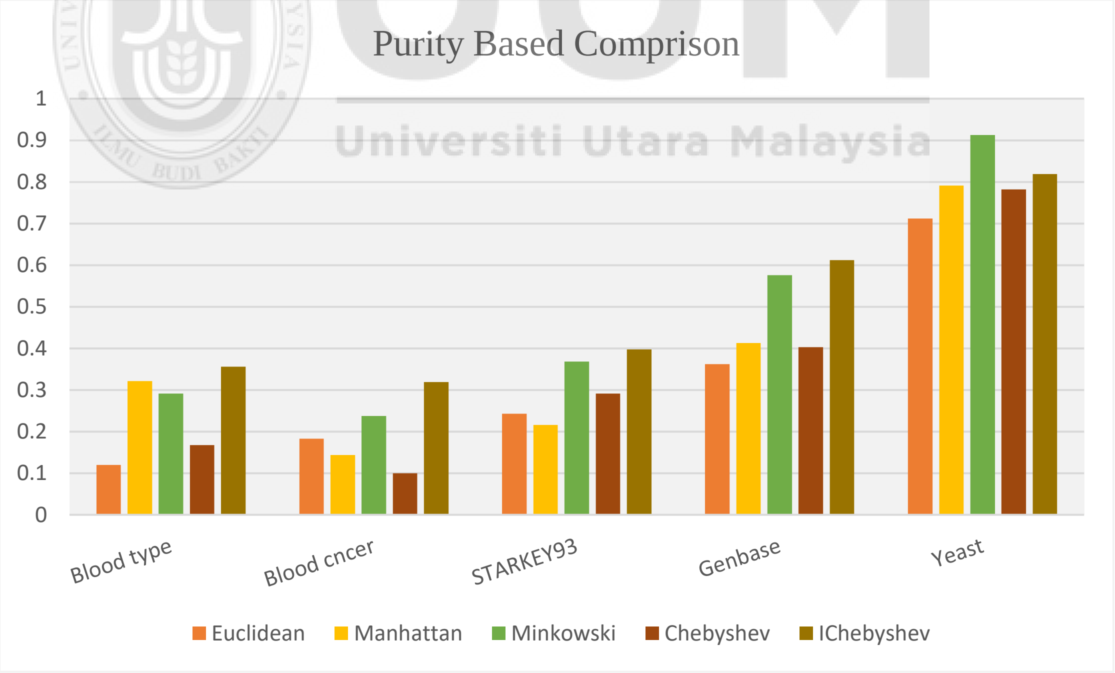
For entropy evaluation value, the smallest entropy value has been registered with Limit-Chebyshev at STARKEY"93 dataset (0.0108**).** Figure 6.12 and 6.13 represents the graphical presentation of the comparison results for purity and entropy. From this figure it is noted that Limit-Chebyshev distance achieve smaller Entropy value for all datasets, while Manhattan achieves highest value (1.0497) as it is learned before in (Grabusts, et al.,2011), that Manhattan distance is not suitable for FCM.

UUM Universiti Utara Malaysia

Table 6.17

Purity evaluation metric for FCM clustering with: Euclidean vs. Manhattan vs. Minkowski vs. Chebyshev vs. Limit-Chebyshev.

Evaluation metric	Dataset	Euclidean	Manhattan	Minkowski	Chebyshev	Limit-Chebyshev
Purity	Blood type	0.1201	0.1832	0.2432	0.3621	**0.7123**
	Blood cancer	0.3213	0.1436	0.2160	0.4130	**0.7912**
	STARKEY"93	0.2912	0.2376	0.3683	0.5761	**0.9125**
	Genbase	0.1673	**0.1001**	0.2912	0.4028	**0.7819**
	Yeast	0.3563	0.3192	0.3976	0.6120	**0.8186**
Entropy	Blood type	0.1238	0.2549	0.1214	0.2092	**0.1142**
	Blood cancer	0.1269	0.1821	0.1623	0.1178	**0.0139**
	STARKEY"93	1.0135	1.0497	0.0325	0.0114	**0.0108**
	Genbase	0.1261	0.1921	0.1451	0.0321	**0.3113**
	Yeast	1.0249	0.2101	1.1381	0.0618	**1.0114**



UUM
Universiti Utara Malaysia

Figure 6.12. Graphical representation of Purity values of FCM with distance metrics: Euclidean vs. Manhattan vs. Minkowski vs. Chebyshev vs. limit-Chebyshev for the datasets under consideration.

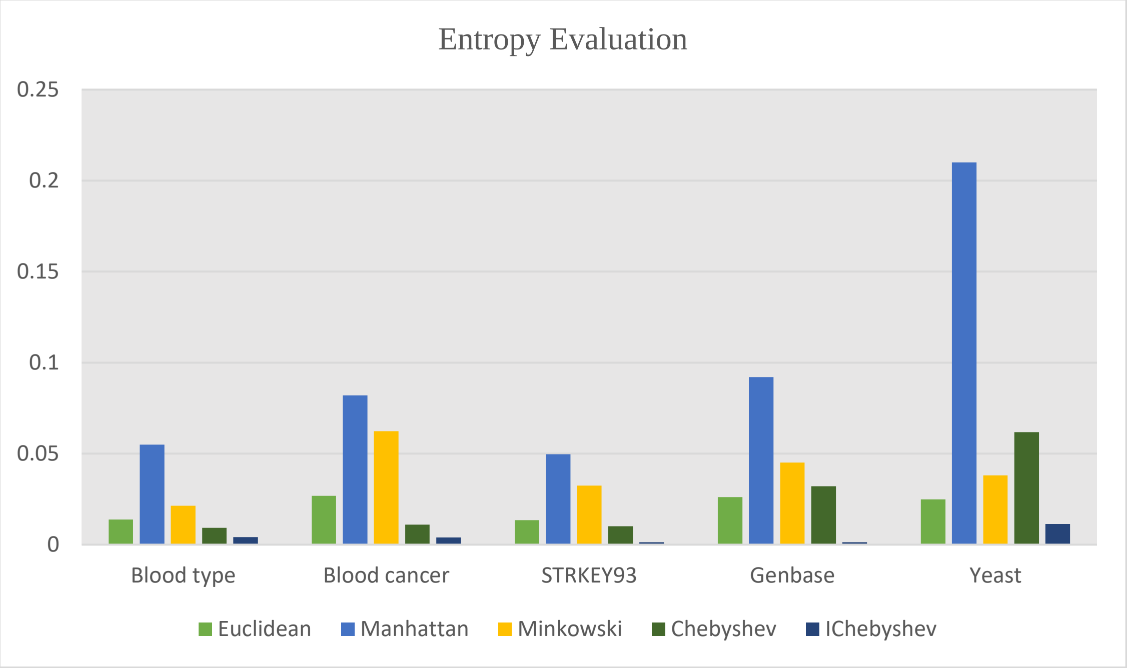


Figure 6.13. Graphical representation of Entropy values of FCM with distance metrics: Euclidean vs. Manhattan vs. Minkowski vs. Chebyshev vs. Limit-Chebyshev for the datasets under consideration.

Universiti Utara Malaysia

Table 6.18 illustrates the results of applying $_T$FCM with the five distances. From this table, it could be noted that applying $_T$FCM with Limit-Chebyshev distance metric will improve the performance of clustering process. It could be seen from the table that the highest purity value which is (0.9783) associated with blood type data and it is higher than the purity value registered by applying FCM with Limit-Chebyshev (0.9125), which indicates that the proposed $_T$FCM with proposed Chebyshev improves the clustering performance of overlapped data that which could be found with medical data more than regular data, as mentioned previously. Figure 6.14 illustrates the graphical representation of the results.

From the previous discussion, the following could be concluded

1- Proposed Limit-Chebyshev achieved very good performance when applied with the proposed $_T$FCM. This is clear from the high purity value for all datasets used in this research.

2- Highest purity values for proposed Limit-Chebyshev with $_T$FCM have been found with medical data than the other, as seen in the table. This indicates that the proposed FCM with Limit-Chebyshev can handle the overlapped problem.

3- Proposed distance with proposed $_T$FCM achieves the highest value of purity for all datasets, which means that it improves the clustering performance which then affect the accuracy of recognition of HTM as it will be seen later in Chapter Seven.

4- Proposed distance metric with $_T$FCM doing well with overlapping in data regardless of the size of data

For Entropy, it is noted from the table that Limit-Chebyshev with $_T$FCM register the smallest value for all datasets. For the sequence of entropy with respect to the distance metrics used in this research, blood cancer has smallest value under Chebyshev distance, which is (0.0015), while blood type has smallest value under the rest three distances: Euclidean (0.0038), Manhattan (0.0344) and Minkowski (0.0154). Figure 6.15 illustrates the graphical representation of the result.

UUM Universiti Utara Malaysia

Table 6.18

Purity evaluation metric for $_T$FCM clustering with: Euclidean vs. Manhattan vs. Minkowski vs. Chebyshev vs. Limit-Chebyshev

Evaluation metric	Dataset	Euclidean	Manhattan	Minkowski	Chebyshev	**Limit-Chebyshev**
Purity	Blood type	0.2061	0.2031	0.2997	0.4201	**0.9783**
	Blood cancer	0.4083	0.1998	0.2890	0.4820	**0.8913**
	STARKEY”9	0.3212	0.2906	0.4081	0.7762	**0.9045**
	Genbase	0.1993	0.1321	0.4576	0.6043	**0.8094**
	Yeast	0.4003	0.3901	0.4312	0.7120	**0.8866**
Entropy	Blood type	0.0038	0.0344	0.0154	0.0072	**0.0030**
	Blood cancer	0.0200	0.0621	0.0433	0.0015	**0.0128**
	STARKEY”9	0.0103	0.1377	0.0256	0.0112	**0.2101**
	Genbase	0.0231	0.0841	0.0341	0.0301	**0.0010**
	Yeast	0.2333	0.2117	0.3420	0.1518	**0.1123**

Universiti Utara Malaysia

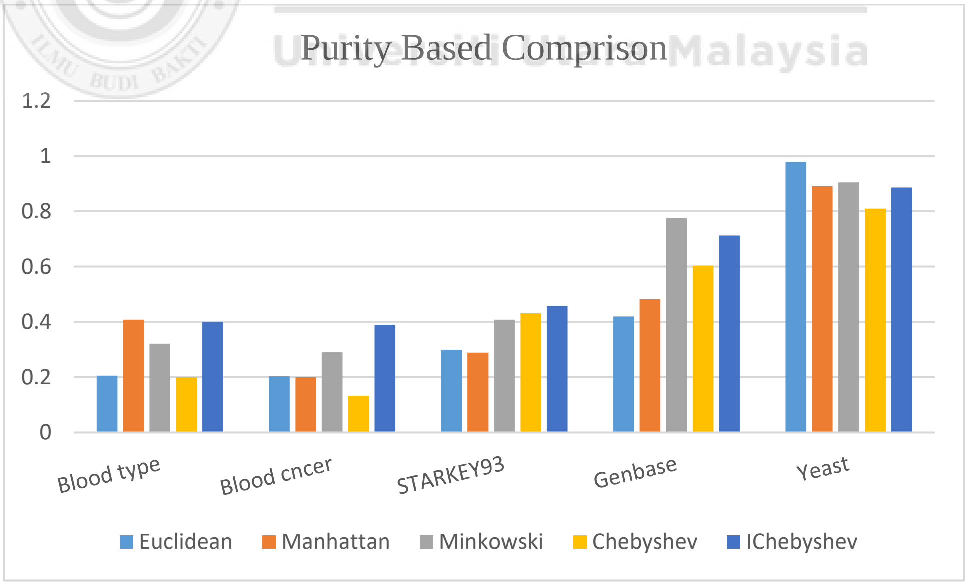


Figure 6.14. Graphical representation of Purity values of $_T$FCM with distance metrics: Euclidean vs. Manhattan vs. Minkowski vs. Chebyshev vs. Limit-Chebyshev for the datasets under consideration.

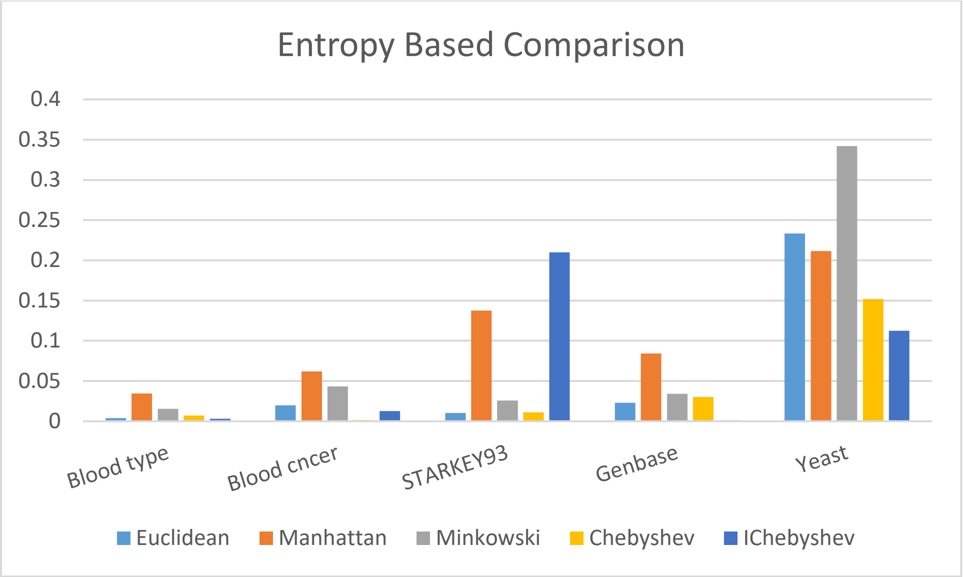


Figure 6.15. Graphical representation of Entropy values of $_T$FCM with distance metrics: Euclidean vs. Manhattan vs. Minkowski vs. Chebyshev vs. Limit-Chebyshev for the datasets under consideration.

UUM Universiti Utara Malaysia

Table 6.19 demonstrates the result of comparison between HTM- $_T$S-layer, HTM -$_T$FCM, HTM-Limit-Chebyshev, HTM -$_T$FCM-Limit-Chebyshev, HTM- $_T$S-layer -$_T$FCM, HTM - $_T$S-layer-$_T$FCM-Limit-Chebyshev. From this table, it is seen that HTM-$_T$S-layer-$_T$FCM-Limit-Chebyshev reach the highest accuracy, which assign in bold.

Table 6.19

Comparison of HTM with improved Chebyshev distance vs. original HTM vs. $_T$S-layer vs. $_T$HTM vs. $_T$HTM with improved Chebyshev

Algorithm		Accuracy in %			
	Blood type	Blood Cancer	STARKEY"93	Genbase	Yeast
Original HTM	84	84.4	87	83.3	88
$_T$S-layer	96.2	96.9	98	97	99
HTM-$_T$FCM	95	96	96.7	98.8	98
HTM-$_T$FCM-$_T$S-layer	97	99	98.2	98.9	97.4
HTM-Limit-Chebyshev	97.5	97.1	97.8	99	99
HTM-$_T$FCM-Limit-Chybyshev	98.6	98.3	97.7	98	98
HTM-$_T$FCM-Limit Chybyshev-$_T$S-layer	**99.3**	**98.8**	**99.5**	**99.3**	**99**

UUM Universiti Utara Malaysia

6.4 Result analysis for cFCM vs. original FCM

In comparisons between improved FCM and original FCM with five experiments on five datasets, it is obvious that the improved FCM is better than the original one. The following conclusions are found from the previous discussion:

1. The improved Chebyshev has the minimum objective function with less iteration count.
2. The size of the dataset affects the number of iteration count. Big dataset results more iteration counts, which slow the clustering operation, but give more accurate clustering.
3. Big datasets may have much overlap in data points; hence, the clustering will be more complex and require an improved distance metric that can process this overlapped, which is achieved by the proposed distance metric.

4. The density of clusters affects the value of objective function, then performance of clustering. The small deference in cluster density, the minimum objective function, which leads to more accurate clustering.

5. Big dataset has high difference in cluster density, which needs improved distance metric for accurate clustering. Table 6.20 shows the difference between objective function of small and big datasets.
   The proposed Limit-Chebyshev registers high purity and small entropy values with medical data, which are overlapping.

   With the previous foundations, the FCM with improved distance metric has improve the performance of the clustering when the dataset is big and when the deference in density of clusters are small; meanwhile, the proposed algorithm has good results with high cluster density deference, better than the original algorithm.

6. The proposed Limit-Chebyshev with proposed temporal HTM and improved Neocognitron achieves the best purity and entropy values, which indicates high clustering performance which leads to high recognition accuracy

UUM Universiti Utara Malaysia

Table 6.20

Difference between objective function of Genbase and Yeast datasets

Objective Function	**Euclidean**	**Manhattan**	**Minkowski**	**Chebyshev**	**Limit-Chebyshev**
Genbase	1.455	1.994467	1.027818098	1.842837	0.230833
Yeast	3.55E+09	6.2E+09	2.91E+09	2.06E+09	0.40E+9

6.5 Summary

This chapter presents image and dataset clustering using an improved distance metric. Experiments were conducted on 500 blood type images, 33 blood cancer images and three multi- label datasets. Comparisons were made against the HTM algorithm presented in Chapter Four, original HTM, combination of HTM+ $_T$FCM and $_T$S-layer algorithm and combination of HTM-Limit Chebyshev-$_T$FCM + and HTM-Limit Chebyshev + HTM-$_T$FCM + HTM-$_T$S-layer, HTM- Limit Chebyshev-$_T$FCM and HTM-Limit Chebyshev-$_T$FCM-$_T$S-layer. Purity and entropy evaluation metrics were applied to verify the performance of $_T$HTM.

Universiti Utara Malaysia

It is learned that the proposed Limit-Chebyshev generates better results than the original one in terms of minimum objective function and least iteration counts. Comparison of accuracy indicates that proposed Limit-Chebyshev with proposed HTM-Limit Chebyshev-$_T$FCM-$_T$S-layer provide better recognition of patterns. By this result, objective three has been achieved.

CHAPTER SEVEN
IMPROVED HIERARCHICAL TEMPORAL MEMORY

7.1 Introduction

In this chapter, a combination of the proposed algorithms demonstrated in Chapter Four, Five and Six is presented. The improvement includes the integration of Improved S-layer Neocognitron neural network ($_T$S-layer) as feature extraction step (as in Chapter Four) and $_T$FCM that includes limit Chebyshev distance as clustering step in HTM (as in Chapter Five and Six). The improved HTM is used to perform recognition in multi-label datasets. Five multi-label datasets are used to test the proposed method. HTM algorithm includes three steps: feature extracting step (memorizing) using improved Neocognitron neural network ($_T$S-layer) as in Chapter Four, and Markov model and temporal FCM clustering with improved Chebyshev distance. Figure 7.1 illustrates this algorithm.

Universiti Utara Malaysia

imHTM Algorithm

Feature extraction

```
Input stimuli of size 4x4 data part is presented to S-layer plain, each plane receives the same input connection
time matrix of same dimension of input stimuli is created. Each element of the time matrix corresponds to one input vector
Input is compared with the S-cells content according to threshold ($\theta$ =0.02)
    If identical then
      Begin
        For each identical (winner) cell
            apply improved S-layer function, and output of winner cells (Us). By following Equation
            $u_{si}(k,n,t) = \frac{\theta}{1-\theta}.\varphi\left[\frac{1+\ \sum_{k=i<AS_l}^{Kc_{l-1}} \sum as_l\,(v,k,K,t)}{1+\theta.bs_i(k,t).v_i(n,K,t)} - 1\right]$
            $\theta$ is the threshold of the S-cell and determines the selectivity in extracting features
    End
    Else
    Generate new plain and store input pattern in new cell in this plain
end if
```

Markov Model

```
Each added pattern (vertex of Markov graph) given a number
```

Universiti Utara Malaysia

```
For each vertex do
    Link between two vertex L is established (number of transitions between vertices)
    Total number of transition N is established
    If new pattern added L= L+1, N=N=+1
    Normalize Markov graph for each vertex by dividing number of transitions out from it over the total number of transitions of that vertex:
        $\frac{L}{N}$
End for
Clustering
Initialize $\mu = [\mu_{ij}]^0$ $matrix$ of membership
Choose parameter $l$ >0 to stop the iteration, set the iteration counting parameter $l$ equal to 1
Set time matrix $t[i] = 0$
At k-step calculate the centers vectors $v^k$=[$v_j$] by
    $\boldsymbol{v_j} = (\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} (\mu_{\boldsymbol{ij}})^{\boldsymbol{m}} \boldsymbol{u_w(t)}_i)/(\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} (\mu_{\boldsymbol{ij}})^{\boldsymbol{m}}),\ \forall_{\boldsymbol{j}} = \boldsymbol{1, 2, 3} \dots .\boldsymbol{c}$
Update the membership matrix $[\mu_{ij}]^k$ , $[\mu_{ij}]^{k+1}$ by
    $\mu_{\boldsymbol{ij}} = \boldsymbol{1}/\sum_{\boldsymbol{k=1}}^{\boldsymbol{c}} (\boldsymbol{d_i}/\boldsymbol{d_{ik}})^{(\frac{\boldsymbol{2}}{\boldsymbol{m-1}})}$
```

Universiti Utara Malaysia

```
22:  If  $[\mu_{ij}]^{k+1}$- $[\mu_{ij}]^{k} > \beta$      where $\beta$ is the termination condition such that
            $\boldsymbol{\beta} = \Delta_{\boldsymbol{j}}^{\boldsymbol{N}} \left|[\mu_{ij}]^{k+1} - [\mu_{ij}]^{k}\right|$
        $\Delta$ supplied a vector of values, returns the largest value in that vector
23:            stop at some iteration $l$
24:        Else
25:        Go to step 16
26:  End if
27: Calculate limit distance
            d$\boldsymbol{xy}$=$\lim_{\boldsymbol{p}\to\infty} (\boldsymbol{max_k}(|\boldsymbol{x_{ik}} - \boldsymbol{x_{jk}}|)^{\boldsymbol{1/p}})$   $\infty < \boldsymbol{p} > \boldsymbol{2}$
28:  Calculate the temporal function and map it to the new membership function
     by
            $\boldsymbol{p_{ij}} = \sum_{\boldsymbol{t=1}}^{\boldsymbol{m}} \boldsymbol{\mu_{it}}$
            $\boldsymbol{\mu}_{\boldsymbol{new_{ij}}} = \frac{\boldsymbol{\mu_{ij}p_{ij}}}{\sum_{\boldsymbol{t=1}}^{\boldsymbol{c}} \boldsymbol{\mu_{tj}p_{tj}}}$
29:  If  $dxy == 0.02$
23 Go to step 33
31:   Else
32:        Goto step 28
33:  End if
```

Universiti Utara Malaysia

Figure 7.1. Improved HTM Algorithm for overlapped data

7.2 Comparison Results

The proposed HTM is compared with other structural pattern recognition methods and statistical methods. This comparison is made to study its performance and accuracy in the recognition of overlapping and shifting in position of data points of multi- label datasets. The proposed HTM has been compared with Hidden Markov Model (HMM) (Jose, Nair, .Biju, Mathew, & Prashanth, et al.,2016) and original HTM (Skrynnik, Petrov, & Panov, et al.,2016) as a structural method, Sum of Absolute Difference (SAD) method of Template matching (Mahalakshmi et al., et al.,2012), Principal Component Analysis (PCA), Support Vector Machine (SVM), and Neocognitron neural network (Gorokhovatskyi, et al.,2016) statistical approaches (Huan et al., et al.,2012). Recognition rate is used for comparison (number of recognized patterns over the whole number of patterns). Table 7.1 illustrates the number of recognized patterns from specific number of input patterns for imHTM and original HTM. It is seen from the table that imHTM achieves the best performance of recognition, where the number of recognized pattern for imHTM is greater than the original HTM for all amount of input patterns (assign in bold).

Table 7.2 gives the result of comparison between imHTM and Neocognitron. It is seen from this table that the proposed method again achieves better performance than the Neocognitron neural network. From tables 7.1and 7.2, the original HTM recognize less patterns than the others. This is due to the memorizing and clustering algorithms applied with original HTM which is sensitive to shift in position and noisy patterns as mentioned in Chapter Four and Five. Neocognitron achieves better results as it recovers the shift in position problem, but still has the noisy problem, imHTM has the highest results for its ability of overcome the previous problems.

Table 7.1

Number of correct recognized patterns: ¡mHTM and original HTM

Dataset	No. of input pattern	Number of recognized patterns	
		Original HTM	imHTM
Blood type	100	20	**96**
	200	100	**147**
	300	110	**270**
	400	135	**390**
	500	175	**480**
	600	185	**495**
	700	305	**698**
Blood cancer	10	3	**9**
	40	15	**34**
	70	33	**69**
	100	48	**90**
	130	75	**120**
	160	79	**157**
	190	83	**188**
Genbase	100	57	**94**
	200	120	**196**
	300	200	**293**
	400	230	**389**
	500	330	**489**
	600	350	**593**
	700	400	**694**
STARKEY”93	100	56	**97**
	200	117	**193**
	300	200	**296**
	400	228	**391**
	500	328	**499**
	600	350	**520**
	700	389	**689**
Yeast	100	54	**98**
	200	116	**180**
	300	200	**290**
	400	223	**380**
	500	320	**482**
	600	350	**540**
	700	382	**690**

UUM Universiti Utara Malaysia

Table 7.2

Number of correct recognized patterns: imHTM vs. Neocognitron

		Number of recognized patterns	
Dataset	No. of input pattern	Neocognitron	imHTM
Blood type	100	50	**96**
	200	140	**147**
	300	160	**270**
	400	150	**390**
	500	275	**480**
	600	270	**495**
	700	375	**698**
Blood cancer	10	6	**9**
	40	20	**34**
	70	40	**69**
	100	56	**90**
	130	70	**120**
	160	89	**157**
	190	90	**188**
Genbase	100	75	**94**
	200	124	**196**
	300	190	**293**
	400	230	**389**
	500	290	**489**
	600	350	**593**
	700	436	**694**
STARKEY”93	100	70	**97**
	200	117	**193**
	300	186	**296**
	400	220	**391**
	500	279	**499**
	600	350	**520**
	700	436	**689**
Yeast	100	72	**98**
	200	112	**180**
	300	190	**290**
	400	240	**380**
	500	290	**482**
	600	368	**540**
	700	469	**690**

UUM Universiti Utara Malaysia

Tables 7.3 and 7.4 summarize the accuracy values of imHTM, HMM, HTM and Neocognitron. It is seen from these tables that imHTM is the most accurate method for recognition than the other method, which is an important factor when used with medical application which needs high level of accuracy. Figures 7.2 and 7.3 represent the comparison graphically. In these figures, the high performance of the proposed HTM is clear.

Table 7.3

Accuracy value for imHTM vs. original HTM vs. HMM

Dataset	Accuracy in %		
	HMM	HTM	imHTM
Blood Type	47	87	**98.8**
Blood Cancer	50	85	**99**
STARKEY"93	68	89	**97.9**
Genbase	62	80	**99.3**
Yeast	54	87	**98.4**

Universiti Utara Malaysia

Table 7.4

Accuracy value for imHTM vs. Neocognitron

Dataset	Accuracy in %	
	Neocognitron	imHTM
Blood type	83	**98.8**
Blood cancer	85	**99**
STARKEY"93	80.8	**97.9**
Genbase	89.3	**99.3**
Yeast	90.8	**98.4**

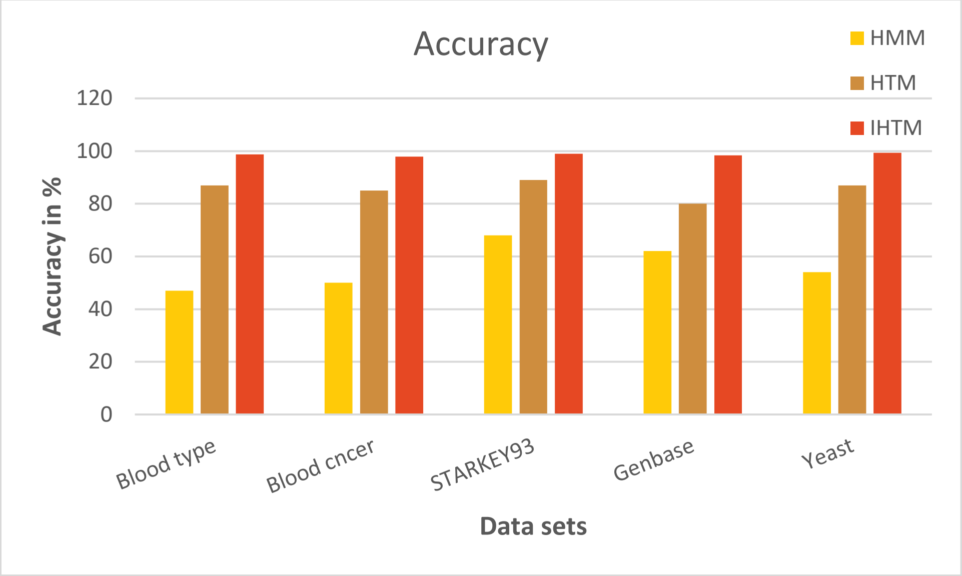
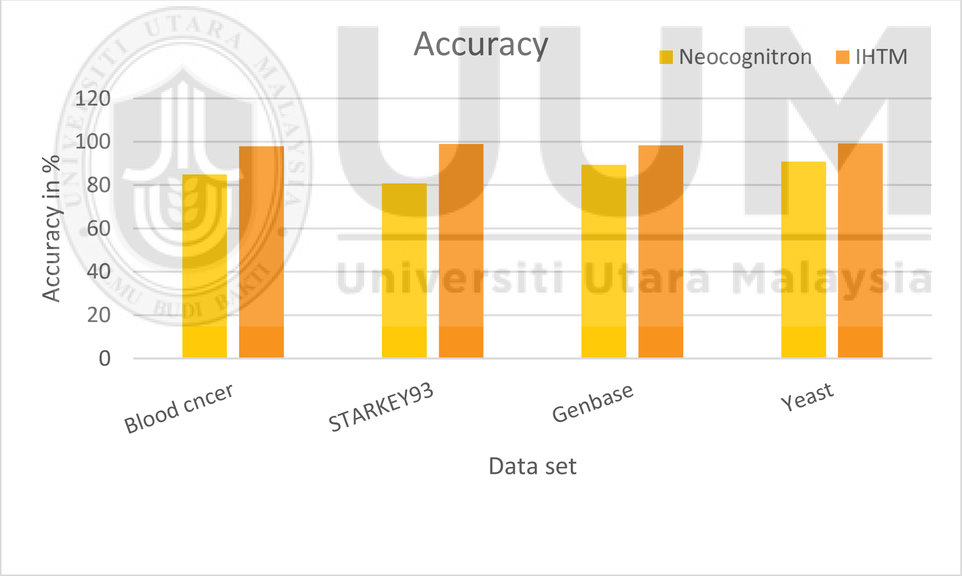


Figure 7.2. Accuracy comparison of HMM vs. HTM vs. imHTM



Universiti Utara Malaysia

Figure 7.3. Accuracy comparison of Neocognitron vs. imHTM

Table 7.5 illustrates the comparison between improved HTM with example methods of the other approaches of pattern recognition mentioned in Chapter two section 2.2. This comparison is made to prove the high performance of improved HTM algorithm in recognizing big number of patterns from input patterns with respect to the other approaches of pattern recognition. From table 7.5, it could be seen that imHTM recognize the greatest number of patterns from the whole pattern presented to the HTM for all datasets. It is also seen that SVM is the second algorithm in recognizing patterns.

Table 7.5

Number of correct recognized patterns: imHTM vs. PCA vs. SAD vs. SVM

No. of input pattern	**Dataset**	**Number of recognized patterns**			
		imHTM	**PCA**	**SAD**	**SVM**
100		**96**	60	40	70
200	Blood Type	**147**	100	80	150
300		**270**	150	100	220
400		**390**	190	150	300
500		**480**	260	200	415
600		**495**	350	278	310
700		**698**	390	468	498
10	Blood Cancer		6	7	8
40		**9**	10	16	25
70		**34**	26	30	40
100		**69**	50	65	69
130		**90**	70	90	100
160		**120**	75	98	105
190		**157**	90	115	130
		188			
100	STARKEY"93		60	70	80
200		**94**	90	100	110
300		**196**	130	140	160
400		**293**	180	200	210
500		**389**	200	230	270
600		**489**	290	310	340
700		**593**	300	350	450
		694			
100	Genbase		30	35	40
200		**97**	80	90	100
300		**193**	100	110	130
400		**296**	140	145	156
500		**391**	200	220	240
600		**499**	230	250	289
700		**520**	230	250	270
		689			
100	Yeast		40	45	60
200		**98**	100	120	150
300		**180**	160	170	178
400		**290**	149	158	190
500		**380**	270	290	384
600		**482**	300	350	400

Universiti Utara Malaysia

700	**540**	400	450	498

Table 7.6 illustrates the accuracy values for imHTM against other pattern recognition approaches under consideration. It is seen from this table that imHTM has the highest recognition accuracy among the other methods for datasets used in this research. It is also seen that there is high difference in accuracy for medical data between imHTM and the other methods and imHTM has the highest accuracy value with this data, demonstrating that the proposed method achieve its goals of shift in position and overlapping in data points.

Table 7.6

Accuracy value for imHTM vs. PCA vs. SAD vs. SVM

	Accuracy in %			
Dataset	**PCA**	**SAD**	**SVM**	**imHTM**
Blood Type	45	64	74	**98.8**
Blood Cancer	40	65	87	**99**
STARKEY”93	69	78	88	**97.9**
Genbase	56	64	70	**99.3**
Yeast	74	80	82	**98.4**

Universiti Utara Malaysia

7.3 Summary

This chapter presents image and dataset recognition by using improved Hierarchical Temporal Memory (imHTM). Experiments were conducted on 500 blood type images, 33 blood cancer images and three multi- label datasets. Comparison was made against HTM algorithm presented in Chapter Four, original HTM, combination of ${}_{T}$HTM and ${}_{T}$S-layer algorithm and combination of ${}_{T}$HTM and ${}_{T}$S-layer algorithm, ${}_{TC}$HTM and ${}_{TCN}$HTM which were presented in Chapter Five and Six. Accuracy evaluation was applied to verify the performance of imHTM, also number of recognized patterns were used to compare imHTM with the mentioned algorithms. Comparisons are made against other pattern recognition approaches (PCA, SAD, and SVM).

It is learned that imHTM generates better results in accuracy and number of recognized patterns. Such results indicate that imHTM is a competitor to other approaches in pattern recognition.

UNIVERSITI UTARA MALAYSIA
ILMU BUDI BAKTI
UUM
Universiti Utara Malaysia

CHAPTER EIGHT
CONCLUSION AND FUTURE WORK

The main goal of this thesis is to propose an improved Hierarchical temporal memory (HTM) for multi-label data recognition. HTM consists of three stages, feature extracting, temporal learning, and clustering phase. To achieve this goal, the first and third phases are improved. First, an improved feature extracting phase has been proposed to extract largest number of features from the input pattern and overcome the shift in position. Later the third phase, temporal information and a new distance measure is employed – an improved fuzzy c-means clustering algorithm that reduces the overlapping effect of data points and improve cluster quality. The proposed algorithms are then combined as temporal-spatial recognizes for multi- label data.

Universiti Utara Malaysia

8.1 Research Contribution

In this thesis, there are four main contributions. The first contribution is the feature extraction step of Hierarchical Temporal Memory. This step improved the S-layer Neocognitron neural network by adding time factor as matching factor for the next stage. Its block diagram is illustrated in Figure 3.6. Experiments in Chapter Four confirm that the proposed feature extraction step can extract a larger number of features than the original HTM and Neocognitron, which support the resolving of shift in position problem.

The second contribution includes the improvement of the clustering algorithm used in the third step of HTM. The improvement is made in two steps. The first, is the improvement of distance measure included in FCM clustering algorithm which is used in clustering step in HTM (illustrated in algorithm 6.1). The proposed distance measure (second contribution) enhances the performance of the clustering algorithm in terms of minimum objective function with less iteration count. Experiments in Chapter six indicates that the quality of clustering enhanced compared with the clustering algorithm with original distance measure.

The second improvement in the HTM clustering is the inclusion of fuzzification function (third contribution) to include temporal information (as illustrated in Figure 3.9). Experiments in Chapter Five indicate that the proposed function supports the clustering process and improves the performance in terms of high objective function with less iteration count, which leads to high membership function for the data points that overlap.

The fourth contribution is the integration of the $_{T}$S-layer and $_{T}$FCM with limit Chebyshev in the HTM. The proposed imHTM has been tested against several techniques of pattern recognition (template, structural and statistical approaches). Based on the results and discussion presented in Chapters 4, 5, 6 and 7, the proposed HTM for pattern recognition in multi-label data is proven to be useful. Results indicates that it is it a competitive method in recognizing patterns.

Universiti Utara Malaysia

8.2 Future Work

As a continuity to the study, an HTM algorithm with various numbers of cluster is suggested to be used in the clustering initial step. This can be represented as an optimization problem that determines the suitable number of cluster prior to the recognition process. Furthermore, analysis on the number of overlap data points should also be carried out in order to better understand the classification solutions for multi- label dataset.

Future work may also be carried in SDM, which has a high dimensional feature set. This can be useful in gaining ecological and evolutionary insights and to predict distributions across landscapes, which is widely used across terrestrial, freshwater, marine realms, plants, and animal distributions. The SDM has become increasingly important in the context of awareness of environmental change and its ecological consequences.

Analysis on number of overlap will be added to the research output to reach the optimum results and approve the effectivity of proposed HTM.

8.3 Limitation of the Study

This research has yet to be tested on other types of multi- label data, besides than the ones mentioned in Chapter 3. As there exist various multi- label data from different application domains, further investigate would generalized the propose algorithm to a more abstract level. In addition, the undertaken study has not considered algorithm complexity such as (processing time, number of cycles or looping and etc.) in measuring its performance.

Universiti Utara Malaysia

REFERENCES

Abdel-Azim, G. (2016). New Hierarchical Clustering Algorithm for Protein Sequences Based on Hellinger Distance. *Applied Mathematics & Information Sciences*, *10*(4), 1541-1549.

Abdulfattah, G. M., & Ahmad, M. N. (2013). Face Localization Based Template Matching Approach Using New Similarity easurements. *Journal of Theoretical and Applied Information Technology*, *57*(3).

Adhikari, S. K., Sing, J. K., Basu, D. K., & Nasipuri, M. (2015, 4-7 Jan. 2015). A spatial fuzzy C-means algorithm with application to MRI image segmentation. Paper presented at the 2015 *Eighth International Conference on Advances in Pattern Recognition (ICAPR).*

Adjouadi, M., Zong, N., & Ayala, M. (2005). Multidimensional Pattern Recognition and Classification of White Blood Cells Using Support Vector Machines. *Part. Part. Syst. Charact.*, *22*, 107-118.

Ahmad, S., & Hawkins, J. (2015). Properties of Sparse Distributed Representations and their Application to Hierarchical Temporal Memory. *arXiv:1503.07469*.

Ahmed, M., & Aradhya, V. N. M. (2016). A Study of Sub-Pattern Approach in 2D Shape Recognition Using the PCA and Ridgelet PCA. *Int. J. Rough Sets Data Anal.*, *3*(2), 10-31. Retrieved from doi: 10.4018/ijrsda.2016040102

Ali, J., Ahma, A., George, L. E., Der, C. S., & Aziz, S. (2013). Red Blood Cell Recognition using Geometrical Features. *International Journal of Computer Science (IJCSI)Issues*, *10*(1), 5.

Amza, C. G., & Cicic, D. T. (2015). Industrial Image Processing Using Fuzzy-logic. *Procedia Engineering*, *100*, 492-498. Retrieved from http://dx.doi.org/10.1016/j.proeng.2015.01.404

Arel, I., Rose, D. C., & Karnowski, T. P. (2010). Deep Machine Learning - A New Frontier in Artificial Intelligence Research [Research Frontier]. *Computational Intelligence Magazine, IEEE*, *5*(4), 13-18. Retrieved from doi: 10.1109/MCI.2010.938364

Arimond, A. (2010). A Distributed System for Pattern Recognition and Machine Learning. (Diploma). Retrieved from TU Kaiserslautern, TU Kaiserslautern.

Asgar, F., & Salehi, A. (2015). The biologically inspired Hierarchical Temporal Memory Model for Farsi Handwritten Digit and Letter Recognition. *International Journal of Computer Applications*, *129*(16).

Aytug Onan, Bulut, H., & Korukoglu, S. (2016). An improved ant algorithm with LDA-based representation for text document clustering. *Journal of Information Science*.

Azad, R., Azad, B., & Brojeeni, H. R. S. (2014). Real-Time and Efficient Method for Accuracy Enhancement of Edge Based License Plate Recognition System Paper presented at the *International Conference on computer, Information Technology and Digital Media*

Abdel-Azim, G. (2016). New Hierarchical Clustering Algorithm for Protein Sequences Based on Hellinger Distance. *Applied Mathematics & Information Sciences*, *10*(4), 1541-1549.

Abdulfattah, G. M., & Ahmad, M. N. (2013). Face Localization Based Template Matching Approach Using New Similarity easurements. *Journal of Theoretical and Applied Information Technology*, *57*(3).

Adhikari, Sing, Basu, & Nasipuri. (2015, 4-7 Jan. 2015). A spatial fuzzy C-means algorithm with application to MRI image segmentation. Paper presented at the

Universiti Utara Malaysia

2015 *Eighth International Conference on Advances in Pattern Recognition (ICAPR).*

Adjouadi, M., Zong, N., & Ayala, M. (2005). Multidimensional Pattern Recognition and Classification of White Blood Cells Using Support Vector Machines. *Part. Part. Syst. Charact., 22*, 107-118.

Ahmad, S., & Hawkins, J. (2015). Properties of Sparse Distributed Representations and their Application to Hierarchical Temporal Memory. *arXiv:1503.07469*.

Ahmed, M., & Aradhya, M. (2016). A Study of Sub-Pattern Approach in 2D Shape Recognition Using the PCA and Ridgelet PCA. *Int. J. Rough Sets Data Anal., 3*(2), 10-31. Retrieved from doi:10.4018/ijrsda.2016040102

Aksoy, S. (2012). Structural and Syntactic Pattern Recognition. Retrieved from doi:10.3422/ jnw.2012.06.014.

Alqaisy, M. A., Yassen, M. T., Khalifa, F. S., & Salim, A. J. (2016, 9-10 May 2016). A dual-band bandpass filter using single unit cell of complementary split ring resonator with third harmonic reduction. Paper presented at the 2016 *Al-Sadeq International Conference on Multidisciplinary in IT and Communication Science and Applications (AIC-MITCSA).*

Amza, C. G., & Cicic, D. T. (2015). Industrial Image Processing Using Fuzzy-logic. *Procedia Engineering, 100*, 492-498. Retrieved from doi:http://dx.doi.org/10.1016/j.proeng.2015.01.404

Arel, I., Rose, D. C., & Karnowski, T. P. (2010). Deep Machine Learning - A New Frontier in Artificial Intelligence Research *Computational Intelligence Magazine, IEEE, 5*(4), 13-18. Retrieved from doi:10.1109/MCI.2010.938364

Arimond, A. (2010). A Distributed System for Pattern Recognition and Machine Learning. (Diploma), TU Kaiserslautern, TU Kaiserslautern.

Asgar, F., & Salehi, A. (2015). The biologically inspired Hierarchical Temporal Memory Model for Farsi Handwritten Digit and Letter Recognition. *International Journal of Computer Applications, 129*(16).

Assis, L. d. S., Soares, A. d. S., Coelho, C. J., & Van Baalen, J. (2016). An Evolutionary Algorithm for Autonomous Robot Navigation. *Procedia Computer Science, 80*, 2261-2265. Retrieved from doi:http://dx.doi.org/10.1016/j.procs.2016.05.404

Aytug Onan, Bulut, H., & Korukoglu, S. (2016). An improved ant algorithm with LDA-based representation for text document clustering. *Journal of Information Science*. Retrieved from doi: 10.3211/22133/j.ins.7002441

Azad, R., Azad, B., & Brojeeni, H. R. S. (2014). Real-Time and Efficient Method for Accuracy Enhancement of Edge Based License Plate Recognition System Paper presented at the *International Conference on computer, Information Technology and Digital Media*

Badyalina, B., Shabri, A., & Samsudin, R. (2014). Streamflow Estimation at Ungauged Site Using Wavelet Group Method of Data Handling in *Int. Journal of Math. Analysis, 8*(11), 513 - 524

Bal, A., & Saha, R. (2016). An Improved Method for Handwritten Document Analysis Using Segmentation, Baseline Recognition and Writing Pressure Detection. *Procedia Computer Science, 93*, 403-415. Retrieved from doi:http://dx.doi.org/10.1016/j.procs.2016.07.227

Bezdek, J. C. (1981). Pattern Recognition With Fuzzy Objective Function Algorithms: *Kluwer Academic* Publishers.

Universiti Utara Malaysia

Bin Wang, Hiroki Yamamoto, Wu, J., & Ejima, Y. (2015). Neuroscience and Biomedical Engineering. *Bentham Science, 1*(2), 102-110. Retrieved from doi: 10.2174/2213385202666140207002441

Bishop, C. M. (1995). Neural Networks for Pattern Recognition. Retrieved from doi:10.4566/j.ins.1995.03.045

Bommes, M., Fazekas, A., Volkenhoff, T., & Oeser, M. (2016). Video Based Intelligent Transportation Systems – State of the Art and Future Development. *Transportation Research Procedia, 14*, 4495-4504. Retrieved from http://dx.doi.org/10.1016/j.trpro.2016.05.372

Bova, N., & G, V. (2016). Deformable models direct supervised guidance. *Neurocomput., 177*(C), 317-333. Retrieved from doi:10.1016/j.neucom.2015.11.023

Boyat, A. K., & Joshi, B. K. (2015). A review Paper: Noise MOodels in Digital Image Prossesing. *Signal & Image Processing : An International Journal (SIPIJ), 6*(2).

Bradbury, J., Merity, S., & Socher, C. X. R. (2017). *QusiI-Recurrent Neural Networks*. Paper presented at *the International Conference on Learning Robotics (ICLR) 2017, Palo Alto, California*.

Byrne, F. (2015a). Encoding Reality: Prediction-Assisted Cortical Learning Algorithm in Hierarchical Temporal Memory. *arXiv:1509.08255v1*.

Byrne, F. (2015b). Encoding Reality: Prediction-Assisted Cortical Learning Algorithm in Hierarchical Temporal Memory. *Neural and Evolutionary Computing, 2*.

Cabral, R. S., De la Torre, F., Costeira, J. P., & Bernardino, A. (2011). Matrix Completion for Multi-label Image Classification. Papper presented to the *6th International Conference on Classification Techniques (ICCT), 201(1).*

Cai, W., Chen, S., & Zhang, D. (2010). Fast and Robust Fuzzy C-Means Clustering Algorithms Incorporating Local Information for Image Segmentation. *Pattern Recognition Letters 0031-3203, 40*(3), 825-838

Chakraborty, S., & Parekh, R. (2015). An Improved Template Matching Algorithm for Car License Plate Recognition. *International Journal of Computer Applications 118*(25).

Chattopadhyay, S., Pratihar, D. K., & Sarkar, S. C. D. (2011). A comparative Study of Fuzzy C-means Algorithm and Entropy-Based Fuzzy Clustering Algorithms. Journal of *Computing and Informatics, 30*, 701-720.

Chen. (2017). Medical Image Segmentation Using Independent Component Analysis-Based Kernelized Fuzzy -Means Clustering. *Mathematical Problems in Engineering, 201721 pages*. Retrieved from doi:10.1155/2017/5892039

Chen, G., Bui, T. D., & Krzyżak, A. (2016). Sparse support vector machine for pattern recognition. *Concurrency and Computation: Practice and Experience, 28*(7), 2261-2273. doi:10.1002/cpe.3492

Chen, Y.-T. (2017). Medical Image Segmentation Using Independent Component Analysis-Based Kernelized Fuzzy -Means Clustering. *Mathematical Problems in Engineering, 201721 pages*. Retrieved from doi:10.1155/2017/5892039

Coutinho, K., & Inoue, T. (2016). Deconstructing and constructing innate immune functions using molecular sensors and actuators. *SPIE 9871*. Retrieved from doi: 10.1117/12.2225185

Das, & De. (2016, 23-25 Sept. 2016). Multilevel color image segmentation using modified genetic algorithm (MfGA) inspired fuzzy c-means clustering. Paper

UUM Universiti Utara Malaysia

presented at the 2016 *Second International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN).*

Das, Pal, Ballester, & Blumenstein. (2014, 9-12 Dec. 2014). *A new efficient and adaptive sclera recognition system.* Paper presented at the *Computational Intelligence in Biometrics and Identity Management (CIBIM), 2014 IEEE Symposium on.*

Das, S. (2013). Pattern Recognition using the Fuzzy c-means Technique. *International Journal of Energy, Information and Communications, 4*(1), 14.

Deepa, & Revathy. (2012). Validation of Document Clustering based on Purity and Entropy measures. *International Journal of Advanced Research in Computer and Communication Engineering, 1*(3), 6.

Diao, J., & Kang, H. (2014, 13-15 July 2014). An Integrated Hierarchical Temporal Memory Network for Real-Time Continuous Multi-interval Prediction of Data Streams. Paper presented at the *2014 Sixth International Symposium on Parallel Architectures, Algorithms and Programming.*

Diplaris, S., Tsoumakas, G., Mitkas, P. A., & Vlahavas, I. (2005). Protein Classification with Multiple Algorithms. In P. Bozanis & E. N. Houstis (Eds.), *Advances in Informatics: 10th Panhellenic Conference on Informatics, PCI 2005, Volas, Greece, November 11-13, 2005. Proceedings* (pp. 448-456). Berlin, Heidelberg: Springer Berlin Heidelberg.

Dobrescu, R., Ichim, L., & Crişan, D. (2013). Diagnosis of Breast Cancer from Mammograms by Using Fractal Measures. *International Journal of Medical Imaging, 1*(2), 32-38. Retrieved from doi:10.11648/j.ijmi.20130102.14

Dornaika, F., Traboulsi, Y. E., & Assoum, A. (2016a). Inductive and flexible feature extraction for semi-supervised pattern categorization. *Pattern Recogn., 60,* 275-285.

Dornaika, F., Traboulsi, Y. E., & Assoum, A. (2016b). Inductive and flexible feature extraction for semi-supervised pattern categorization. *Pattern Recognition, 60,* 275-285. Retrieved from doi:http://dx.doi.org/10.1016/j.patcog.2016.04.024

Doulaty, M., Rose, R., & Siohan, O. (2016). Automatic Optimization of Data Perturbation Distributions for Multi-Style Training in Speech Recognition. Paper presented at the *International Conference on Sequence Learning Techniqes 2016.*

Duan, Chen, X., Houthooft, R., Schulman, J., & Abbeel, P. (2016). Benchmarking Deep Reinforcement Learning for Continuous Control. Paper presented at the *International conference on Memory learning ICML 2016.*

Duan, Yao, M., Wang, J., Bai, T., & Yue, J. (2016). Integrative intrinsic time-scale decomposition and hierarchical temporal memory approach to gearbox diagnosis under variable operating conditions. *Advances in Mechanical Engineering, 1*(1).

Duygulu, P., Barnard, K., Freitas, J. F. G. d., & Forsyth, D. A. (2002). Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary. Paper presented at the ECCV '02 Proceedings of the *7th European Conference on Computer Vision-Part IV, London, UK*

Eyupoglu. (2016, 2-3 Feb. 2016). Implementation of color face recognition using PCA and k-NN classifier. Paper presented at the 2016 *IEEE NW Russia Young Researchers in Electrical and Electronic Engineering Conference (EIConRusNW).*

Fan, & Lin, C. J. (2007). A study on threshold selection for multi-label classification, 1-23. Retrieved from doi:10.1232/j.ins.2007.23.05.

Universiti Utara Malaysia

Fan, Sharad, Sengupta, & Roy. (2016). Hierarchical Temporal Memory Based on Spin-Neurons and Resistive Memory for Energy-Efficient Brain-Inspired Computing. *IEEE Trans Neural Netw Learn Syst, 27*(9). Retrieved from doi:10.1109/TNNLS.2015.2462731

Fathima, N. (2013). Classification of Blood Types by Microscope Color Images. *International Journal of Machine Learning and Computing, 3*(4), 4.

Flores, A., Linguraru, M. G., & Okada, K. (2010). *Boosted-LDA for Biomedical Data Analysis*. Retrieved from doi:10.1223/j.ins.2010.02.06

Fouda. (2015). A Robust Template Matching Algorithm Based on Reducing Dimensions. *Journal of Signal and Information Processing, 6*, 109-122.

Fukushima. (2013a). Artificial Vision by Multi-Layered Neural Networks: Neocognitron and its Advances. *Neural Networks, 37*, 103-119. Retrieved from doi: 10.1016/j.neunet.2013.09.016

Fukushima. (2013b). Training Multi-Layered Neural Network Neocognitron. *elsevier, 40*, 18-31.

Fukushima. (2016, 24-29 July 2016). Margined Winner-Take-All: New learning rule for pattern recognition. Paper presented at the 2016 *International Joint Conference on Neural Networks (IJCNN).*

Fukushima, & Shouno. (2015, 12-17 July 2015). Deep convolutional network neocognitron: Improved Interpolating-Vector. Paper presented at the 2015 *International Joint Conference on Neural Networks (IJCNN).*

Gabrielsson, J., Meibohm, B., & Weiner, D. (2016). Pattern Recognition in Pharmacokinetic Data Analysis. *AAPS Journal*. Retrieved from doi:10.1208/s12248-015-9817-6

Garg, V. (2014). Inductive Group Method of Data Handling Neural Network Approach to Model Basin Sediment Yield *Journal of Hydrologic Engineering, 20*(6).

George. (2008). How the Brain Might Work A Hierarchical and Temporal Model for Learning and Recognition. (PhD), Stanford.

George, & Hawkins. (2005a, 31 July-4 Aug. 2005). A hierarchical Bayesian Model of Invariant Pattern Recognition in the Visual Cortex. Paper presented at the *Neural Networks, 2005. IJCNN '05. Proceedings. 2005 IEEE International Joint Conference on Neural Network*

George, & Hawkins. (2005b). Invariant Pattern Recognition using Bayesian Inference on Hierarchical Sequences. *International Conference on Neural Networks, 2005. IJCNN '05. Proceedings. 2005 IEEE International Joint Conference*

George, & Hawkins. (2012). USA Patent No.: I. Numenta.

Gorokhovatskyi. (2016, 23-27 Aug. 2016). Neocognitron as a tool for optical marks recognition. Paper presented at the 2016 *IEEE First International Conference on Data Stream Mining & Processing (DSMP).*

Gottschlich, C. (2016). Convolution Comparison Pattern: An Efficient Local Image Descriptor for Fingerprint Liveness Detection. *PLOS ONE*.

Grabusts, P. (2011). The Choice of Metrics for Clustering Algorithms. Paper presented at the *International Scientific and Practical Conference., Izdevniecība,.*

Guarneros, M. J., Ochoa, J. A. C., & Trinidad, J. F. M. (2015). Prototype Selection for Graph Embeding Using Instance Selection. Paper presented at the *7th Mexican Conference on Computer Science , maxico.*

Guichard., D. (2015). *Calculus Early Transcendentals*. USA: Creative Commons,.

H, T. A., & Sachin, D. (2015). Dimensionality Reduction and Classification through PCA and LDA. *International Journal of Computer Applications 122*(17).

Han, Kamber, & Pei. (2011). Data Mining: Concepts and Techniques, *3rd Edition* Elsevier (Ed.)

Hawkins, & George. (2006). Hierarchical Temporal Memory--Concepts, Theory, and Terminology. *Whitepaper, Numenta Inc, May*. doi:citeulike-article-id:7020873

He, Z., Liu, L., Zhou, S., & Shen, Y. (2016). Learning group-based sparse and low-rank representation for hyperspectral image classification. *Pattern Recogn., 60*, 1041-1056.

Heinbockel, J. H. (2012). Introduction to Calculus *Volume I* (Vol. 1): Old Dominion University.

Horvath, L., & Khoshnevisan, D. (2010). Weight Functions and Pathwise Local Central Limit Theorems. *Stochastic Processes and Their Applications, 59*(12), 105-123. Retrieved from doi:10.1016/0304-4149(95)00021-X

Hou, S.-w., Feng, S., & Wang, H. (2016). Intelligent Process Abnormal Patterns Recognition and Diagnosis Based on Fuzzy Logic. *Computational Intelligence and Neuroscience, 2016*, 8. Retrieved from doi:10.1155/2016/8289508

http://www.dti.unimi.it/fscotti/all.

Huan, X., Caramanis, & Sanghavi. (2012). Robust PCA via Outlier Pursuit. *Information Theory, IEEE Transactions, 58*(5), 3047-3064. Retrieved from doi:10.1109/TIT.2011.2173156

Jaros, George, Hawkins, & Astier. (2014). Sequence learning in a hierarchical temporal memory based system: Google Patents.

Jasmin Léveillé, Isao Hayashi, & Fukushima, K. (2014). A Probabilistic WKL Rule for Incremental Feature Learning and Pattern Recognition. *Journal of Advanced Computational Intelligence and Intelligent Informatics, 18*(4), 672-681.

Jiang, Trundle, & Ren. (2010). Medical image analysis with artificial neural networks. *Computerized Medical Imaging and Graphics, 34*, 617-631.

Jose, Nair, .Biju, Mathew, & Prashanth. (2016, 18-19 March 2016). Hidden Markov Model: Application towards genomic analysis. Paper presented at the 2016 *International Conference on Circuit, Power and Computing Technologies (ICCPCT).*

Ju, F., Sun, Y., Gao, J., Liu, S., Hu, Y., & Yin, B. (2016). Mixture of Bilateral-Projection Two-Dimensional Probabilistic Principal Component Analysis. Paper presented at the *IEEE Conference on Computer Vision and Pattern Recognition.*

Jun, W., & Shi-Tong, W. (2010). Double Indices FCM Algorithm Based on Hybrid Distance Metric Learning. *Journal of Software, 21*(8), 1878-1888.

Jurie, F., & Dhome, M. (2002). Real Time Robust Template Matching, 10. Retrieved from doi:10.3554/j.ins.2010.02.04

Juvela1. (2016). Template matching method for the analysis of interstellar cloud structure. *Worldwide astronomical and astrophysical research, 593*, 12. Retrieved from doi:10.6554/ jaciii.2016.02.05

Kafrawy, P. E., Mausad, A., & Esmail, H. (2015a). Experimental Comparison of Methods for Multi-Label Classification in Different Application Domains. *International Journal of Computer Applications, 114*(19), 0975 – 8887.

Kafrawy, P. E., Mausad, A., & Esmail, H. (2015b). Experimental Comparison of Methods for Multi-label Classification in different Application Domains. *IJCA Journal, 114*(19). Retrieved from doi:10.5120/20083-1666

Universiti Utara Malaysia

Kahkashan Kouser1, & Sunita. (2013). A comparative study of K Means Algorithm by Different Distance Measures. *International Journal of Innovative Research in Computer and Communication Engineering, 1*(9).

Kalmar, A., & Vida, R. (2013). Extracting High Level Context Information using Hierarchical Temporal Memory *Asian Journal Of Computer Science And Information Technology, 3*(7), 27-34.

Kannan, Ramathilagam, & Pandiyarajan. (2011). Modified bias field fuzzy C-means for effective segmentation of brain MRI *Transactions on computational science VIII* (pp. 127-145). Marina L. Gavrilova: Springer-Verlag.

Karczmarek, P., Kiersztyn, A., Pedrycz, W., & Dolecki, M. (2017). An application of chain code-based local descriptor and its extension to face recognition. *Pattern Recognition, 65*, 26-34. Retrieved from http://dx.doi.org/10.1016/j.patcog.2016.12.008

Kaur, N., & Kaur, U. (2013). Survey of Pattern Recognition Methods. *International Journal of Advanced Research in Computer Science and Software Engineering, 3*(2), 317-319.

Khandelwal, C. S., Maheshewari, R., & Shinde, U. B. (2016). Review Paper on Applications of Principal Component Analysis in Multimodal Biometrics System. *Procedia Computer Science, 92*, 481-486. Retrieved from http://dx.doi.org/10.1016/j.procs.2016.07.371

Kim, D. Y., & Jeon, M. (2014). Data fusion of radar and image measurements for multi-object tracking via Kalman filtering. *Information Sciences, 278*, 641-652. Retrieved from:http://dx.doi.org/10.1016/j.ins.2014.03.080

Klimt, B., & Yang, Y. (2004). The Enron Corpus: A New Dataset for Email Classification Research. *springer*, 217-226.

Kondo. (2011, 11-15 April 2011). *Revised GMDH*-type neural network using artificial intelligence and its application to medical image diagnosis. Paper presented at *the Hybrid Intelligent Models And Applications (HIMA), 2011 IEEE Workshop*

Kostavelis, I., & Gasteratos, A. (2012). On the optimization of Hierarchical Temporal Memory. *Pattern Recognition Letters, 33*(5), 670-676.

Kour, A. (2017). Face Recognition using Template Matching. *International Journal of Computer Applications, 115*(8). Retrieved from doi:10.5120/20170-2329

Kouser1, K., & Sunita. (2013). A comparative study of K Means Algorithm by Different Distance Measures. *International Journal of Innovative Research in Computer and Communication Engineering, 1*(9), 2443-2447.

Kravanja, J., Žganec, M., Žganec-Gros, J., Dobrišek, S., & Štruc, V. (2016). Robust Depth Image Acquisition Using Modulated Pattern Projection and Probabilistic Graphical Models. *Sensors, 16*(10).

Kumar, R. (2011). *Pattern Classification*. Retrieved from Dept.of Computer Science & Engg. S.J.C.E,Mysore:

Kunihiko, F. (2016). Artificial Vision by Deep CNN Neocognitron.I*ntrntional Journal of Cognetive Neural Network* (*IJCNN*), 977-984.

Lai, D., & Garibaldi, J. (2013). Investigating Distance Metrics in Semi-supervised Fuzzy c-Means for Breast Cancer Classification. In L. Peterson, F. Masulli, & G. Russo (Eds.), *Computational Intelligence Methods for Bioinformatics and Biostatistics* (Vol. 7845, pp. 147-157): Springer Berlin Heidelberg.

Larsen, P. M., Schmidt, S., & Schiøtz, J. (2016). Robust Structural Identification via Polyhedral Template Matching. *Modelling and Simulation in Materials Science and Engineering, 24*.

Universiti Utara Malaysia

Léveillé, J., Hayashi, I., & Fukushima, K. (2014). A Probabilistic WKL Rule for Incremental Feature Learning and Pattern Recognition. *JACIII, 18*(4), 672-681. Retrieved from doi:10.20965/jaciii.2014.p0672

Li, & Khashanah, M. (2015, 7-10 Dec. 2015). The Predictive Power of Volatility Pattern Recognition in Stock Market. Paper presented at *the 2015 IEEE Symposium Series on Computational Intelligence.*

Lipton, Z. C., & Berkowitz, J. (2015). A Critical Review of Recurrent Neural Networks for Sequence Learning. *arXiv:1506.00019v4 2015*.

Liu, Lu, Feng, & Zhou. (2017). Learning Deep Sharable and Structural Detectors for Face Alignment. *IEEE Transactions on Image Processing, PP*(99), 1-1. Retrieved from doi:10.1109/TIP.2017.2657118

Lopez-Perez, J. J., Ayala-Ramirez, V., & Hernandez-Belmonte, U. H. (2016). Dynamic Object Detection and Representation for Mobile Robot Application. In J. F. Martínez-Trinidad, J. A. Carrasco-Ochoa, V. Ayala Ramirez, J. A. Olvera-López, & X. Jiang (Eds.), *8th Mexican Conference on Pattern Recognition, MCPR 2016, Guanajuato, Mexico, June 22-25, 2016. Proceedings* (pp. 84-93). Cham: Springer International Publishing.

Lou, X., Li, J., & Liu, H. (2012). Improved Fuzzy C-means Clustering Algorithm Based on Cluster Density. *Journal of Computational Information Systems, 8*(2), 727-737.

Maciej Wielgosza, b., & Pietron´, M. (2016). Noisy video classification with Spatial Pooler of Hierarchical Temporal Memory. Retrieved from doi:10.1201/j.ins.2016.01.05.

Mahalakshmi, Muthaiah, & Swaminathan. (2012). Review Article: An Overview of Template Matching Technique in Image Processing. *Research Journal of Applied Sciences, Engineering and Technology, 4*(24), 5.

Maltoni, D. (2011). Pattern Recognition by Hierarchical Temporal Memory, 45. Retrieved from doi:10.1311/j.ins.2011.22.34.

Maltoni, D., & Lomonaco, V. (2016). Semi-Supervised Tuning from Temporal Cohorence. Paper presented at the *International Conference on Learning Robotics ICLR 2016.*

Manivannan, S., Li, W., Akbar, S., Wang, R., Zhang, J., & McKenna, S. J. (2016). An automated pattern recognition system for classifying indirect immunofluorescence images of HEp-2 cells and specimens. *Pattern Recognition, 51*, 12-26. Retrieved from http://dx.doi.org/10.1016/j.patcog.2015.09.015

Mariyama, T., Fukushima, K., & Matsumoto, W. (2016). Automatic Design of Neural Network Structures Using AiS. In A. Hirose, S. Ozawa, K. Doya, K. Ikeda, M. Lee, & D. Liu (Eds.), *23rd International Conference on Neural Information Processing, (ICONIP) 2016, Kyoto, Japan, October 16–21, 2016, Proceedings, Part II* (pp. 280-287). Cham: Springer International Publishing.

Marqu´es, I. (2010). *Face Recognition Algorithms.* (phd), euskal herriko, euskal herriko.

Mehdizadeh, E., Golabzaei, A., & Chen, K. (2016). Electrical fuzzy C-means: A new heuristic fuzzy clustering algorithm. *Cogent Engineering, 3*(1), 1208397. Retrieved from doi:10.1080/23311916.2016.1208397

Meyer, & Fingscheidt. (2016, 20-25 March 2016). Soft linear discriminant analysis (SLDA) for pattern recognition with ambiguous reference labels: Application to social signal processing. Paper presented at the 2016 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).*

Universiti Utara Malaysia

Min, Z., & Wenming, C. (2015). Recognition of Mixture Control Chart Pattern Using Multiclass Support Vector Machine and Genetic Algorithm Based on Statistical and Shape Features. *Mathematical Problems in Engineering, 2015*, 10. Retrieved from doi:10.1155/2015/382395

Mnatzaganian, J., Fokoué, E., & Kudithipudi, D. (2016). A Mathematical Formalization of Hierarchical Temporal Memory's Spatial Pooler. (MS), Computer Engineering (KGCOE).

Mokeev, & Mokeev. (2015). Pattern recognition by means of linear discriminant analysis and the principal components analysis. *Pattern Recognition and Image Analysis, 25*(4), 685-691. Retrieved from doi:10.1134/s1054661815040185

Mullner, D. (2011). Modern Hierarchical, Agglomerative Clustering Algorithms. *Modern Hierarchical, Agglomerative Clustering Algorithms*. arXiv:1109.2378v1.

Murali. (2015). Principal Component Analysis based Feature Vector Extraction. *Indian Journal of Science and Technology, 8*(35). Retrieved from doi: 10.17485/ijst/2015/v8i35/77760

Mussarat Yasmin, Sharif, M., & Mohsin, S. (2013). Neural Networks in Medical Imaging Applications: A Survey. *World Applied Sciences Journal, 22*(1), 12. Retrieved from doi:10.5829/idosi.wasj.2013.22.01.72188

Mygdalis, V., Iosifidis, A., Tefas, A., & Pitas, I. (2016). Graph Embedded One-Class Classifiers for media data classification. *Pattern Recogn., 60*, 585-595.

Nai, W., Liu, Y., Rempel, D., & Wang, Y. (2017). Fast hand posture classification using depth features extracted from random line segments. *Pattern Recognition, 65*, 1-10. Retrieved from http://dx.doi.org/10.1016/j.patcog.2016.11.022

Nasierding, & Kouzani. (2012). Comparative evaluation of multi-label classification methods. Paper presented at the *2012 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*.

Ng, M., Li, M. J., Huang, J. Z., & He, Z. (2007). On the Impact of Dissimilarity Measure in k-Modes Clustering Algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(3), 503-507.

Numenta, I. (2009). Numenta Node Algorithms Guide – NuPIC 1.7. Retrieved from doi:10.1134/s1054661815040185

Olah, C., & Carter, S. (2016). Attention and Augmented Recurrent Neural Networks. *Distill*.

Osegi, N. (2016). An Improved Intelligent Agent for Mining Real-Time Databases Using Modified Cortical Learning Algorithms *arXiv:1503.07469*.

Otahal, M., Najman, M., & Stepankova, O. (2016). Design of Neuromorphic Cognitive Module based on Hierarchical Temporal Memory and Demonstrated on Anomaly Detection. *Procedia Computer Science, 88*, 232-238. Retrieved from http://dx.doi.org/10.1016/j.procs.2016.07.430

Pablo, Menichini, Larese, M. G., & Riquelme, B. D. (2015). Automatic analysis of microscopic images of red blood cell aggregates. *SPIE, 9531*. Retrieved from doi: 10.1117/12.2181110

Park;, H., Ahn, T., Kim, K., Lee, S., Kook, S.-y., Lee, D., . . . Park, Y. (2015). Three-dimensional refractive index tomograms and deformability of individual human red blood cells from cord blood of newborn infants and maternal blood. *Journal of Biomedical Optics, 20*(11).

Universiti Utara Malaysia

Paul, S., Magdon-Ismail, M., & Drineas, P. (2016). Feature selection for linear SVM with provable guaranteesOriginal Research Article. *Pattern Recognition, 60*, 205-214.

Pechenizkiy, M. (2005). The Impact of Feature Extraction on the Performance of a Classifier: kNN, Naïve Bayes and C4.5. In B. Kégl & G. Lapalme (Eds.), *Advances in Artificial Intelligence* (Vol. 3501, pp. 268-279): Springer Berlin Heidelberg.

Pelillo, M., Elezi, I., & Fiorucci, M. (2016). Revealing structure in large graphs: Szemerédi's regularity lemma and its use in pattern recognition. *Pattern Recognition Letters*. Retrieved from http://dx.doi.org/10.1016/j.patrec.2016.09.007

Peng, L., & Liu, Y. (2018). Feature Selection and Overlapping Clustering-Based Multilabel Classification Model. *Mathematical Problems in Engineering, 2018*, 12. Retrieved from doi:10.1155/2018/2814897

Peter Mahler, L., Søren, S., & Jakob, S. (2016). Robust structural identification via polyhedral template matching. *Modelling and Simulation in Materials Science and Engineering, 24*(5), 055007.

Poletaev, Pervunin, K. S., & Tokarev, M. P. (2017). Artificial Neural Network for Bubbles Pattern Recognition on the Images. *Journal of Physics*. Retrieved from doi:10.1088/1742-6596

Pouliakis, A., Karakitsou, E., Margari, N., Bountris, P., Haritou, M., Panayiotides, J., . . . Karakitsos, P. (2016). Artificial Neural Networks as Decision Support Tools in Cytopathology: Past, Present, and Future. *Biomedical Engineering and Computational Biology, 7*, 1-18. Retrieved from doi:10.4137/BECB.S31601

Rammal, A., Perrin, E., Vrabie, V., Bertrand, I., & Chabbert, B. (2017). Classification of lignocellulosic biomass by weighted-covariance factor fuzzy C-means clustering of mid-infrared and near-infrared spectra. *Journal of Chemometrics*, e2865-n/a. doi:10.1002/cem.2865

Read, J., Martino, L., & Hollmén, J. (2017). Multi-label methods for prediction with sequential data. *Pattern Recognition, 63*, 45-55. Retrieved from http://dx.doi.org/10.1016/j.patcog.2016.09.015

Reddy, K. S. P., & Raju, D. K. N. (2016). Design and Implementation of an Algorithm for Face Recognition by using Principal Component Analysis (PCA) in MATLAB. *International Journal of Advanced Research in Computer Science and Software Engineering, 6*(10).

Rehman, H. U., Azam, N., Yao, J., & Benso, A. (2017). A three-way approach for protein function classification. *PLoS ONE, 12*(2).

Romanuke, V. V. (2016). Optimal Pixel-to-Shift Standard Deviation Ratio for Training 2-Layer Perceptron on Shifted 60 × 80 Images with Pixel Distortion in Classifying Shifting-Distorted Objects. *Applied Computer Systems, 19*(1), 61–70. Retrieved from https://doi.org/10.1515/acss-2016-0008,

Rowland, M. M., Coe, P. K., Stussy, R. J., Ager, A. A., Cimon, N. J., Johnson, B. K., & Wisdom, M. J. (1998). *The Starkey Habitat Database for Ungulate Research: Construction, Documentation, and Use*. Pacific Northwest, Research Station, USA.

Rozado, D. (2011). Analysis and Extension of Hierarchical Temporal Memory for Multivariable Time Series. (PhD), Universidad Autonoma de Madrid

Universiti Utara Malaysia

Ryu, S., Kim, S., Choi, J., Yu, H., & Lee, G. G. (2017). Neural sentence embedding using only in-domain sentences for out-of-domain sentence detection in dialog systems. *Pattern Recognition Letters, 88*, 26-32. Retrieved from http://dx.doi.org/10.1016/j.patrec.2017.01.008

Sao, P., Hegadi, R., & Karmakar, S. (2014). A Literature Review on Approaches of ECG Pattern Recognition *International Journal of Information Science and Intelligent System, 3*(2), 79-90.

Sechidis, K., Tsoumakas, G., & Vlahavas, I. (2011). On the Stratification of Multi-label Data. In D. Gunopulos, T. Hofmann, D. Malerba, & M. Vazirgiannis (Eds.), *European Conference on Machine Learning and Knowledge Discovery in Databases, ECML PKDD 2011, Athens, Greece, September 5-9, 2011, Proceedings, Part III* (pp. 145-158). Berlin, Heidelberg: Springer Berlin Heidelberg.

Serb, A., Bill, J., Ali Khiat, R. B., Legenstein, R., & Prodromakis, T. (2016). Unsupervised learning in Probabilistic Neural Networks with Multi-state Metal-oxide Memristive Synapses. *Nature Communications,, 7*(1261), 1-9.

Shahrour. (2014). Virtual DMA Municipal Water Supply Pipeline Leak Detection and Classification Using Advance Pattern Recognizer Multi-Class SVM. *Journal of Pattern Recognition Research, 9*(1).

Shwartz, S. S., & Ben-David, S. (2014). Understanding Machine Learning: From Theory to Algorithms. Retrieved from doi:10.2344/j.ins.2014.44.55.

Singh, S. P., & Nair, K. (2013). Intelligent Controller for Reduction of Total Harmonics In Single Phase Inverters *American Journal of Applied Sciences, 10*(11), 1378-1385. Retrieved from doi:10.3844

Singla, S., & Khera, B. S. (2016). A Comparative Study of K-Means, Fuzzy C-Means and Possibilistic Fuzzy C-Means Algorithm on Noisy Grayscale Images. Paper presented at the *International Conference on Advances in Emerging Technology*.

Skrynnik, A., Petrov, A., & Panov, A. (2016). Hierarchical Temporal Memory Implementation with Explicit States Extraction. In Samsonovich, A. V. Klimov, V. V. Rybina, & G. V (Eds.), *Biologically Inspired Cognitive Architectures (BICA)* (pp. 219-225): Springer International Publishing.

Soliman, A., Khalifa, F., Elnakib, A., El-Ghar, M. A., Dunlap, N., Wang, B., . . . El-Baz, A. (2017). Accurate Lungs Segmentation on CT Chest Images by Adaptive Appearance-Guided Shape Modeling. *IEEE Transactions on Medical Imaging, 36*(1), 263-276. Retrieved from doi:10.1109/TMI.2016.2606370

Soltani, R., & Jiang, H. (2017). Higher Order Recurrent Neural Networks. Paper presented at the *International Conference on Learning Robotics (ICLR) 2017, Torento*

Song, J. P., Zhu, Z., Scully, P., & Price, C. (2013). Selecting Features for Anomaly Intrusion Detection : A novel Method using Fuzzy C-Means and Dicision Tree Classification. Paper presented at the 5*th International Symposium, CSS 2013, Zhangjiajie, China*.

Streich. (2010). Multi-label Classification and Clustering for Acoustics and Computer Security*: American Journal of Applied Sciences, 12(*11), 1378-1385. Retrieved from doi:10.1244/j.ins.2010.02.44.

Sutskever, I., & Hinton, G. (2009). Temporal Kernel Recurrent Neural Networks, 15. Retrieved from doi:10.1522/j.ins.2009.02.45

Universiti Utara Malaysia

Tabacchi, Asensio, Pavón, Recuero, Mir, & Artal. (2013). A statistical pattern recognition approach for the classification of cooking stages. The boiling water case. *Applied Acoustics, 74*(8), 1022-1032.

Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*. Retrieved from doi:10.1186/s12880-015-0068-x

Tan, L., McGarry, M. D. J., Houten, E. E. W. V., Ji, M., Solamen, L., Weaver, J. B., & Paulsen, K. D. (2017). Gradient-Based Optimization for Poroelastic and Viscoelastic MR Elastography. *IEEE Transactions on Medical Imaging, 36*(1), 236-250. Retrieved from doi:10.1109/TMI.2016.2604568

Thornton, J., Main, L., & Srbic, a. A. (2012). Fixed Frame Temporal Pooling. Retrieved from doi:10.1233/s14880-010-0023-x

Tokuda, K., & Zen, H. (2016). Directly Modeling Voiced and Unvoiced Components in Speech Waveforms by Neural Networks. Paper presented at the *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP).*

Tsoumakas, & Katakis. (2015). Multi-Label Classification: An Overview. *International Journal of Data Warehousing and Mining, 3*(3), 1-13.

Tu, Z., Xie, W., Cao, J., van Gemeren, C., Poppe, R., & Veltkamp, R. C. (2017). Variational method for joint optical flow estimation and edge-aware image restoration. *Pattern Recognition, 65*, 11-25. Retrieved from http://dx.doi.org/10.1016/j.patcog.2016.10.027

Uga, M., Dan, I., Sano, T., Dan, H., & Watanabe, E. (2014). Optimizing the general linear model for functional near-infrared spectroscopy: an adaptive hemodynamic response function approach. *Neurophotonics, 1*(1).

Wielgosz, M., & Pietro´n, M. (2016). OpenCL-accelerated object classification in video streams using Spatial Pooler of Hierarchical Temporal Memory. Retrieved from doi:10.1255/j.ins.2016.44.28.

William, R. (2012). Hierarchical Temporal Memory Cortical Learning Algorithm for Pattern Recoginition: *ProQuest*, UMI Dissertation Publishing (October 17, 2012).

Wong. (2017). Parametric methods for comparing the performance of two classification algorithms evaluated by k-fold cross validation on multiple data sets. *Pattern Recognition, 65*, 97-107. Retrieved from http://dx.doi.org/10.1016/j.patcog.2016.12.018

Wong, Summers, R. M., Kebebew, E., & Yao, J. (2017). Pancreatic Tumor Growth Prediction With Elastic-Growth Decomposition, Image-Derived Motion, and FDM-FEM Coupling. *IEEE Transactions on Medical Imaging, 36*(1), 111-123. Retrieved from doi:10.1109/TMI.2016.2597313

Wu, & Deng, W. (2015). Adaptive Quotient Image with 3D Generic Elastic Models for Pose and Illumination Invariant Face Recognition. In J. Yang, J. Yang, Z. Sun, S. Shan, W. Zheng, & J. Feng (Eds.), *10th Chinese Conference on Biometric Recognition, (CCBR), Tianjin, China, November 13-15, 2015, Proceedings* (pp. 3-10). Cham: Springer International Publishing.

Wu, & Toet, A. (2014). Speed-Up Template Matching Through Integral Image Based Weak Classifiers *ournal of Pattern Recognition Research, 9*(1), 12.

Xia, X., Chen, Z., Luan, F., & Song, X. (2017). Signature alignment based on GMM for on-line signature verification. *Pattern Recognition, 65*, 188-196. Retrieved from http://dx.doi.org/10.1016/j.patcog.2016.12.019

UUM Universiti Utara Malaysia

Xie, L., Yang, K., Peng, J., Gao, X., & Tan, Y. (2016). Application of Object Recognition in Locomotive Components Monitoring. Paper presented at the *19th World Conference on Non-Destructive Testing 2016.*

Xu, R., & Wunsch, D. (2005). Survey of Clustering Algorithms. *IEEE Transactions On Neural Networks, 16*(3), 645-678.

Yu, Y.-F., Dai, D.-Q., Ren, C.-X., & Huang, K.-K. (2017). Discriminative Multi-Scale Sparse Coding for Single-Sample Face Recognition with Occlusion. *Pattern Recognition*. Retrieved from http://dx.doi.org/10.1016/j.patcog.2017.01.021

Zambrano, J. G., Guzmán-Ramírez, E., & Pogrebnyak, O. (2013). Search Algorithm for Image Recognition Based on Learning Algorithm for Multivariate Data Analysis: Intech publishing. Retrieved from doi: 10.5772/52179

Zanaty, E. A. (2013). An Adaptive Fuzzy C-Means Algorithm for Improving MRI Segmentation. *Open Journal of Medical Imaging, 3*(4), 125-135. Retrieved from doi: 10.4236/ojmi.2013.34020.

Zhang, An, L., Xu, J., Zhang, B., Zheng, W. J., Hu, M., . . . Yue, F. (2018). Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus. *Nature Communications, 9*(1), 750. Retrieved from doi:10.1038/s41467-018-03113-2

Zhang, Liu, F., Liu, J., Luo, J., Xie, Y., Bai, J., & Xing, L. (2017). Cone Beam X-ray Luminescence Computed Tomography Based on Bayesian Method. *IEEE Transactions on Medical Imaging, 36*(1), 225-235. Retrieved from doi:10.1109/TMI.2016.2603843

Zhang, & Shen, L. (2014). An Improved Fuzzy C-Means Clustering Algorithm Based on Shadowed Sets and PSO. *Computational Intelligence and Neuroscience, 7*, 368-388. Retrieved from doi:10.1155/2014/368628

Zhang, Z., Zhao, M., & Chow, T. W. S. (2013). Binary- and Multi-class Group Sparse Canonical Correlation Analysis for Feature Extraction and Classification. *IEEE Transaction on Knowledge and Datsa Engineering, 25*(10), 2192-2205.

Zhou. (2000). Digital Image Processing and Interpritation. Retrieved from doi:10.1186/s12880-015-0068-x

Zhou, Xie, L., Shen, W., Fishman, E., & Yuille, A. (2016). Pancreas Segmentation in Abdominal CT Scan: A Coarse-to-Fine Approach. *arXiv:1612.08230v, 2016.*

Zhu, L., & Ma, L. (2016). Class centroid alignment based domain adaptation for classification of remote sensing images. *Pattern Recognition Letters, 83, Part 2*, 124-132. Retrieved from http://dx.doi.org/10.1016/j.patrec.2015.12.015

Zipfel. (2014). Plant pattern-recognition receptors. *Trends Immunol, 35*(7). Retrieved from doi:10.1016/j.it.2014.05.004

Zytoon, A. A. (2014). Standardized Uptake Value Variations of Normal Glandular Breast Tissue at Dual Time Point FDG-PET/CT Imaging. *International Journal of Medical Imaging, 1*(3), 56-65. Retrieved from doi:10.11648/j

Universiti Utara Malaysia