MODERN FUZZY MIN MAX NEURAL NETWORKS FOR PATTERN CLASSIFICATION

OSAMA NAYEL AHMAD AL SAYAYDEH

DOCTOR OF PHILOSOPHY

UNIVERSITI MALAYSIA PAHANG

UNIVERSITI MALAYSIA PAHANG

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name : OSAMA NAYEL AHMAD AL SAYAYDEH

Date of Birth : 13 MAY1985

Title : MODERN FUZZY MIN MAX NEURAL NETWORK FOR PATTERN CLASSIFICATION

Academic Session : SEM 2 2018/2019

I declare that this thesis is classified as:

☐	CONFIDENTIAL	(Contains confidential information under the Official Secret Act 1997)*
☐	RESTRICTED	(Contains restricted information as specified by the organization where research was done)*
☑	OPEN ACCESS	I agree that my thesis to be published as online open access (Full Text)

I acknowledge that Universiti Malaysia Pahang reserves the following rights:

1. The Thesis is the Property of Universiti Malaysia Pahang
2. The Library of Universiti Malaysia Pahang has the right to make copies of the thesis for the purpose of research only.
3. The Library has the right to make copies of the thesis for academic exchange.

Certified by:

_______________________ (Student's Signature)

_______________________ (Supervisor's Signature)

____
New IC/Passport Number
Date:

____
Name of Supervisor
Date:

NOTE : * If the thesis is CONFIDENTIAL or RESTRICTED, please attach a thesis declaration letter.

MAKLUMAT PANEL PEMERIKSA PEPERIKSAAN LISAN

Tesis ini telah diperiksa dan diakui oleh
This thesis has been checked and verified by

Nama dan Alamat Pemeriksa Dalam Name and Address Internal Examiner	: Assoc. Prof. Dr. Mohd Nizam Bin Mohmad Kahar Faculty of Computer Systems & Software Engineering, Universiti Malaysia Pahang
Nama dan Alamat Pemeriksa Luar *Name and Address External Examiner*	: Prof. Dr. Nor Ashidi Mat Isa School of Electrical and Electronic Engineering, Engineering Campus Universiti Sains Malaysia
Nama dan Alamat Pemeriksa Luar *Name and Address External Examiner*	: Prof. Dr. Ali Bin Selamat *Malaysia Japan International Institute of Technology (MJIIT),Universiti Teknologi Malaysia KL*

Disahkan oleh Timbalan Pendaftar di IPS
Verified by Deputy Registrar IPS

Tandatangan : Tarikh:
Signature *Date:*

Nama :
Name

UMP

SUPERVISOR'S DECLARATION

We hereby declare that We have checked this thesis and in our opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Doctor of Philosophy.

____

(Supervisor's Signature)

Full Name : MOHAMMAD FALAH MOHAMMAD

Position : SENIOR LECTURER

Date :

____

(Co-supervisor's Signature)

Full Name : LIEW SIAU CHUIN

Position : SENIOR LECTURER

Date :

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

____

(Student's Signature)

Full Name : OSAMA NAYEL AHMAD AL SAYAYDEH

ID Number : PCC16014

MODERN FUZZY MIN MAX NEURAL NETWORK FOR PATTERN CLASSIFICATION

OSAMA NAYEL AHMAD AL SAYAYDEH

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Doctor of Philosophy/Master of Science/Master of Engineering

Faculty of Computer Systems & Software Engineering

UNIVERSITI MALAYSIA PAHANG

AUGUST 2019

ACKNOWLEDGEMENTS

First of all, many thanks to my supervisor, Dr. Mohammad Falah Mohammad for his great patience and knowledgeable guidance during the past three years PhD study. His professional research experience teaches me how to be creative, how to find problems and solve them. His active attitude of life encourages me working hard towards my destination. In addition, I would like to thank him for improving and polishing my writing skills. I would like to express the deepest appreciation also for Dr. Liew Siau Chun for advising support.

My distinctive thanks to my beloved wife- Eng.Abeer binti Moh'd; you were a complete package of family and love me strength, support, and love during tough times. My daughter Retan and son Aws, you are my world, my words, my strength. Thank you

I am greatly grateful to my parents, Mr. Nayel bin Ahmad, Ibtehaj binti Abdullah, for their endless love, unconditional support, and encouragement through all these years. I am also thankful to my sisters and my brothers, Eng.Ansam, Ashwaq, Dr. Wasn, Dr. Mohammed, Dr. Ehab, and Dr. Anas for their support during many years of my studies.

Finally, I would like to thank all staff at the faculty of Computer Systems & software Engineering, University of Malaysia Pahang for their support.

ABSTRAK

Kebelakangan ini, terdapat tumpuan yang mendalam terhadap kaedah pengkomputeran lembut bagi mengatasi masalah dunia sebenar yang kompleks. Rangkaian neural dan logik kabur merupakan antara kaedah pengkomputeran lembut yang sering digunakan dalam bidang klasifikasi corak. Dalam membina sebuah model pengelas yang berkesan, para penyelidik memperkenalkan model hibrid yang menggabungkan kedua-dua logik kabur dan rangkaian neural buatan. Antara algoritma yang diperkenalkan, algoritma rangkaian neural Min-Max Kabur (FMM) terbukti sebagai salah satu rangkaian neural terulung dalam mengatasi masalah klasifikasi corak. FMM mempunyai pelbagai fitur penting, berkebolehan untuk menyediakan proses pembelajaran dalam talian dan menangani masalah kelupaan. Namun sedemikian, algoritma ini turut menghadapi beberapa batasan, khususnya dalam proses pembelajarannya, iaitu proses pengembangan, proses ujian bertindih dan proses pengecutan. Oleh itu, rangkaian neural Min-Max Kabur Moden (MDFMM) diperkenalkan dengan tujuan mengatasi batasan-batasan FMM asal. MDFMM membawa kepada beberapa sumbangan seperti mengubah suai fungsi pengaktifan pengembangan FMM asal dan menggantikannya dengan Min-Max Kabur Tertingkat (EFMM) untuk menyingkirkan kes bertindih. Pertama, kajian ini mencadangkan kaedah pengembangan baru untuk mengatasi masalah kelonggaran dan ketaksamaan pengembangan hiperboks yang bertindih. Akibatnya, kaedah ini dapat mengurangkan proses pengecutan. Kedua, kajian ini mengusulkan satu formula ujian bertindih baru yang mempermudahkan proses ujian bertindih FMM/EFMM dengan meliputi semua kes bertindih yang mungkin dengan sempurna. Ketiga, kajian ini mencadangkan satu proses pengecutan baru yang memberikan gambaran hiperboks yang lebih tepat dan menghindari masalah herotan data (kehilangan maklumat hiperboks). Keempat, kajian ini mengusulkan strategi ramalan baru dalam fasa ujian dengan menyepadukan persamaan jarak bersama fungsi keahlian untuk menangani masalah pembuatan keputusan rawak. Strategi ini dapat membantu untuk memberikan ramalan yang lebih tepat ketika sampel input mempunyai nilai kesesuaian yang sama dengan kelas lain. Bagi mengatasi kerumitan struktur rangkaian MDFMM, satu penambahbaikan diperkenalkan dengan meningkatkan pemilihan hiperboks yang menang dalam proses pengembangan menggunakan algoritma k-terdekat. Model cadangan baru dinamakan sebagai MDFMM-Kn. Prestasi MDFMM dan MDFMM-Kn dinilai menggunakan pelbagai set data tanda aras UCI dan set data kecerdasan buatan 2D. Di samping itu, tiga kaedah analisis statistik, iaitu kaedah bootstrap, pengesahsahihan silang k-fold dan ujian taraf bertanda Wilcoxon, digunakan untuk menyatakan kuantiti prestasi secara statistik. Berdasarkan penilaian empirik, MDFMM yang dicadangkan didapati lebih baik berbanding model sedia ada (rangkaian neural Min-Max Kabur Ubah Suai, atau MFMMN) dari segi ketepatan dengan peratusan peningkatan sebanyak 35.42%. Tambahan pula, prestasi purata MDFMM-kn berbanding model FMM dan MDFMM menunjukkan prestasi yang lebih baik dari segi kerumitan dengan peratusan sebanyak 62%.

ABSTRACT

In the recent years, the world has demonstrated an increasing interest in soft computing techniques to deal with complex real world problems. Neural network and fuzzy logic are considered to be one of the most popular soft computing techniques that applied in pattern classification domain. To build an efficient classifier model, researchers have introduced hybrid models that combine both fuzzy logic and artificial neural networks. Among these algorithms, Fuzzy Min Max (FMM) neural network algorithm has been proven to be one of the premier neural networks for undertaking the pattern classification problems. Although the FMM has many important features with the ability to provide online learning process and can handle the forgetting problem, it suffers from a number of limitations, especially in its learning process i.e., expansion process, overlapping test process, and contraction process. Therefore, Modern Fuzzy Min Max neural network is introduced with aim of overcoming the specified limitations of the original FMM. The MDFMM introduces a number of contributions in addition to modify the original FMM expansion activation function by replace it with that from the Enhanced Fuzzy Min Max (EFMM) to eliminate the overlapping cases. First, this study proposed a new expansion technique to overcome both overlap leniency and irregularity of hyperbox expansion problems, as a result, reducing the number of contraction processes. Secondly, proposing a new overlapping test formula that simplify the FMM/EFMM overlap test process with perfectly covers all the possible overlapped cases. Thirdly, proposing a new contraction process that provides more accurate hyperboxes description and avoid data distortion problem (hyperbox information losses). Fourthly, proposing a new prediction strategy in the test phase by integrating the distance equation with membership function in order to solve the randomization decision making problem, which helps to provide more accurate prediction when input sample has same fitness values with different classes. To overcome the network structure complexity of MDFMM, a further improvement is introduced by improving the selection of the winning hyperbox during the expansion process using the k-nearest neighbours algorithm (MDFMM-Kn). The performance of MDFMM and MDFMM-Kn was evaluated using different UCI benchmark datasets and 2D artificial intelligence dataset. Furthermore, three statistical analysis techniques, namely, bootstrap method, k-fold cross-validation and the Wilcoxon signed-rank test, were utilized to statistically quantify the performances. From the empirical evaluation, the proposed MDFMM is better than the recent existing model modified FMM network (MFMMN) in terms of accuracy at an improvement percentage of 35.42%. Furthermore, the average performance of the MDFMM-Kn against the FMM and MDFMM models is better than that of the existing techniques in terms of complexity at a percentage of 62%.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

θ	User defined expansion parameter
Δ	Value of overlap
γ	Sensitive parameter

LIST OF ABBREVIATIONS

AFMN	Adaptive Fuzzy Min Man Neural
AI	Artificial Intelligence
ANN	Artificial Neural Network
ART	Adaptive Reasoning Theory
CLN	Classify Neuron
CNN	Containment Compensation Neuron
DCFMN	Data Core Fuzzy Min Max Neural Network
EFC	Inclusion/Exclusion Fuzzy Classifier
EFMM	Enhanced Fuzzy Min Max
EGFMM	Enhanced General Fuzzy Min Max
ES	Evolutionary Strategies
FAM	Fuzzy ARTMAP
FMCN	Fuzzy Min Max Compensatory neurons
FL	Fuzzy Logic
FMM	Fuzzy Min Max
MFMM-GA	Modified Fuzzy Min Max – Genetic Algorithm
FMM-KN	Fuzzy Min Max- Key Nearest
FNN	Fuzzy Neural Network
GA	Genetic Algorithm
GFMM	General Fuzzy Min Max
GP	Genetic Programming
GRFMN	General Reflex Fuzzy Min Max Neural
MDFMM	Modern Fuzzy Min Max
MDFMM-Kn	Modern Fuzzy Min Max Enhanced Key Nearest
M-FMCN	Modified- Fuzzy Min Max Compensatory neural
MFMM1	Modified Fuzzy Min Max 1
MFMM2	Modified Fuzzy Min Max 2
MFMM-GA	Modified Fuzzy Min Max with genetic algorithm
MFMMN	Modified Fuzzy Min Max Neural Network
MLF	Multilevel Fuzzy
MLP	Multi-Layer Perceptron

NFS	Neuro Fuzzy Systems
OCN	Overlap Compensation Neuron
OCR	Optical Character Recognition
PR	Pattern Recognition
RBFN	Radial Basis Function Network
SLP	Single Layer Perceptron
WFMM	Weighted Fuzzy Min Max

CHAPTER 1

INTRODUCTION

1.1 Background

Numerous species inhabit the earth and are stuck in a struggle for survival. Although not all strong, some of these species manage to stay alive and adapt to their surroundings whilst others die in huge numbers to the point of extinction. Amongst these living beings, only humans are given a rational mind that has transformed them into the most powerful creature on earth with full control over other species. When facing hardships, humans have sought for various means to alleviate their situation. These challenges have driven them to engage in research and development across various fields, especially in science. This constant engagement in research has inspired many innovations that are continuously being developed up to this day. However, with the constant emergence of these innovations, the amount of relevant data produced each day has increased to a point that they can no longer be stored in the human mind. As the traditional means for storing data are becoming obsolete, humans have resorted to computers as advanced technologies for storing and retrieving information in large amounts. Computers are also able to simulate the human memory, thereby allowing them not only to store information but also make rational decisions.

Researchers in computer science have attempted to develop a smart machine that can engage in a learning process similar to humans. However, several challenges have delayed the development of such technology, including the complex processes in a human brain when making decisions. Therefore, researchers have devoted much effort in studying the decision-making process of the human brain. One of the pioneering researchers in this field was Alan Turing, whose paper served as an introduction to the new concept of Artificial Intelligence (AI) (Alan 1950; Jain et al. 2000), which, over the past few decades, has become one of the most important fields in computer science and applied mathematics.

AI has a wide range of applications, including in diagnosing patients (Begum and Devi 2011; Umoh 2012), forecasting the weather (Barde and Patole 2016), automating industrial processes and performing agricultural and geological research (Raj et al. 2015). AI is also considered a key component of an intelligent machine that can simulate the human learning process through the use of algorithms. With the recent research advancements in this field, people have begun to apply AI in various areas, including in data mining (Kamruzzaman and Jehad Sarkar 2011), robotics, email spam filtering (Shrivastava and Bindu 2014), financial analysis (Tkáč and Verner 2016) and especially pattern recognition (PR) (Liu et al. 2018).

PR is deemed useful in many applications, including optical character recognition (Mithe et al. 2014), industrial inspection, speech recognition (Hau 2015; Saksamudre et al. 2015), biomedical studies (i.e. neuroscience, ECG monitoring, drug development and DNA sequences), biometrics (i.e. face, fingerprint and iris recognition) (Unar et al. 2014) and military applications (Raj et al. 2015).

Pattern classification has eventually become a popular area in PR given its important role in several applications, including engineering control, medical diagnosis (Begum & Devi, 2011), signal processing (Reaz et al. 2006), speech recognition and data mining (Qu 2009). Given their increasing significance, researchers have resorted to using various classification approaches to develop high-accuracy classifiers. Classification problems are often addressed by using artificial neural networks (ANNs), which are particularly useful for managing noisy data collected from real environments. Several types of ANNs are often utilised in pattern classification, such as Hopfield network, neuro-fuzzy network, multilayer perceptron, Boltzmann machine and radial basis function network (Oludolapo et al. 2012; Yilmaz et al. 2010; Yilmaz and Kaynar 2011).

However, the extant neural networks can only support offline (batch) learning (Mohammed and Lim 2015), which examines how a specific task can be performed. No further improvements or changes can be observed in the system after the learning phase, during which no task can be performed. Therefore, offline learning faces a problem called '*catastrophic forgetting*', where the system cannot add or update its knowledge whenever new information is available. In other words, the ANNs previous knowledge will be forgotten whenever adding a new information.

Even though they can manage noisy and nonlinear data from real environments, ANNs face the catastrophic forgetting problem (also known as the stability–plasticity dilemma) when dealing with new information in an incremental manner (Mohammed and Lim 2015; Simpson 1992). When facing such problem, the system requires plasticity and stability in order to combine new information and to retain previous information, respectively (Mohammed and Lim 2015, 2017b; Simpson 1992). In other words, to integrate new knowledge and retain previous knowledge, a parallel and distributed system must possess plasticity and stability, respectively. When increasing its plasticity, the system forgets its previously learned knowledge; meanwhile, increasing stability will affect the efficient learning of the system at the level of the synapses (Mermillod et al. 2013). This dilemma poses several problems, including

- how can the learning system stay plastic or adaptive when receiving significant input patterns and stay stable in response to irrelevant input patterns?
- how can the system learn when to switch from the plastic mode to the stable mode and vice versa?
- how can the system adapt to a new environment whilst retaining its previously learned information?

Several ANNs with incremental learning abilities have been proposed to address these questions and to solve the stability–plasticity dilemma. These ANNs include the FMM network (Simpson, 1992) and adaptive reasoning theory (ART, including its various forms such as fuzzy ART, ARTMAP and ART1; (Carpenter et al. 1991a; Carpenter et al. 1991b; Moore 1988).

To overcome the stability–plasticity dilemma, ART networks insert a feedback mechanism between the input and competitive layers of a network to enable the system to (1) learn new information without necessarily forgetting its previously stored information and (2) automatically switch between the plastic and stable modes. Accordingly, several types of ART neural networks have been introduced, including fuzzy ARTMAP (FAM); (Kim Hua et al. 2004; Kuan et al. 2003). Despite effectively handling classification problems in different domains, FAM has several limitations, the most important of which is the overlapping of hyperboxes (Simpson 1992). In other words, an input pattern may be granted full class membership to various classes, thereby negatively affecting the classification rate. To address this problem, Simpson introduced supervised and unsupervised FMM neural networks. This work selects the supervised FMM network as its backbone model given its focus on the classification task.

In general, an FMM neural network consist of three layers: input, hidden, and output layers. FMM able to absorb information online, and there is no need to retrain the network when new information is available. The FMM has the following key features that enable the network to address the problems in pattern classification (Mohammed and Lim 2015; Simpson 1992):

- *Online learning*: The FMM can quickly learn new classes and refine existing ones without forgetting the previous information. Online learning contrasts offline learning such that both the old and new information are retained in the former.
- *Nonlinear*: The FMM builds a nonlinear decision boundary to separate different classes from one another irrespective of their size and shape.
- *Overlapping classes*: The FMM builds a decision boundary to reduce the number of misclassification patterns within the overlapping region.
- *Training time*: The FMM can learn and revise the nonlinear decision boundary with one-pass learning within a short training time.
- *Soft and hard decisions:* The FMM can support hard and soft classification decisions, of which the former means that the pattern is within the class (0) or not (1) whilst the latter quantifies the fitness degree of a pattern within a class.

In general, an FMM neural network consists of two stages i.e., learning and testing stage. Where the learning involves three main processes, namely, the expansion process in which a new input pattern is included by expanding the hyperbox, the overlap test process to identify the overlapped areas between hyperboxes belongs to various classes, and the contraction process in which the overlapped area is eliminated if deemed necessary. All three processes allow the elimination of overlapped areas between hyperboxes belonging to different classes and subsequently address the limitations of FAM. While testing stage involves the membership function, which uses to measure the degree to which input sample fits within a hyperbox.

1.2 Problem statement and Motivations

Despite these salient features, the FMM neural network can still be improved in many ways. Specifically, some room for improvement can be observed in the overlapping test, expansion, membership function and contraction processes of this algorithm. The expansion process is utilized to extend the wining hyperbox such that it contains the new input pattern. However, applying expansion equation to activate the expansion process can distort the network structure due following reason. Firstly, there is leniency in creating an overlap between hyperboxes from different classes whenever the expansion constraint is satisfied. In this case, the expanded hyperbox will cause an overlap region with another hyperbox regardless whether both are from the same classes or otherwise. Accordingly, more overlap regions will be generated, which negatively affect the classifier performance. Secondly, there is irregularity of the hyperbox expansion procedure that leads one or more hyperbox dimensions to occupy a large area at the expense of other dimensions, thereby increasing the misclassification rate by generating unnecessary overlapped areas between classes. In fact, most of FMM variants such as: data core FMM (DCFMN), multilevel FMM (MLF), MFMMN, enhanced general FMM (EGFMM), and k-nearest FMM (FMM-Kn) still suffering from this limitation which adversely affects the network performance and generate inaccurate predication (Davtalab et al. 2014; Mohammed and Lim 2017a; Shinde and Kulkarni 2016; Zhang et al. 2011).

The overlap test is used to specify whether an overlapped region occurs between existing hyperbox and expanded hyperbox which belongs to different classes. Most of FMM variants, which include MLF, EGFMM, enhanced FMM (EFMM), FMM-Kn, EFMMII, DCFMN, and MFMMN use this process suffers from two main limitations: a complex overlap test process and the boundary points are considered non-overlapped regions in the FMM. That could affect the network decision making natively during the test stage (Davtalab et al. 2014; Donglikar and Waghmare 2017; Mohammed and Lim 2015, 2017a, 2017b; Shinde and Kulkarni 2016; Zhang et al. 2011).

Contraction process is utilized to eliminate the overlap between hyperboxes that belong to different classes. Most of FMM variants which include: EGFMM, adaptive FMM (AFMM), EFMM, EFMMII, FMM-kn, and MFMMN use the FMM contraction process suffer from the data distortion problem (Liu et al. 2012; Mohammed and Lim 2015, 2017a, 2017b; Shinde and Kulkarni 2016). Data distortion happens due to lose of part of contracted hyperbox during the contraction producer (Bargiela et al. 2004; Kim et al. 2004). Consequently, inaccurate decision boundary is generated, thus decreasing classification accuracy.

Another shortcoming of FMM and its variants, During the testing stage, the membership function of FMM and its variants can result in a random prediction especially when the fitness value of the input sample is the same across different classes. This type of predication could negatively affect the classification performance. All these limitations can influence the performance of the classifier and introduce additional misclassification errors. Meanwhile, learning with large datasets may increase the complexity of the network structure by generating more number of hyperboxes. To address these issues, this thesis searches for ways to enhance the accuracy and reduce the complexity of the FMM classifier and to subsequently improve its performance in pattern classification.

1.3 Research Questions

The following research questions (REQ) are addressed in this work:

1- REQ1: What is the significance of the FMM neural network in the area of pattern classification?

2- REQ2: What are the limitations of the FMM neural network and its variants?

3- REQ3: How can these limitations be addressed to improve the classification accuracy of the FMM neural network?

4- REQ4: How can the complexity of the network structure be addressed whilst simultaneously maintaining or improving the classification performance of the FMM neural network?

5- REQ5: How can the recent modifications influence the performance of classifiers in processing datasets, specifically their accuracy, complexity and noise tolerance?

1.4 Research Objectives

This work aims to propose new models improve the pattern classification efficiency of the FMM network. The specific objectives of this work are

1. ROB1: To analyse the weaknesses, drawbacks and limitations of the extant FMM model and its variants when attempting to solve problems in pattern classification.
2. ROB2: To introduce a new expansion technique, overlap test formula, contraction process and prediction strategy for building the modern fuzzy min–max (MDFMM) model.
3. ROB3: To further improve the MDFMM model via the integration of the k-nearest technique, which can reduce the complexity of the network structure and improve the noise tolerance of the model.
4. ROB4: To assess the usefulness of the proposed models in solving problems related to pattern classification by using different benchmarks and 2D AI (noisy/noise-free) datasets, and to utilise various statistical indicators to quantify the performance of these models and to compare and analyse the effectiveness with various classifiers.

Figure 1.1 presents the mapping structure among the stated research questions, objectives and the defined limitations in the problem statement section. ROB1 of this work is mapped to the REQ1 and REQ2 that intend to perform a comprehensive investigation of the supervised FMM network with specifying the limitations in FMM variants.

Further, ROB2 is associated to REQ3 to introduce MDFMM as a new model with the aim of handling the limitations of leniency in creating an overlap, irregularity of hyperbox expansion, complex overlap test process, boundary points and the data distortion problem. The proposed MDFMM introduces four new contributions. Firstly, a new expansion technique that can address the problems in hyperbox expansion, including irregularity and overlap leniency. Secondly, a new overlap test formula is proposed to address the potential overlaps and reduce the complexity of the original overlapping rules. Thirdly, a new contraction process is proposed in order to generate highly accurate decision boundaries and address the data distortion problem. Fourthly, a new predication strategy that integrates the distance formula, a membership function to improve the decision making during the testing stage. The ROB3 is linked to the REQ4 that aims to further improve the performance of the proposed MDFMM in terms of catering for the complexity issue by developing a highly comprehensive model called MDFMM-Kn.

Furthermore, ROB4 is linked to the REQ5, in which the performance of the proposed MDFMM and MDFMM-Kn models are assessed with respect to accuracy and complexity of the network structure.

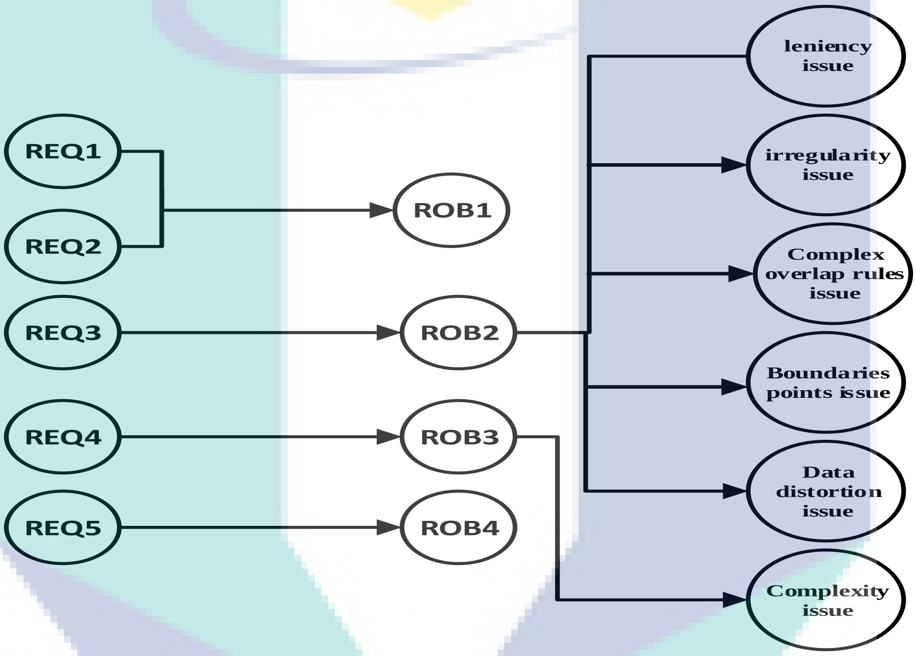


Figure 1.1 The mapping structure among the Research Objectives, Questions and problem statement

1.5 Research Scope

To examine pattern classification problems, this work applies the FMM neural network as its backbone model due to its unique and salient features (i.e. nonlinearity, online learning capability, insulation of overlapping classes, support for soft and hard decisions, ability to overcome the stability–plasticity dilemma and short training time). The FMM neural network is thoroughly analysed in this work. A total of 15 UCI benchmark datasets, including Wine, Heart, Glass, Iris, Liver, Thyroid, Seed, Zoo, Ionosphere, PID, WBC, Page blocks, Ozone level, Spambase and Parkinson’s, as well as 4 different 2D AI datasets, including half kernel, two spirals, outliers and corners, were used to evaluate the performance of both MDFMM and MDFMM-Kn. Furthermore, three statistical methods have been used i.e., bootstrap method, k-fold cross validation, and Wilcoxon signed rank test to evaluate the performance of both MDFMM and MDFMM-Kn models.

1.6 Research Methodology

The main aim of this work is to introduce a new model that can overcome the limitations of extant FMM models. Figure 1.2 illustrates the research methodology, which is also outlined below:

1. Review the FMM model and its variants, analyse the problems in the learning and testing stages and identify the necessary required improvements.
2. Examine the testing and learning stages of MDFMM and apply modifications to each of these stages. In the learning stage, this study will introduce a new expansion technique to address the problems relating to overlapping cases, develop an overlap test formula that is applicable to any possible overlap case and can also simplify the original overlap test rules, and propose a new contraction process that can address the data distortion problem and reduce the complexity of the original contraction process. In the testing stage, the study will integrate the distance equation into the membership function to develop a new strategy that can overcome the problems related to randomised decision making.
3. Use different benchmark datasets to assess the performance of MDFMM and then analyse the results and compare them with those of other classifiers proposed in the literature. Conduct a statistical analysis by using k-fold cross validation, bootstrap methods and Wilcoxon signed rank test.
4. Reduce the complexity of the MDFMM network and improve its noise tolerance when processing different noisy and/or noisy free datasets. Introduce the k-nearest method to improve the structure of the MDFMM model and build the MDFMM-Kn model.
5. Use different benchmarks as well as noisy and noise-free datasets to assess the performance of MDFMM-Kn and then analyze and compare the results with those of MDFMM and other classifiers reported in the literature. Conduct a statistical analysis by using k-fold cross validation and bootstrap methods.

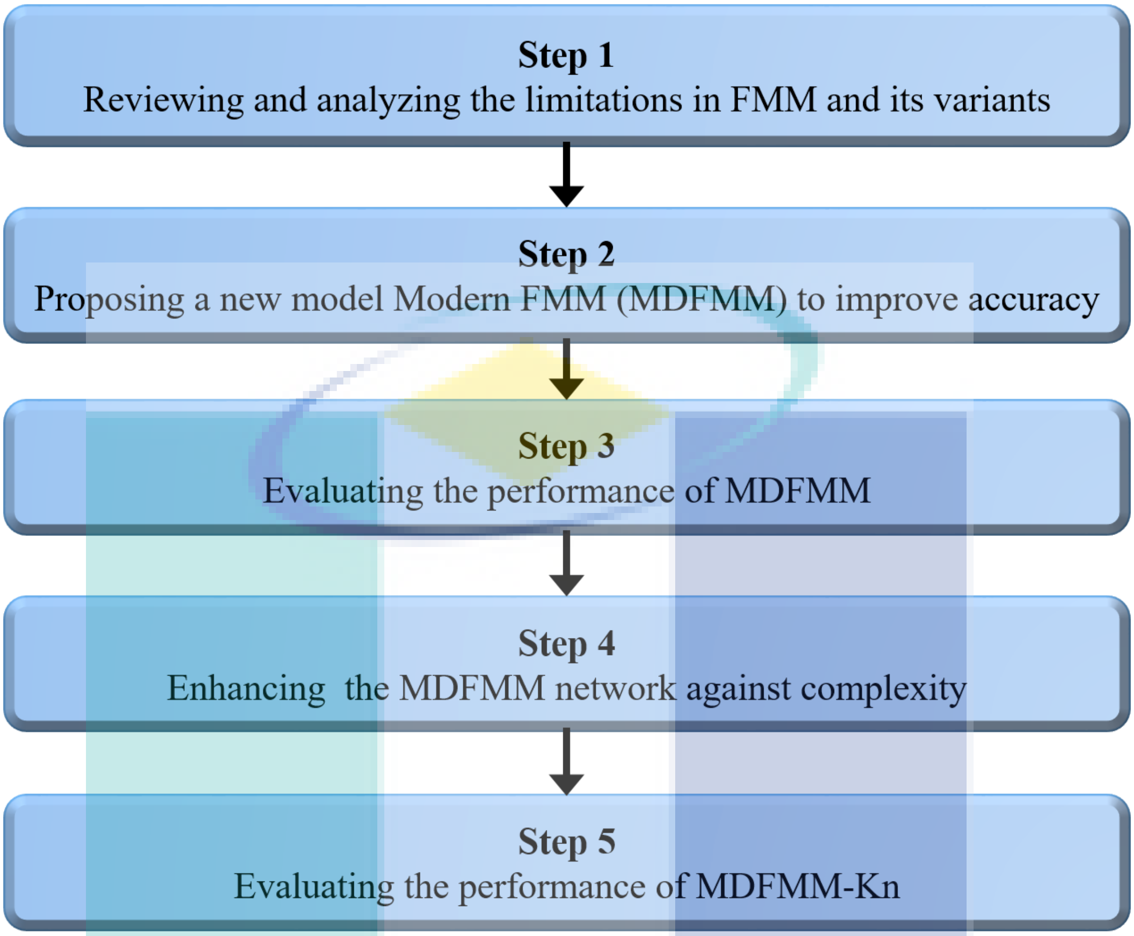


Figure 1.2 Research Methodology

1.7 Thesis Outline

The arrangement of this thesis is patterned against its objectives. Chapter 1 presents the introduction to this work whilst Chapter 2 reviews the pattern classification techniques, including FMM, and highlights the methods, modifications and techniques that are applied to improve the performance of the original FMM model.

Chapter 3 describes the FMM network and its limitations in detail. The modifications applied to the existing network with an aim to address its limitations and improve its performance are also discussed. The MDFMM neural network and its performance are then evaluated by using different benchmark datasets. The evaluation results are presented and compared with those of FMM variants or other classifiers.

Chapter 4 presents those problems relating to the complexity of the MDFMM network. Some existing methods for addressing the issues in FMM are also reviewed, including the k-nearest method. The MDFMM network is then modified to develop MDFMM-Kn, which can address the problems related to the complex structure and noise tolerance of the former. Some experiments are also conducted by using various benchmarks as well as noisy and noise-free datasets to evaluate the performance of MDFMM-Kn. The evaluation results are statistically analysed via bootstrap and k-fold validation methods and compared with those of other FMM variants.

Chapter 5 summarises the entire thesis and its contributions as well as presents several areas that may be explored in future research.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

As explained in Chapter 1, the main focus of this work is to improve the fuzzy min max neural network (FMM) performance for tackling the pattern classification problems. Based on that, an investigation about the efficiency of the FMM-based network as a useful and usable pattern classification system is covered. As such, this chapter presents a highlight on the pattern classification techniques, as well as, a detail elaboration of the FMM neural network learning process. In addition, a review on the FMM network and its variants has been introduced in this chapter. A research gap and summary of the review has been presented at the end of this chapter, respectively.

2.2 Background

Pattern classification is concerned with the ability to find categorical labels for a set of observation. It is an active area because it has been widely used in a number of real world applications including weather forecasting (Barde and Patole 2016), medical diagnosis (cancer, heart diseases)(Begum and Devi 2011; Mohammed and Lim 2015; Quteishat et al. 2013; Umoh 2012), Optical character recognition (OCR) (Naz et al. 2014), fault diagnosis industrial (cooling system, induction motors, robotics)(Duan et al. 2007; Gao et al. 2015; Quteishat et al. 2010; Seera and Lim 2014; Seera et al. 2013), face detection(Kim et al. 2006), intrusion detection(Azad and Jha 2017), object recognition (Pawar 2015), and speaker recognition (Jawarkar et al. 2011). Over the past years, many techniques have been developed for pattern classification. Among the neuromas techniques that are dedicated to tackling pattern classification problems, two main techniques, namely, ANNs and FL, are the most popular.

ANNs are computational models that able to learn, handle large sample, and tackle noisy data that are obtained from real world environments (Genaro et al. 2009). However, several drawbacks have limited their application in practice such as: ANNs lack a formal approach to identifying the optimal structure of a network, difficulty tackling imprecise or ambiguous information and suffer from the black box problem. To overcome these shortcomings, researchers integrated the ANN with FL to form the fuzzy neural network. This integration led to the generation of more robust classifier designs even when applied to real world problems. e.g., ANFIS Fuzzy (Jang 1993), Fuzzy ARTMAP (FAM); (Carpenter et al. 1991a), and the fuzzy min-max (FMM) neural networks (Simpson 1992; Simpson 1993).

FNN is one of the most popular hybrid methods because it combines the main advantages of ANNs and FL: ability to learn and deal to the imprecise data in constructing robust classifiers. In 1985, Keller and Hunt were the first one to work on the integration of fuzzy logic within the neural network (Keller and Hunt 1985). They suggested a way of incorporating the concept of fuzzy sets into perceptron (single-layer) for pattern recognition. The fuzzification of the single-layer perceptron improves the convergence problem when the datasets are nonlinearly separable. However, the fuzzification of the single-layer perceptron was applied on a single layer, which means that the computation will be simply done for this reason. Another work was proposed in 1992 by Pal and Mitra to overcome the limitations of the Keller and Hunt model (Pal and Mitra 1992). Pal and Mitra worked to extend the fuzzification from single layer to the MLP network. In fuzzy MLP, the input vectors consist of membership values while the output vector is defined in terms of the fuzzy class membership value.

Although the use of fuzzy within the neural network led to many successes, especially for real applications(Viharos and Kis 2015), it still inherits the limitations of the MLP. One of the limitations is related to "catastrophic forgetting". The catastrophic forgetting problem happens when the neural network loses all the previous information while trying to learn a new set of information. This problem has motivated Carpenter & Grossberg in 1987 in order to tackle the problem and led them to introduce the adaptive resonance theory (ART) network that provides online learning (Carpenter and Grossberg 1987). Carpenter et al., later developed fuzzification of ART, and suggested the Fuzzy ART in 1991 (Carpenter et al. 1991b), and Fuzzy ARTMAP (FAM) networks in 1992 (Carpenter et al. 1991a). Fuzzy ART and FAM are able to learn both binary and analog input patterns. Intersection and union operators in the ART and FAM network are replaced by the maximum and minimum operators in the fuzzy set theory. Although the FAM solved the "catastrophic forgetting" problem, it faced a problem that allowed it to overlap between the hyperboxes, despite their classes being different. For this reason Simpson in (1992 and 1993) proposed two versions of fuzzy min max neural network (FMM) to tackle the limitation in the ARTMAP network (Simpson 1992; Simpson 1993).

The FMM network was introduced in 1992, which is the backbone for this research, deals with pattern classification, while the FMM was proposed in 1993 deals with pattern clustering. The FMM network uses fuzzy sets as pattern classes where each fuzzy set is an aggregate of fuzzy set hyperboxes. The learning algorithm of FMM uses series of activities on creating, expanding, and contracting hyperboxes. Further details about the FMM for pattern classification are explained in Section 2.3. Over past few years, a number of modifications were proposed in the FMM network; these modifications will be explained in Section 2.4.

2.3 Overview of the Fuzzy Min-Max Neural Networks

The FMM for pattern classification is a supervised learning model proposed by Simpson in 1992 (Simpson 1992). This network builds decision boundaries by creating hyperboxes in the pattern space. Each hyperbox is characterized by a pair of minimum and maximum points in n-dimensional space with a membership function. The training algorithm in FMM consist of three stages, i.e., expansion, overlap test, and contraction. During learning, an input pattern with its class is selected. Then, a hyperbox from the same class that has the highest membership degree is selected and expanded to include the input pattern. If hyperbox fails to expand and absorb the input pattern a new hyperbox is created. Overlapping among hyperboxes of the same class is allowed, while overlapping among hyperboxes from different classes is forbidden. A contraction process is triggered when overlapping between two hyperboxes belonging to two different classes occurs. The FMM learning process is shown in Figure 2.1.

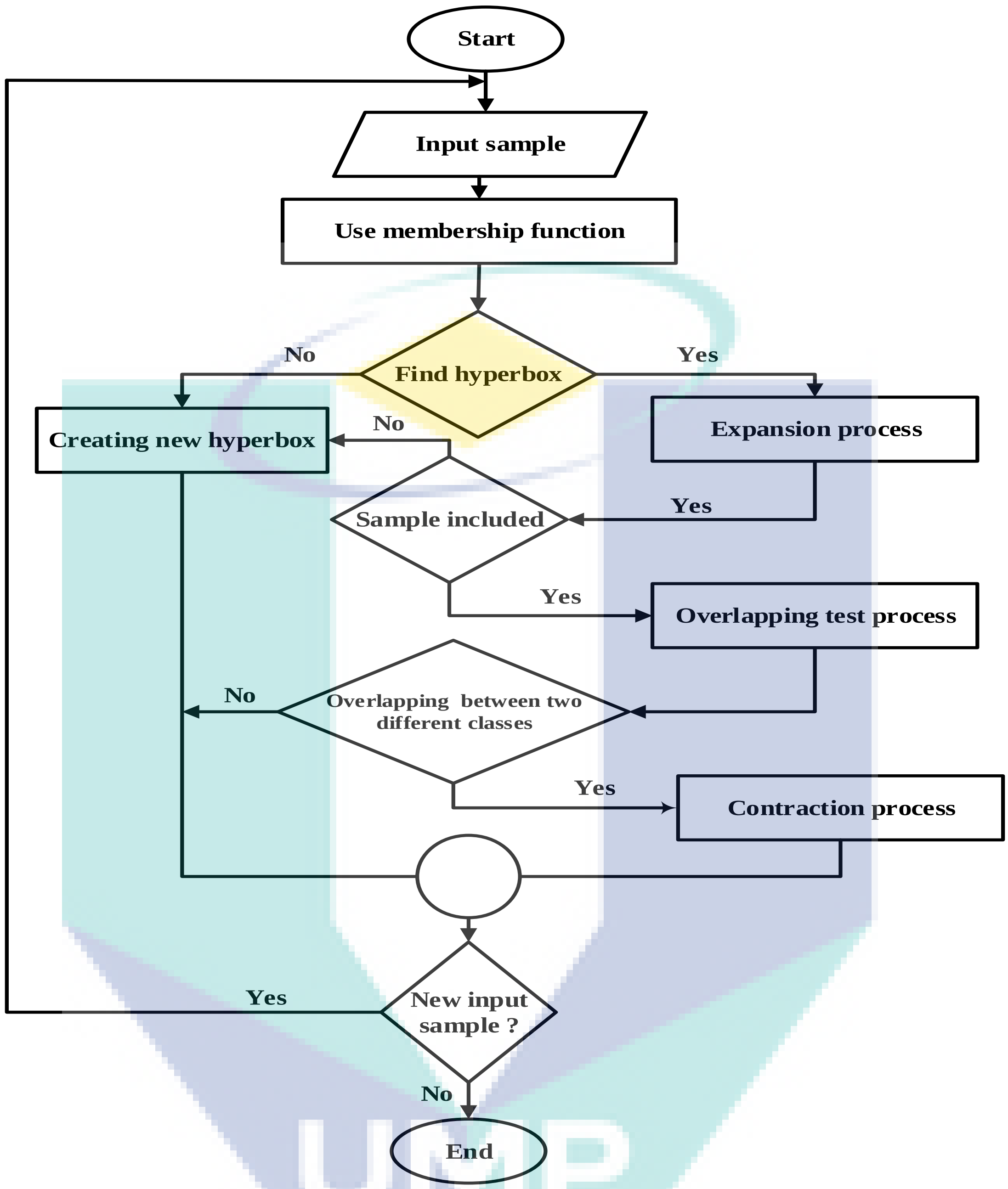


Figure 2.1 The FMM learning process

In general, the FMM classifier consists of three-layers i.e., input, hidden, and output layers. In the first layer, the number of neurons is equal to the number of features, while the number of neurons in output layer is equal to the number of classes. The number of neurons in the hidden layer is equal to the number of hyperboxes that are generated during the learning process, Figure 2.2 shows the FMM network structure (Mohammed and Lim 2017a). Each hyperbox in the hidden layer also represents a fuzzy set, and a hyperbox is normally characterized as a pair of minimum (V_j) and maximum (W_j) points in n-dimensional space, as shown in Figure 2.3.

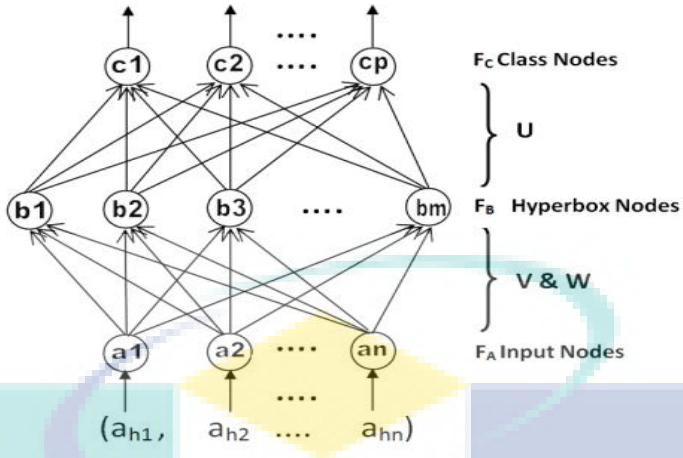


Figure 2.2 The FMM structure

Source: Mohammed and Lim (2017)

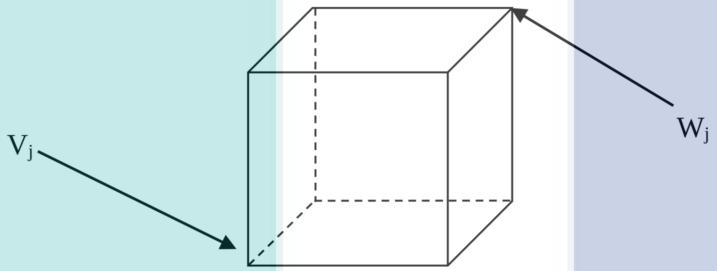


Figure 2.3 A 3-D hyperbox structure

FMM uses a membership function to measure the belongingness of the input pattern with respect to the *j*-th hyperbox. The membership value ranges from 0 to 1. In this case, the input pattern contained in the hyperbox has a full class membership. The membership degree of the input pattern with respect to the j^{th} hyperbox decreases whenever the distance between the input pattern and the j^{th} hyperbox increases. The membership value is calculated using:

$$b_j(A_h) = \frac{1}{2n}\sum_{i=1}^{n}[\max(0, 1-\max(0, \gamma\min(1, a_{hi}-w_{ji}))) + \max(0, 1-\max(0, \gamma\min(1, v_{ji}-a_{hi})))] \quad 2.1$$

Where $b_j(A_h)$ represents the membership function; $A_h = (a_{h1}, a_{h2}, a_{h3}, \ldots, a_{hn}) \in I^n$ is the h^{th} input pattern; γ is the sensitivity parameter that controls how fast the membership value decreases as the distance between input pattern A_h and hyperbox B_j increases; and $V_j = (v_1, v_2, v_3, \ldots, v_n)$ and $W_j = (w_1, w_2, w_3, \ldots, w_n)$ are the minimum and maximum points of the hyperbox, respectively. Mathematically, the *j*-th hyperbox fuzzy set B_j is defined as follows (Simpson 1992):

$$B_j = \{A_h, V_j, W_j, f(A_h, V_j, W_j)\} \forall A \in I^n \qquad 2.2$$

In the FMM, the hyperbox size is regulated by an expansion parameter (θ) that varies between 0 and 1. A smaller expansion parameter leads to generating a smaller hyperbox size, which, as a result, creates a larger number of hyperboxes, and otherwise. As shown in Figure 2.2, the FMM network structure consists of three layers, i.e., input layer, F_A, hidden (hyperbox) layer, F_B, and output layer, F_C. The number of nodes in F_A equals to the dimension of the input pattern, while the number of nodes in F_C equals to the number of target classes. Each node in F_B represents a hyperbox (fuzzy rule). The connection between F_A and F_B nodes are the minimum and maximum points. The connection between F_B and F_C is in binary values and is stored in a matrix, *U*, described as follows:

$$u_{jk} = \begin{Bmatrix} 1 & if \;\; b_j \; is \; a \; hyperbox \; for \; class \;\; c_k \\ 0 & otherwise \end{Bmatrix} \qquad 2.3$$

Where b_j is the jth F_B node and c_k is the kth F_C node with. Given an input pattern, the node in the F_C layer that has the highest membership degree is selected as the winner. The transfer function for each node in the F_C layer is defined as:

$$c_k = \max_{j=1}^{m} b_j u_{jk} \qquad 2.4$$

Where c_k is the k^{th} output layer node , u_{jk} is a binary value that represents the connections between F_b nodes and F_c nodes, and *m* represent the number of hyperboxes in the F_b layer (as shown in Figure 2.2). The FMM network uses either a hard or soft decision to determine the winning class. If a soft decision is required, the output is used directly. Correspondingly, the output of the F_C layer can be 0 or 1 (hard decision), where one indicates the winning c_k class. As discussed previously, the FMM learning process consist of three steps: the expansion process, the overlap test, and the contraction process, if necessary (if all dimensions from hyperboxes that belong to different classes are overlapped). When a new input sample is provided during the learning process, the degree of membership is computed via Eq. 2.1 to determine the winning hyperbox. The winning hyperbox is the hyperbox that has the highest membership degree and contains the input sample; a new hyperbox is constructed if it does not contain the input sample. Typically, the expansion process generates various types of overlapped areas. Partially overlapped dimensions between hyperboxes from the different classes are allowed, whereas fully overlapped dimensions are forbidden. To solve the overlap problem, an overlap test process is performed to identify the dimension that has the minimum overlapped region. Then, a contraction process is activated to eliminate the identified region. The FMM learning phase can be described as follows:

a. **Expansion***:* Before the expansion process begins, the membership degrees are computed to determine the winning hyperbox. The winning hyperbox is expanded to contain the input sample if the expansion coefficient satisfies Eq. 2.5; otherwise, a new hyperbox is created.

$$n\theta \geq \sum_{i=1}^{n} (\max(w_{ji}, a_{hi}) - \min(v_{ji}, a_{hi})) \quad 2.5$$

Where θ is a user-defined parameter that is used to control the maximum size of hyperbox. The hyperbox value range is ($0 \leq \theta \leq 1$). If Eq. 2.5 is satisfied, the winning hyperbox min and max points should be modified via the following two equations:

$$v_{ji}^{new} = \min(v_{ji}^{old}, a_{hi}) \quad 2.6$$

$$w_{ji}^{new} = \max(w_{ji}^{old}, a_{hi}) \quad 2.7$$

Where v_{ji}^{new} and w_{ji}^{new} are new minimum and maximum points of the hyperbox that is expanded.

b. **Overlap Test***:* The overlap test is utilized to check if the hyperbox expansion caused overlap with a hyperbox from a different class. FMM uses four rules to determine the overlapped area, where each dimension in the expanded hyperbox is compared with the same dimension in the hyperbox that is from a different class. The hyperboxes are determined to overlap if any of equations (Eq. 2.8, Eq. 2.9, Eq. 2.10, or Eq. 2.11) is satisfied.

$Case\ 1:$

$$V_{ji} < V_{ki} < W_{ji} < W_{ki}\ , \delta^{new} = min(W_{ji} - V_{ki}\ , \delta^{old}\ ) \quad 2.8$$

$Case\ 2:$

$$V_{ki} < V_{ji} < W_{ki} < W_{ji}\ , \delta^{new} = min(W_{ki} - V_{ji}\ , \delta^{old}\ ) \quad 2.9$$

$Case\ 3:$

$V_{ji} < V_{ki} < W_{ki} < W_{ji}\ ,$

$$\delta^{new} = min(min(W_{ji} - V_{ki}, W_{ki} - V_{ji})\ , \delta^{old}\ ) \quad 2.10$$

$Case\ 4:$

$V_{ki} < V_{ji} < W_{ji} < W_{ki}\ ,$

$$\delta^{new} = min(min(W_{ji} - V_{ki}, W_{ki} - V_{ji})\ , \delta^{old}\ ) \quad 2.11$$

Initially assume that δ^{old} =1. If δ^{old} - δ^{new} > 1, then Δ = *i* and $\delta^{new} = \delta^{old}$. This shows that overlap detects for Δth dimension and test proceeds for next dimension. Otherwise, testing stops and goes for contraction test with Δ= -1. If the hyperboxes from dissimilar classes overlaps with each other, then contraction process is started so that, the overlap of hyperboxes can be eliminated from overlapping regions (Simpson 1992). Nevertheless, overlapping areas occurred by hyperboxes from the similar class are acceptable.

c. **Contraction:** The contraction process is initiated if all dimensions that belong to different classes are overlapped; hence, the contraction process is used to create one pure dimension by eliminating the minimally overlapped dimension. The contraction rules are defined as follows:

$Case\ 1:$

$$V_{j\Delta} < V_{k\Delta} < W_{j\Delta} < W_{k\Delta}\ .\ W_{j\Delta}^{new} = V_{k\Delta}^{new} = \frac{W_{j\Delta}^{old} + V_{k\Delta}^{old}}{2} \quad 2.12$$

$Case\ 2:$

$$V_{k\Delta} < V_{j\Delta} < W_{k\Delta} < W_{j\Delta}\ . W_{k\Delta}^{new} = V_{j\Delta}^{new} = \frac{W_{k\Delta}^{old} + V_{j\Delta}^{old}}{2} \quad 2.13$$

$Case\ 3a:\ V_{j\Delta} < V_{k\Delta} < W_{k\Delta} < W_{j\Delta}\ \ and$ 2.14

$$(W_{k\Delta} - V_{j\Delta}) < (W_{j\Delta} - V_{k\Delta})\ . V_{j\Delta}^{new} = W_{k\Delta}^{old}$$

$Case\ 3b: V_{j\Delta} < V_{k\Delta} < W_{k\Delta} < W_{j\Delta}\ \ and$ 2.15

$$(W_{k\Delta} - V_{j\Delta}) > (W_{j\Delta} - V_{k\Delta})\ . W_{j\Delta}^{new} = V_{k\Delta}^{old}$$

$Case\ 4a: V_{k\Delta} < V_{j\Delta} < W_{j\Delta} < W_{k\Delta}\ \ and$ 2.16

$$(W_{k\Delta} - V_{j\Delta}) < (W_{j\Delta} - V_{k\Delta}). W_{k\Delta}^{new} = V_{j\Delta}^{old}$$

$Case\ 4b: V_{k\Delta} < V_{j\Delta} < W_{j\Delta} < W_{k\Delta}\ \ and$ 2.17

$$(W_{k\Delta} - V_{j\Delta}) > (W_{j\Delta} - V_{k\Delta}). V_{k\Delta}^{new} = W_{j\Delta}^{old}$$

2.4 FMM variants

Although the FMM network is considered an effective online learning model, there is still plenty of room to enhance the FMM network(Mohammed and Lim 2015). In particular, the expansion, overlap test, contraction process, and membership function need further improvements in order to make FMM more powerful and accurate in classification. Researchers have proposed many FMM variants to enhance the classification performance. in this research, FMM and its variants have been classified into two groups as shown in Figure 2.4. The first group covers FMM variants that keep the original FMM learning stages (expansion, overlap test, and contraction) along with modification or enhancements, while the second highlights FMM variants that eliminate the contraction process.

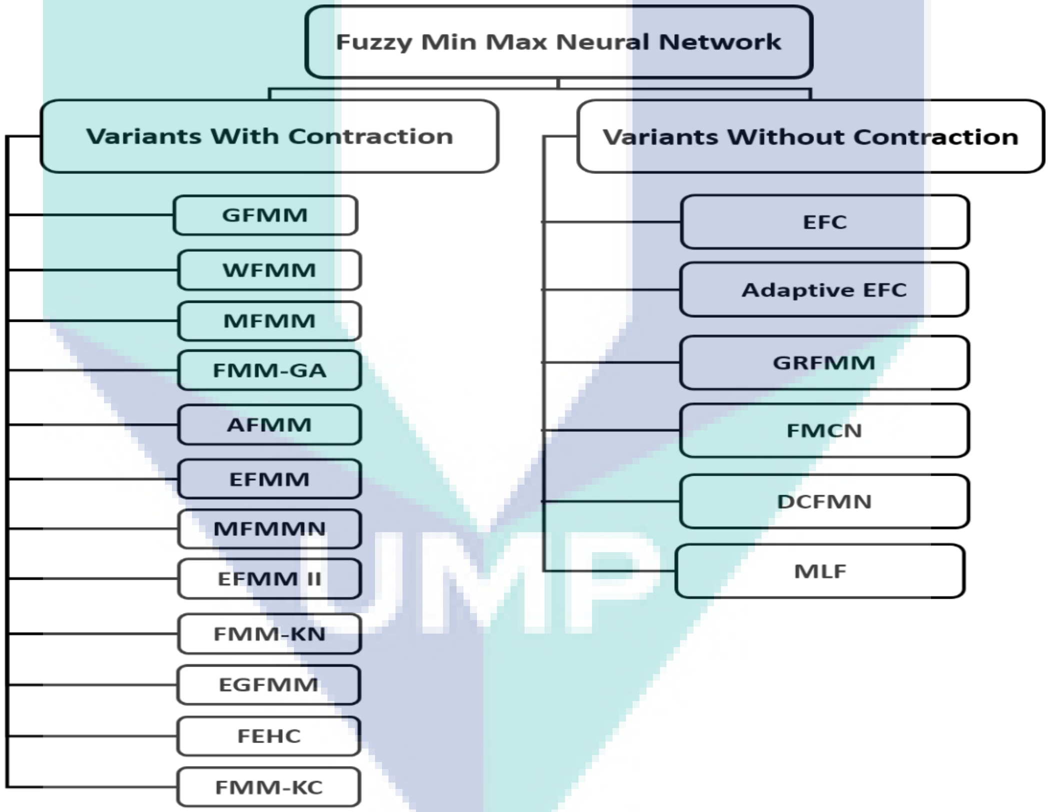


Figure 2.4 FMM variants with and without contraction procedure

2.4.1 FMM variants with contraction

There are a number of FMM variants that maintain the same learning stage, and apply some modifications to overcome the FMM shortcomings or improve its performance by modifying the expansion procedures and the membership function, as well as extending the overlap test rules, or contraction rules. This group of variants utilizes the contraction process to obtain a pure dimension between overlapped hyperboxes belong to different classes. Several FMM variants using contraction process are explained as follow:

In 2000, Gabrys and Bargiela proposed a general FMM model called GFMM (Gabrys and Bargiela 2000). GFMM appears to be the first variant to improve the performance of the original FMM network by addressing the following issues: the inability to distinguish between ignorance and equal interpretation of membership degrees, the inability to simultaneously address labelled and unlabelled data and interval analysis.

GFMM can simultaneously process labelled and unlabelled input patterns by combining supervised and. unsupervised learning into a single algorithm. This feature enables the use of GFMM in three different modes: pure classification, clustering, and hybrid modes (partial supervision). Several changes are introduced to improve efficiency of FMM. The modified fuzzy membership function of GFMM differs from that of FMM, which is a new formulation to compute the membership values. Simultaneously, the sensitive parameter for regulating the maximum hyperbox size can be changed adaptively during the learning phase of GFMM. The input patterns can be fuzzy hyperboxes or crisp points in the pattern space.

Moreover, a change in hyperbox expansion is observed in GFMM, as compared with that in FMM. The GFMM algorithm defines a new constraint, which ensures that the differences between the minimum and maximum points of the individual dimension do not exceed a user-specified limit. GFMM has been compared with FMM in tackling classification and clustering tasks in a single-pass training scenario. Comparatively, GFMM produces fewer hyperboxes and exhibits lower misclassification rates. GFMM uses the same contraction process as supervised learning FMM and assists in achieving the minimal overlapped dimensions of different classes. However, the application of this contraction process results in classification errors for labelled data.

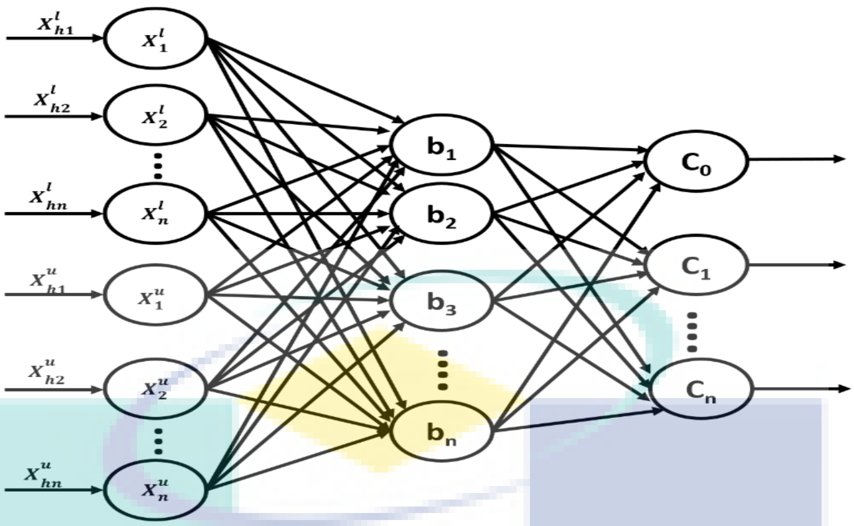


Figure 2.5 GFMM network topology
Source: Gabrys and Bargiela (2000)

In 2004, Kim and Yang proposed a new variants of FMM named as a weighted FMM network (WFMM)(Kim et al. 2004). A hyperbox is not subject to expansion, either with the consideration of its contraction process or the overlap test. The feature distribution information is utilized in the course of the learning process in order to compensate for the distortion of the hyperbox, which may be caused due to the elimination of the overlapping area of hyperboxes during the contraction process. The weight concept is added for the purpose of reflecting the frequency factor of the feature values. According WFMM, because the weight factor effectively reflects the relationship between the feature range and its distribution, the model can prevent an undesirable performance degradation caused by noisy patterns. However, this method also inherits the limitations related to expansion, overlap test, and contraction producers of the original FMM model (Mohammed and Lim 2015).

In 2008, Quteishat and Lim designed a modified version of the FMM, named as (MFMM) (Quteishat and Lim 2007). MFMM enhances the performance of FMM when the size of the expansion parameter is large. It uses the Euclidian distance and membership function to select the wining hyperbox for predicting the winning target class, as shown in Figure 2.6. MFMM also reduces the complexity of FMM using pruning strategy. The drawbacks of MFMM are the same as the FMM learning phase as mentioned earlier. Later, the same authors developed a hybrid model MFMM and the Genetics algorithm (MFMM-GA) for pattern classification and rule extraction (Quteishat et al.

2010). The first stage is used to reduce the MFMM complexity by applying a pruning strategy. In second stage, a “don’t care “strategy is applied by a genetic rule extractor for decreasing the number of features in the extracted rules. Evaluated using benchmark data sets from the UCI machine learning repository and a real medical diagnosis task, the classification performance of MFMM-GA is better than that of MFMM and pruned MFMM. Although it obtained significant result, it relied on the FMM rules (expansion, overlap test, and contraction), which themselves have a number of limitations.

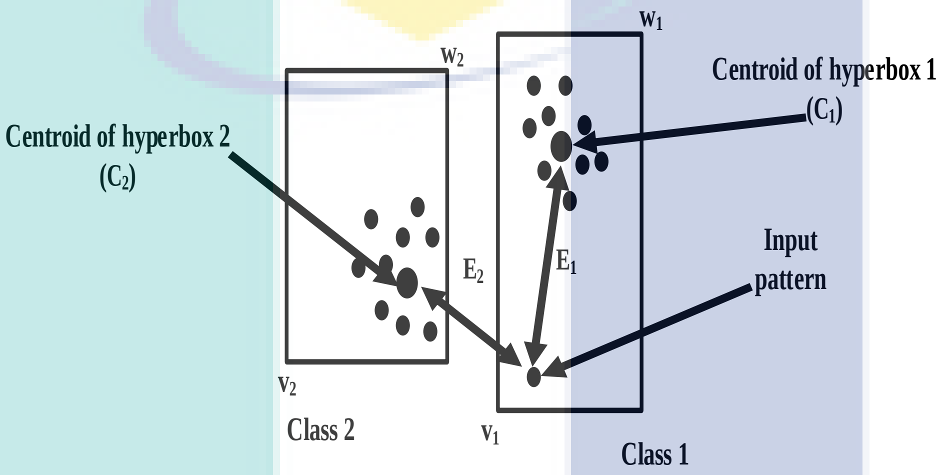


Figure 2.6 The classification process using the Euclidean distance
Source: Quteishat and Lim (2007)

An Adaptive Fuzzy Min Max Neural Network (AFMN) classifier based on the principle component analysis (PCA) and adaptive genetic algorithm (GA) was developed by J. Liu et al. to improve the classification performance of FMM (Liu et al. 2012). AFMN uses the PCA as a pre-processing step to reduce the input dimension and extract only useful information. The confidence coefficient for each hyperbox is calculated for enhancing the classification rate. The GA is used for parameter optimization, which is able to improve the speed and precision of AFMM. Although this variant can reduce the complexity of FMM neural network, the use of expansion, overlap test, and contraction process of original FMM generates misclassification results.

In 2015, Mohammed and Lim presented an Enhanced FMM (EFMM) model to overcome a number of limitations in the original FMM network (Mohammed and Lim 2015). Three heuristic rules to improve the training phase of FMM are introduced, namely the modified expansion, overlap test, and contraction procedures. Firstly, the expansion procedure is updated using a new expansion rule, which employs a dimension-by-dimension scheme rather than the summation of all dimensions. Using the new expansion rule leads to reducing the overlap areas between hyperboxes belonging to different classes. Secondly, the overlap test rules are extended to cover all overlap cases. Finally, a new contraction procedure is introduced to eliminate all overlapped cases. EFMM contracts all overlapped dimensions, rather than only the minimal overlapped dimension as in FMM. The empirical results shows EFMM efficiency as compared with FMM, GFMN, FMCN, and the Support Vector Machine (SVM) classifiers (Mohammed and Lim 2015). However, EFMM inherits the limitations related to contraction process from original FMM model, which leads to data distortion problem.

In 2016, a Modified FMM neural network (MFMMN) was introduced for pattern classification (Shinde and Kulkarni 2016). The main contribution of the MFMM network is the ability to deal with both discrete and continuous data at the same time. In MFMMN, the pruning algorithm was used in order to reduce the network complexity. However, this it still inherits a few limitations i.e., expansion, overlap test, contraction of the original FMM, which affects the classification performance.

In 2017, Mohammed and Lim proposed a new improvement on the FMM network using a K-nearest (Kn) hyperbox expansion rule named as FMM-Kn (Mohammed and Lim 2017a). The K-nearest hyperbox expansion rule is used to increase classification accuracy through reducing the FMM network complexity. The concept of selecting the winner hyperbox is modified. Instead of selecting only one winner, the K-nearest hyperboxes capable of expanding and including the input pattern without expansion rule are selected as the winners. Using different benchmark datasets, the proposed model reveals good classification accuracy with less network complexity as compared with the original FMM model. Nevertheless, it did not address the expansion process and did not cover all the overlap cases.

Mohammed and Lim further presented an extension of EFMM, known as EFMM-II (Mohammed and Lim 2017b). EFMM-II enhanced the performance of EFMM with two strategies, K-nearest hyperboxes and a pruning strategy. Firstly, the K-nearest rule is used to reduce the number of established hyperboxes in the hidden layer. Secondly, a new pruning strategy is used to reduce the noise effect by removing weak hyperboxes that affect the network efficiency. The experimental results show the efficiency of the proposed strategies in improving the EFMM performance in terms of network complexity and classification accuracy. However, EFMMII utilizes a contraction process to eliminate the overlapped regions, which can lead to an increase in the misclassification rate.

The EGFMM network was proposed with the aim to handle both labelled and unlabelled data by combining the advantages of the GFMM membership function and the EFMM equations (expansion, overlap test, and contraction) (Donglikar and Waghmare 2017). The performance of EGFMM network was evaluated using different UCI benchmark datasets and the results are better than those from GFMM neural network. However, EGFMM suffers from two aspects. Firstly, EGFMM uses the Simpson contraction process, which leads to data distortion problem. Secondly, the network complexity, where the network complexity for EGFMM is higher than that of EFMM and GFMM.

Improved data classification using fuzzy Euclidian hyperbox classifier (FEHC) was introduced by Azad et al. with aim of enhancing the classification performance of FMM neural network (Azad et al. 2018). A new way to calculate the membership value for hyperboxes based on the Euclidian distance is proposed. Thereby the process of computing the membership value of each hyperbox is computed with consideration to the centroids of the hyperboxes. However, FEHC inherits the limitations related to expansion, overlap test, and contraction from original FMM model, which can affect the classification performance.

In 2018, an optimized FMM neural network with knowledge compaction (FMM-KC) was introduced for supervised outlier detection (Upasani and Om 2018). The FMM-KC structure added a new phase named as knowledge compaction to be implemented after learning phase. The compaction phase represents the hyperbox that are purely created during the learning phase. This can help to enhance the recall time without decreasing classification performance. However, this method inherits the limitation related to expansion, overlap test, and contraction of the original FMM, which affected negatively on the classification rate.

Although significant improvements on the original FMM model have been proposed over the past years, FMM variants with contraction suffer from the data distortion problem. Data distortion happens due to lose of part of contracted hyperbox information during the contraction procedure (Bargiela et al. 2004; Kim et al. 2004). Table 2.2 shows how the limitations of original FMM are handled by different FMM variants. From Table 2.1, it is evident that FMM and its variants inherit at least one limitation in the learning stage, i.e., expansion, missing overlap rule, missing contraction rules, data distortion, susceptible to noise, and membership function.

Table 2.1 A summary of FMM variants with the contraction procedure

Model	Limitations					
	Expansion	Missing overlap rules	Missing contraction rules	Data distortion	Affected by Nosie	Membership function
FMM	Y	Y	Y	Y	Y	Y
GFMM	Y	Y	Y	Y	Y	Y
WFMM	Y	Y	Y	Y	Y	Y
MFMM	Y	Y	Y	Y	N	Y
FMM-GA	Y	Y	Y	Y	N	Y
AFMN	Y	Y	Y	Y	N	Y
EFMM	N	N	N	Y	Y	Y
MFMMN	Y	N	N	Y	Y	Y
EFMMII	N	N	N	Y	N	Y
FMM-Kn	Y	Y	Y	Y	Y	Y
EGFMM	N	N	N	Y	Y	Y
EGFMM	Y	Y	Y	Y	Y	Y
FEHC	Y	Y	Y	Y	Y	Y
FMM-KC	Y	Y	Y	Y	Y	Y

Y= Yes indicates the limitation still exists.
N= No indicates the limitations has been overcome.

2.4.2 FMM variants without contraction

Another group of FMM improvements focus on eliminating the contraction procedure from the learning stage. The aim is to improve classifier accuracy by avoiding data distortion that leads to increasing misclassification cases by removing the contraction process from learning stage, which generated during contraction procedure. The negative effect of this group is a more complex network structure due to adding more new neurons (in case of overlap) to cover omission of the contraction procedure. A review of this group of FMM variants is as follows.

In 2004, Bargiela et al. proposed an inclusion/exclusion fuzzy classifier (EFC) (Bargiela et al. 2004). EFC introduced two types of hyperboxes, inclusion and exclusion, to overcome the problem of the contraction procedure. The inclusion hyperbox represents the input patterns with the same classes, while the exclusion hyperbox represents the overlap region between hyperboxes belonging to different classes. In this model, two fuzzy sets are used for expressing the class set, rather than using one fuzzy set as in FMM. Each class is represented by taking union of inclusion of hyperboxes of the same class minus the exclusion one. The empirical results show better results than those of FMM and GFMM (Gabrys and Bargiela 2000; Simpson 1992). An issue occurs when the size of exclusion hyperbox is relatively large compared with those of the inclusion hyperboxes, leading to a high ratio of patterns that cannot classify as belonging to a specific class. In the same year, another researcher study proposed an improved model, known as adaptive exclusion/inclusion, by updating the expansion parameter to overcome the weakness of inclusion and exclusion model (Andrzej Bargiela 1 2004).

To enhance classification accuracy in overlap region, Nandedkar and Biswas proposed a new model with Compensatory Neuron, known as (FMCN) in (2007) (Nandedkar and Biswas 2007a). FMCN uses three types of neurons: 1) the classified neuron (CLN) that represent a pure hyperbox; 2) the overlap compensation neuron (OCN) that represents the overlap region; and 3) the containment compensation neuron (CCN) that represents a hyperbox inside a hyperbox. Figure 2.7 shows FMCN network structure. Two activation functions for OCN and CNN neurons are derived. The empirical results show that FMCN performs better than FMM and GFMN (Gabrys and Bargiela 2000; Simpson 1992). The primary weakness of FMCN is its complexity. Besides that, the membership function of the compensatory neuron often does not yield a correct decision.

In the same year, Nandedkar and Biswas proposed a General Reflex FMM network to handle the overlap problem (Nandedkar and Biswas 2007b). GRFMN is capable of clustering and classification in a single pass. It gives good classification accuracy as compared with that of GFMM. However, FMCN inherits the limitations related to missing overlap test rules and expansion process, thereby compromising is classification accuracy.

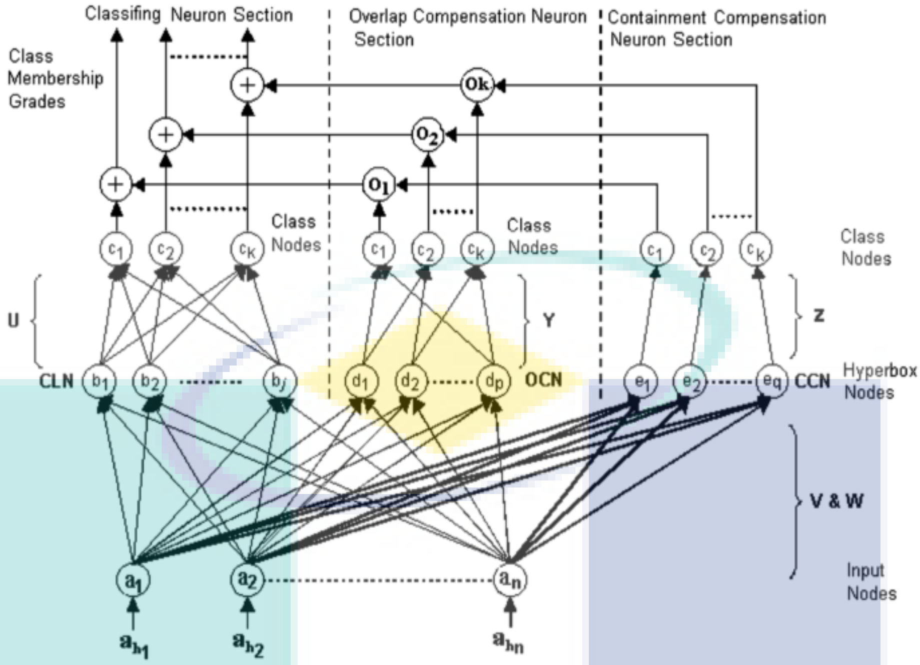


Figure 2.7 Architectuer of the FMCN

Source: Nandedkar and Biswas (2007)

Later, in 2011, Zhang et al. proposed a data core FMM (DCFMN) model for pattern classification (Zhang et al. 2011). DCFMN updates the FMM structure using two types of neurons, i.e., classifying neuron (CN) and overlapping neurons (OLN), as shown in Figure 2.8. Furthermore, a new membership function for the classifying and overlapping neurons is designed, which takes into consideration noise, geometric centre, and data core. The empirical results show that DCFMN outperforms FMM, GFMM, and FMCN with decreased computation time. However, the DCFMN has the following shortcomings: firstly, it is unable to classify all learning samples correctly and cannot correctly classify the samples located in the overlap region (Davtalab et al. 2014); secondly, the DCFMN uses the Simpson expansion rule, which leads to further overlaps between the hyperboxes that belong to different classes (Mohammed and Lim 2015).

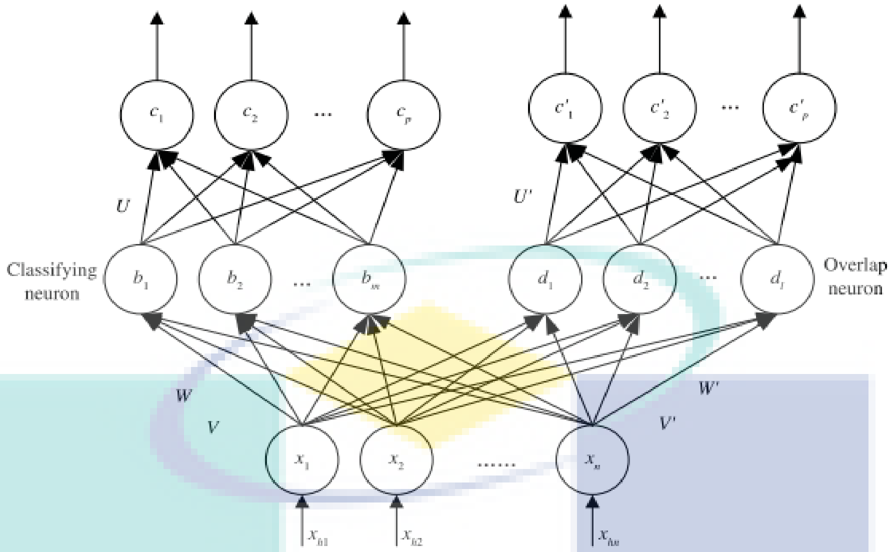


Figure 2.8 Architectuer of the DCFMN

Source: Zhang et al. (2011)

In 2014, Dvtalab et al. proposed a multilevel FMM (MLF) using two types of subnets to improve classification accuracy in the overlap regions (Davtalab et al. 2014). Each node in MLF is known as a subnet, and acts as an independent classifier, as shown in Figure 2.9(a). Each subnet has two types of hyperboxes, i.e., HBS and OLS that represent the hyperbox segment and the overlap hyperbox segment, respectively, Figure 2.9 (b) shows the structure of subnet (S_net). HBS is created during the training stage, while OLS is utilized to classify the input pattern in the overlap region. The transaction function node G determines output of the subnet relying on c_a and o_a, which represent output of HBS and OLS, respectively. For enhancing classification accuracy in the overlap region, different sizes of hyperboxes are created in different network levels. MLF shows high performance in training accuracy with low sensitivity with respect to the expansion parameter, as compared with FMM, FMCN, GFMM, and DCFMN. However, this method will create a larger number of hyperboxes than the FMM and DCFMN (Simpson 1992; Zhang et al. 2011), because the number of hyperboxes depends on the overlap region. Whenever the overlap region increases, the number of hyperboxes will be increased, and thus, the complexity of this method will also be increased.

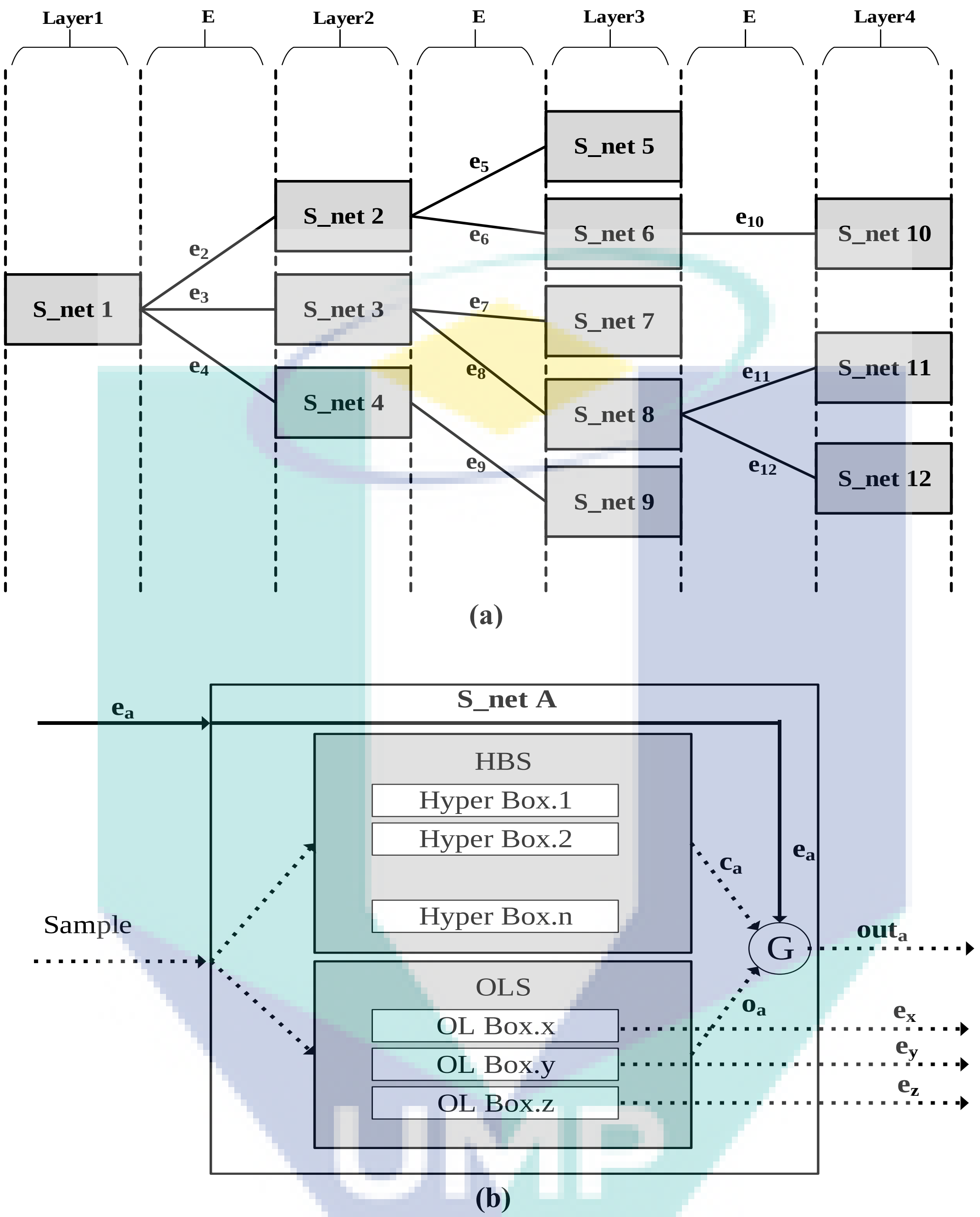


Figure 2.9 MLF network. (a) MLF structure. (b) Structure of subnet

Source : Davtalab et al (2014)

Table 2.2 summarizes the strengths and weaknesses of each proposed model. From Table 2.2, all these FMM variants still suffer from numerous limitations in the learning phase. These include a shortage of overlap rules in addition to the expansion limitation. Certain FMM variants also suffer from the complexity problem because the high number of nodes in the hidden layer.

Table 2.2 A summary of FMM variants without the contraction procedure overcome

Model	Limitations				Membership function
	Expansion	Missing overlap rule	Affected by Nosie	Complexity	
EFC	Y	Y	Y	Y	Y
Adaptive EFC	Y	Y	Y	Y	Y
GRFMM	Y	Y	Y	Y	Y
FMCN	Y	Y	Y	Y	Y
DCFMN	Y	Y	N	Y	Y
MLF	Y	Y	N	Y	Y

Y= Yes indicates the limitation still exists.
N= No indicates the limitations has been overcome.

In general, there is another classification for the development of FMM, i.e., accuracy and complexity. The first category focused on improving the accuracy through the development of the learning process i.e., expansion, overlap test, contraction, and membership function. Examples of these models GFMM, WFMM, EFMM, EGFMM, EFC, Adaptive EFC, GRFMM, FMCN, DCFMN, and MLF. However, these models are still suffering from a number of limitations described in the previous two Tables (2.2,2.3). Therefore, a new model will be introduced with useful modifications for overcoming the existing limitation and realizing better classification performance, as detailed in chapter three.

The other category focused on reducing the complexity and improving the accuracy at the same time, because the modifications conducted on the original model led to the creation of a problem called complexity (extra hyperboxes are generated). The problem of complexity means creating more hyperboxes, which negatively affects the efficiency of the network by increasing computation cost (Liu et al. 2012). The models that have reduced the complexity are: MFMM, MFMM-GA, AFMN, EFMMII, and FMM-Kn (Liu et al. 2012; Mohammed and Lim 2017a, 2017b; Quteishat and Lim 2008; Quteishat et al. 2010). All these models except the FMM-Kn worked to solve the problem of complexity by using pruning algorithm. The pruning algorithm works to delete less efficient hyperboxes after the completion of the training process based on the confidence factor. During the training process the confidence factor for each hyperbox is calculated. Hyperboxes that have a confidence factor less than user defined pruning threshold are deleted. We note that the use of pruning algorithm depends on a confidence factor identified by the user before the starts of the training process. The greater the value of the pruning threshold is the greater the number of pruned hyperboxes is; and vice versa. This means that the pruning threshold must be chosen strictly so as not affect the performance of the network. However, the use of pruning algorithm will make the network more dependent on the user by using user defined parameters. In addition, the hyperboxes that were constructed during the training process represent knowledge. Deleting any hyperbox after completion of the training process is a deletion of the knowledge. This explains why the accuracy of classification is reduced when using pruning algorithm. For this reason, the FMM-Kn has been introduced for pattern classification as a suitable solution to the problem of complexity to ensure not deleting any hyperbox that was constructed as explained in section 2.3.1. However, referring to Table 2.2 in Section 2.3.1, the FMM-Kn is still experiencing a number of limitations, especially during the learning process. Therefore, this research will reduce the complexity in chapter 4 via use of the concept of Kn used by Mohammad and Lim with some improvements that will be clarified in chapter 4.

2.5 Research Gap

Based on the literature review reported in this chapter, there have been many improvements to the FMM network over the past years. Researchers have enhanced the learning process in various stages (membership, expansion, overlap rules, and contraction) to improve accuracy and reduce complexity. In this chapter, it is noticed that several researchers have focused on the development of the membership function. This is because the FMM network does not observe the problem associated with data distortion. Researchers have therefore conducted their experiments by adding distortion to the data samples. Other researchers who studied the disadvantages of the expansion procedure argue that it leads to more overlap regions, and they develop different expansion equations. Several researchers mentioned that the overlap test rules proposed by Simpson are not able to cover all overlap cases. As such, different overlap test rules have been proposed to enhance those proposed by Simpson (Simpson 1992). Researchers also reasoned that the method of contraction proposed by Simpson distorts the hyperboxes and leads to the loss of certain key segments of a hyperbox. As a result, the contraction procedure is omitted, in order to prevent data distortion. The network structure is modified by the addition of neurons representing the overlap zone, which in turn need to be adjusted for their membership function.

In this chapter, FMM and its variants have been classified into two groups: with and without the contraction procedure. The first group of variants aims to obtain pure, no overlapping dimensions. Table 2.1 shows six characteristics of this group of FMM variants: five of which are concerned with the learning algorithm (expansion, missing overlap rules, missing contraction rules, data distortion, and membership function) and the last is concerned with the capability of the network to classify distorted data. Table 2.1 also shows that all the models still suffer from two problems, i.e., data distortion and membership function. The lack of overlap and contraction rules is another issue faced by most of these FMM variants, except EFMM, EFMM-II, MFMMN, and EGFMM, which extended the rules in both procedures, however, that extension led to increase the process complexity due to the number of overlap rules. As for the expansion procedure, most of the variants use the original one proposed by Simpson, although it leads to more overlap regions between hyperboxes belonging to different classes. For the group of FMM variants without contraction, the aim is to preserve the hyperbox structure and prevent the distortion by modifying the components pertaining to the overlapping hyperboxes. Therefore, in the case of an overlap between hyperboxes belonging to different classes, a neuron is added to the network to represent the overlap area. However, adding neurons to the network increases the degree of complexity. As for the expansion and overlap test procedures, these variants apply Simpson's equations, although the equations of overlap are known to be sub-optimal, while the equation of expansion could result in more overlap regions. Table 2.2 summarizes the limitations still existing in the group of FMM variants without contraction.

FMM variants have been introduced to overcome the limitations of FMM on various levels in the learning and test phase stages. However, all the proposed methods inherit at least three limitations from the original FMM model. Table 2.3shows three main characteristics of FMM variants, namely, expansion, overlap test and contraction, where each characteristic presents two limitations. In the expansion column, all the FMM variants, except for FMCN, suffer from the leniency hyperbox problem, whereas irregularity expansion has been handled by EFMM, EGFMM, and EFMMII through the new expansion rule. In the overlap test, all the FMM variants, except for MFMMN, still suffer from the boundary overlap problem. EFMM, EFMMII, EGFMM, and MFMMN have dealt with missing overlap rules via extending the original FMM rules. However, this extension leads to an increase in the complexity of overlap process.

Furthermore, all FMM variants with contraction process suffer from data distortion problem, leading to inaccurate decision boundaries

Table 2.3 FMM variants Limitations

Model	Expansion		Overlap Test		Contraction	
	Leniency overlap	irregularity expansion	Boundary overlap	Missing rules	Missing rules	Data distortion
FMM	H	H	H	H	H	H
GFMM	H	H	H	H	H	H
EFC	H	H	H	H	______	______
WFMM	H	H	H	H	H	H
FMCN	N	H	H	H	______	______
GRFMM	H	H	H	H	______	______
MFMM	H	H	H	H	H	H
DCFMN	H	H	H	H	______	______
MLF	H	H	H	H	______	______
EFMM	H	N	H	N	N	H
MFMMN	H	H	N	N	N	H
EFMMII	H	N	H	N	N	H
KN	H	H	H	H	H	H
EGFM	H	N	H	N	N	H
FEHC	H	H	H	H	H	H
FMM-KC	H	H	H	H	H	H

H= Have Limitation N =No Limitation ______ = process omitted

As shown in Table 2.3, all FMM variants inherit at least three limitations from the original model. Thus, this work aims to enhance the original FMM performance to overcome the limitations of existing FMM variants. This study introduces the MDFMM model by addressing limitations of expansion (leniency overlap and irregularity expansion), overlap test (boundary overlap and missing rule) and contraction process (missing rule and data distortion). Detailed analysis of these core limitations of FMM will be presented as a follow:

a. **Hyperbox Expansion**: The expansion process is used to extend the winning hyperbox to contain a new input pattern that belongs to the same class. However, applying Eq. 2.5 to activate the expansion process could distort the network structure due to the following: First, there is ***leniency*** in creating an overlap between hyperboxes from different classes whenever the constraint in (Eq. 2.5) is satisfied. In this case, the hyperbox will be expanded and will cause overlap regardless of whether the classes are different or not. Accordingly, more overlap regions will be generated, which will negatively affect the classifier performance. Second, the problem of ***irregularity*** of the hyperbox expansion leads one or more hyperbox dimensions to occupy a large area at the expense of other dimensions, thereby increasing the misclassification rate by generating unnecessary overlapped areas between classes. The expansion process calculates the total difference between minimum and maximum points of all dimensions of hyperbox expansion and compare outcome with ($n^* \theta$), as expressed in Eq. 2.5. Although some dimensions of hyperbox expansion exceed the user-defined hyperbox size (θ) during the expansion process, it will be expanded because of the constraint in (Eq. 2.5) is still satisfied. The two expansion problems, namely, the ***leniency*** of overlapping and the ***irregularity*** of the hyperbox expansion, can be explained using a 2D example, as illustrated in Figure 2.10.

Suppose there are two hyperboxes that belong to different classes: hyperbox $H_1 \in C_1$ (class 1) with minimum point $V_1 = (0.1,0.1)$ and maximum point $W_1= (0.2,0.3)$ and hyperbox $H_2 \in C_2$ (class 2) with minimum and maximum points $V_2 = (0.4,0.5)$ and $W_2 = (0.5,0.6)$, respectively. Suppose the expansion parameter is set to $\theta =0.7$ (hyperbox size) and the input sample are provided, which are denoted as $P_1 \in C_1$ (0.45,1), as shown in Figure 2.10 (a). When P_1 is provided, FMM uses Eq. 2.1 to specify the nearest hyperbox that belongs to the same P_1 class as a winning hyperbox; according to that, H_1 is selected to include P_1. To apply the expansion process, while the constraint of Eq. 2.5 is satisfied, namely, $1.25 \leq (2*0.7)$, H_1 is expanded and contains P_1. H_1 now has the same minimum point, namely, $V_1 = (0.1,0.1)$, while its maximum point is updated to $W_1= (0.45,1)$. The expanded H_1 causes the generation of an overlap region between $H_1 \in C_1$ and $H_2 \in C_2$; as a result, the overlap ***leniency*** problem is encountered, as shown in Figure 2.10 (b). Furthermore, the second dimension (Y) of H_1 occupies a large area by exceeding the size of θ (0.9>0.7), which will generate additional overlapping cases and increase the ratio in the contraction process when receiving additional input patterns.

These limitations lead to an increase in the overlap regions between the hyperboxes that belong to different classes. Consequently, this will affect the network performance negatively. Therefore, a new expansion process is needed to overcome these limitations, as discussed in Section 3.5.

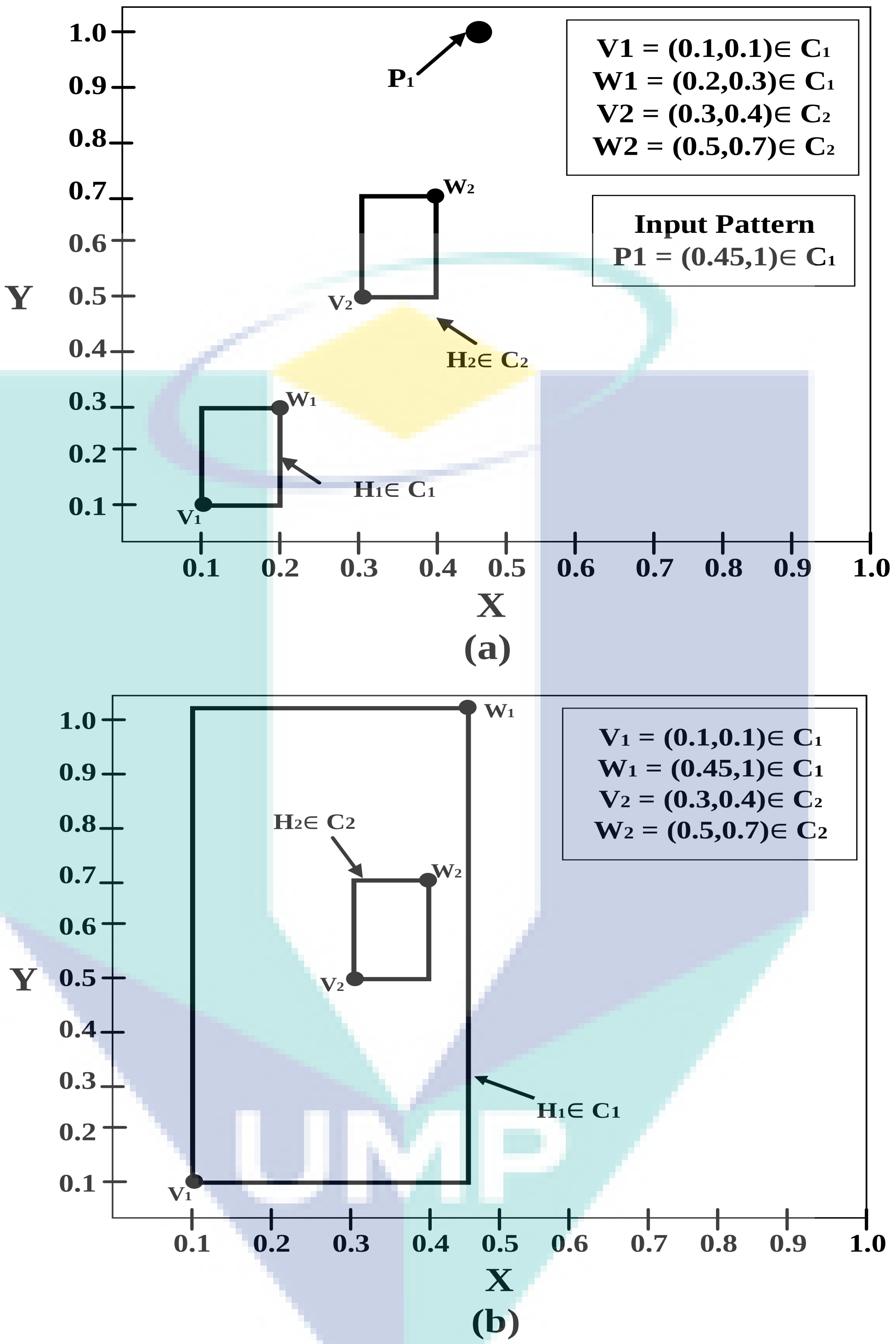


Figure 2.10 The expansion problems

b. **Hyperbox overlap test**: This step is used to examine whether the dimensions of the newly created hyperbox or expanded hyperbox overlap with those of any hyperboxes that belong to different classes. Despite the significant of this stage, it suffers from insufficient overlap rules. The FMM network utilizes four rules to identify the overlap cases; however, these rules are insufficient for covering all the overlapped cases; there are a few cases in which hyperboxes are shown to be pure even though they are overlapped. The EFMM model extended the number of overlap rules to cover all the possible overlapped cases (Mohammed and Lim 2015); however, that extension led to an increase in the process complexity due to the larger number of overlap rules. Therefore, a new overlap test formula is needed to simplify the current process that has the ability to cover all possible overlapped cases.

Furthermore, boundary points are considered non-overlapped regions in the FMM. That could keep overlapping between hyperboxes from different classes from occurring and cause the network to randomly select the winning hyperbox during the test phase. This problem can be explained using a 2D example, as shown in Figure 2.11. Assume there are two hyperboxes that belong to different classes: H_1 belongs to class (C_1) with minimum and maximum points V_1 = (0.1,0.1,0.2) and W_1= (0.4,0.4,0.5), respectively, and H_2 belongs to class (C_2) with minimum and maximum points V_2= (0.3,0.2,0.5) and W_2= (0.5,0.5,0.5). The x-axis and y-axis represent features and hyperbox min-max-point weights, respectively. As Figure 2.11(a) shows, both hyperboxes $H_1 \in C_1$ and $H_2 \in C_2$ shared the boundary region with the same weight value (0.5) at the dimension (F3), while the reset dimensions (F1 and F2) are overlapped.

FMM considers the boundary region to be a non-overlap case; therefore, no contraction process will be applied. That could affect the knowledge presentation through preventing overlap insulation and, as a result, affect the network decision-making natively during the test stage. In this case, when a test sample such as P_1= (0.35, 0.3, 0.6) is provided, as Figure 2.11(b) shows, FMM uses the membership function to identify the winning hyperbox. The provided test sample has the same membership value for both hyperboxes (0.9833); hence, a random decision is made to select the winning class between these hyperboxes, which will have a 50% possibility of generating an incorrect prediction. Therefore, a new process is needed to overcome this problem: the shared boundary area that belongs to different classes can be treated as an overlapped region, as highlights in Section 3.4.

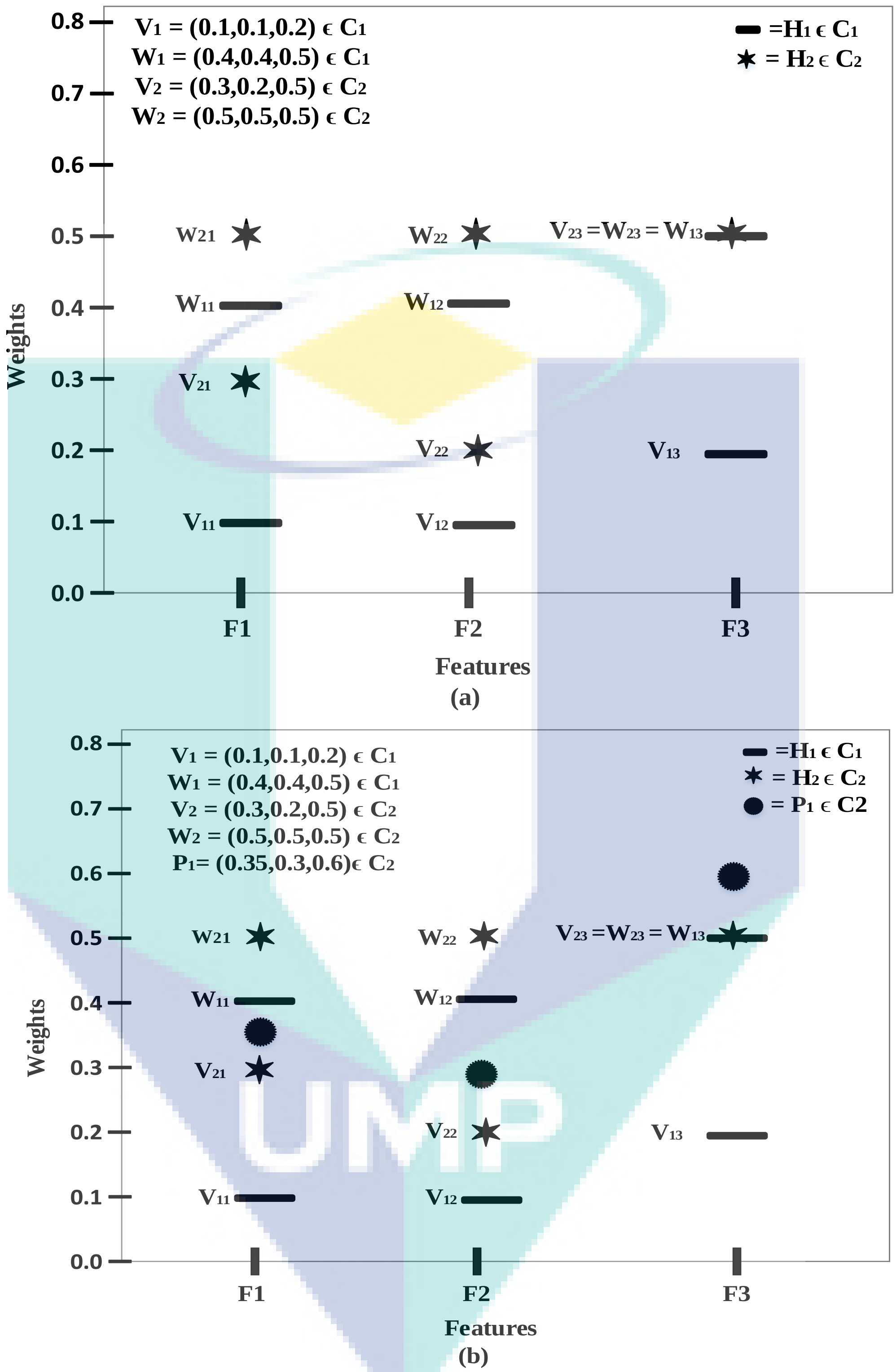


Figure 2.11 The overlap test boundary problem

c. **Hyperbox Contraction**: the contraction step is used to eliminate the overlap between the hyperboxes that belong to different classes. In FMM, this process suffers from two main drawbacks: missing contraction rules and the data distortion problem.

First, the FMM classifier uses six contraction rules, which are constructed based on four test cases, to eliminate the overlapped regions. The overlap test process has limitation in discovering all overlapped cases, and the contraction process inherited those limitations. To solve this problem, EFMM extended the overlapping test rules and the contraction rules. Although the proposed method in the EFMM model could overcome the FMM rule limitations, it could not avoid the data distortion problem. The data distortion problem refers to the lose of partial hyperbox information during the contraction process. Typically, min-max hyperbox points represent the information from the data. During the contraction process, the hyperbox minimum and/or maximum points are updated in the way that causes distortion of the hyperbox data presentation; as a result, an inaccurate decision boundary is generated, which increases the misclassification rate. The contraction problem can be described using the example that is illustrated in Figure 2.12. Assume that there are two hyperboxes that belong to different classes in 2D space: H_1 belongs to class (C_1) with minimum and maximum points V_1= (0.1,0.1,0.2) and W_1 = (0.4,0.4,0.6), respectively, and H_2 belongs to class (C_2) with minimum point V_2= (0.3,0.2,0.4) and maximum point W_2 = (0.6,0.6,0.7). The hyperbox size is θ =0.4, and the x-axis and y-axis represent features and hyperbox min-max point weights, respectively. As shown in Figure 2.12 (a) and according to Eq. 3.8, there is an overlap between H_1 and H_2, where the first dimension (F1) has a minimal overlapped area. Therefore, Eq.3.12 is applied to eliminate the overlapped region in (F1), as Figure 2.12(b) shows. This process creates a data distortion problem, whereby $V_{21} \in C_2$ becomes a full member of $H_1 \in C_1$, even though it represents an actual point of H_2. The same problem occurs with $W_{11} \in C_1$, which becomes a member of $H_2 \in C_2$ even though it represents an actual point of H_1. In addition, the shaded areas, which are denoted as Dr_1 and Dr_2 in Figure 2.12(b), represent the amounts of distorted information that belong to different classes in each hyperbox. As Figure 2.12(b) shows, both hyperboxes have the same amount of distorted area, where according to Eq. 2.12, both hyperbox points V_{21} and W_{11} are shifted to the centre of the overlapped region. Therefore, the amount of distorted area depends on the size of the overlapped region, where a larger overlapped region leads to higher distortion of hyperbox information.

Using the current contraction process led to the loss of part of the overlapped hyperbox information, which could negatively affect the network performance. Furthermore, the FMM network eliminates the overlapped points by scaling them to the hyperbox boundary region; hence, both contracted points will be full members of both hyperboxes (overlap still exists), as shown in Figure 2.12(b), where $W_{11}=V_{21} \in C_1 \in C_2$ in the first dimension (F1). Therefore, it is necessary to introduce a new contraction technique and process for solving the data distortion problem and the boundary region case, as discussed in Section 3.4.

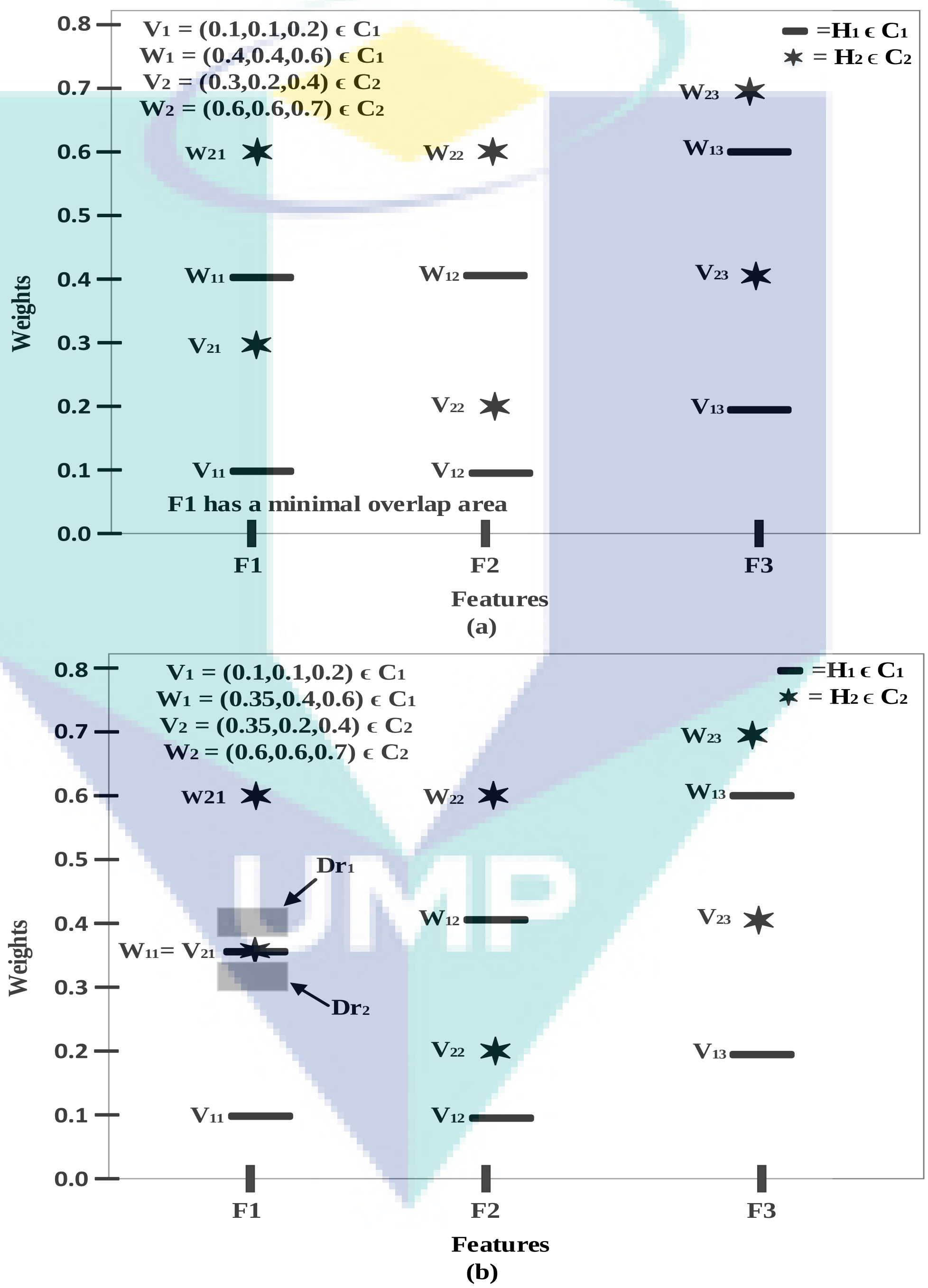


Figure 2.12 The Contraction Process problem

d. **Membership function in testing**: the FMM network uses the membership function to measure the degree to which an input sample fits within a hyperbox. FMM uses the same membership function for the training and testing stages. During the testing stage, there are a few cases in which the test sample could have the same membership degree values with two or more hyperboxes that belong to different classes. In this case, FMM randomly selects one of the winning hyperboxes from the top hyperboxes that have similar fitness values in response to the test sample. This type of prediction could negatively affect the classification performance. This problem can be explained using an example as shown in Figure 2.13. Suppose there are two hyperboxes that belong to different classes: $H_1 \in C_1$ with minimum point V_1= (0.1,0.1,0.3) and maximum point W_1= (0.3,0.2,0.55), and $H_2 \in C_2$ with V_2= (0.15,0.35,0.1) and W_2= (0.4,0.5,0.15). Suppose the hyperbox size is θ = 0.3 and the x-axis and y-axis represent dimensions (features/attributes) and hyperbox min-max point weights, respectively. When a test sample, namely, $P_1 \in C_2$ is specified as P_1= (0.2,0.25,0.2), the membership function is used to identify the fitness of the input sample P_1 relative to H_1 and H_2. According to the specified scenario, P_1 has equal fitness value (0.9750) relative to H_1 and H_2. Consequently, FMM will randomly select the winning hyperbox and there is 50% probability of selecting the incorrect class in this case.

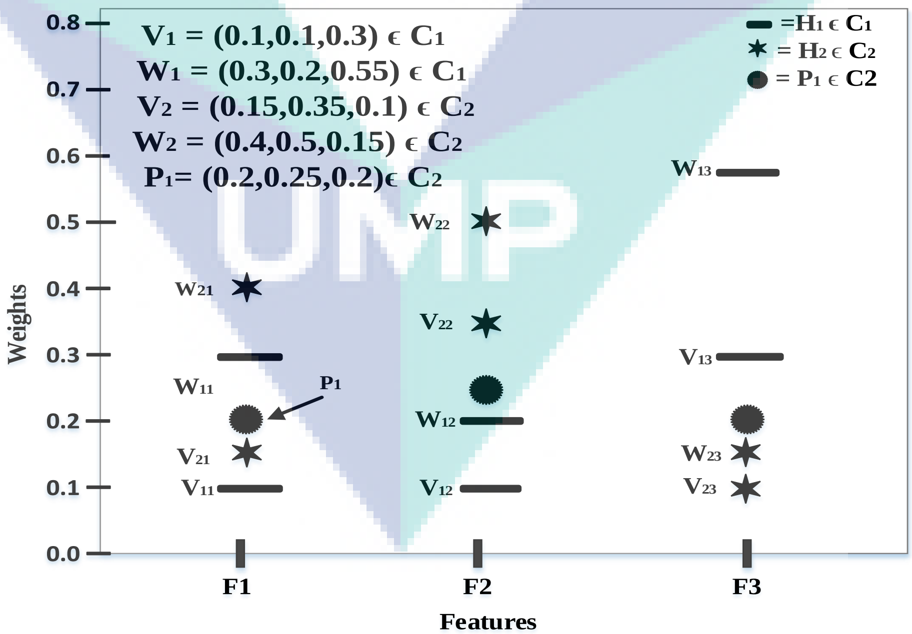


Figure 2.13 Test stage decision

Hence, using only the membership function to make the decision during the test stage will sometimes yield an inaccurate decision. Therefore, it is a necessary to avoid the randomized decision-making problem process by incorporating multiple factors instead of using only the membership function.

Therefore, this research will develop a new model that is capable of addressing these limitations efficiently and effectively named as Modern FMM (MDFMM). This will be achieved by introducing a new expansion process, and using a new formula that is covering all overlapping cases, in addition to introducing a new rule of contraction that does not distort the data in the hyperboxes, in addition that, improving the membership function decision through the testing phase.

2.6 Summary

In this chapter, the FMM neural network has been explained. A comprehensive review on some related works, as well as, recent methods that have been implemented in order to improve the FMM performance for solving pattern classification problems. where FMM and its variants have been classified into two groups, i.e., models with contraction process, and models without contraction process. Despite the possession of many robust characteristics of FMM and its variants, there are a number of existing limitations still exist in the both groups of FMM (i.e., expansion process, overlap test, contraction, process, membership function), as explained highlighted in Table 2.3 . The next Chapter 3 will demonstrate the MDFMM model, which is designed to overcome the limitations still existing in the FMM and its variants.

CHAPTER 3

MODERN FUZZY MIN MAX NEURAL NETWORK

3.1 Introduction

Based on the literature review chapter, we noticed that there is an increasing interest on the FMM neural networks because of its robust characteristics, i.e., online learning, nonlinear, overlapping classes, and training time. However, and depending on the limitations Table 2.3, Section 2.5,Chapter 2, the FMM and its variants still suffer from number of limitations that are existed in the learning process (Mohammed and Lim 2015). Therefore, in this research the Modern Fuzzy Min Max (MDFMM) is introduced to overcome the limitation of FMM and its variants. These include a new expansion technique, a new overlapping test formula, a new contraction technique, and enhancing the decision making during testing phase.

This chapter is organized as follows. The proposed modifications (MDFMM) are illustrated in Section 3.2. In Section 3.3 the performance evaluation of the MDFMM using different benchmarks dataset from UCI machine learning repository is introduced. Finally, the summary of results is given in the Section 3.4.

3.2 The MDFMM Neural Network model

The MDFMM model introduces salient solutions for overcoming limitations of the original FMM and its variants. The following subsections provide details about the MDFMM learning process (the expansion, overlap test, and contraction process) and the proposed decision-making process for the test phase.

3.2.1 Learning in MDFMM

The learning stage in the MDFMM model consists of three main steps, namely, expansion, overlap test, and contraction, which are described as follows:

3.2.1.1 Expansion Rule

Each record in training set (D) consists of a set of n ordered pairs (A_h, c_k), where A_h and c_k are the input pattern and class label, respectively. When A_h is applied, the membership function is used to determine the wining hyperbox that belongs to the same class as A_h. If A_h is not a full member of the selected hyperbox, the expansion process is applied to contain the new input sample. As described in Section 2.5 (a), the original FMM expansion process suffers from two limitations: ***overlap leniency*** and ***irregularity of the hyperbox expansion.*** These limitations negatively affect the network performance by increasing the size of the overlapped region between hyperboxes that belong to different classes. To overcome these limitations, this study propose a new expansion method that combines Eq. 3.1, which is inherited from EFMM (Mohammed and Lim 2015), and a newly proposed method, which is defined in Eq.3.2. MDFMM uses Eq.3.1 to overcome the ***irregularity of hyperbox expansion*** problem, where each dimension will be checked according to the hyperbox size (θ), instead of considering the cumulative value over all dimensions. Eq. 3.1 can be used to overcome the irregular dimension shape; however, it is unable to overcome the overlap leniency problem.

Hence, we use a new overlap test equation, namely, Eq.3.2, which can identify all overlap cases, and take into consideration the boundaries of overlapped areas. Thus, this approach overcomes the high complexity of the EFMM process, as highlighted in Section 2.5 (a), and helps generate more accurate hyperbox decision boundaries.

The MDFMM network will allow the winning hyperbox to be expanded to include the new input sample if the constraints of equations (Eq. 3.1 and Eq. 3.2) are satisfied. Otherwise, the expansion process will be terminated and a new hyperbox will be created from the input sample.

Expansion rule:

$$Max(W_{ji}, a_{hi}) - Min(V_{ji}, a_{hi}) \leq \theta \qquad 3.1$$

Overlap rule:

$$Min(\mathrm{Max}(W_{ji}, a_{hi}), W_{ki}) - Max(Min(V_{ji}, a_{hi}), V_{ki}) \geq 0 \quad 3.2$$

The new expansion process can be further clarified by revisiting the example in Section 2.5 (a) (Figure 2.10) and comparing it with Figure 3.1 which illustrates the new expansion process. As shown in Figure 3.1, assume there are two hyperboxes: $H_1 \in C_1$ with minimum points and maximum points, V_1 = (0.1,0.1) and W_1 = (0.2,0.3), respectively, and $H_2 \in C_2$ with minimum points V_1= (0.3,0.4) and maximum points W_2 = (0.5,0.7), as shown in Figure 3.1 (a). The expansion coefficient is set to θ = 0.7.

When input pattern P_1 = (0.45,1) $\in C_1$ is provided, according to Eq.3.1, H_1 is selected as a the winning hyperbox. Both Eq. 3.1 and Eq. 3.2 are utilized to evaluate the ability of H_1 to contain P_1. According to Eq. 3.1, H_1 is prohibited from being expanded due to the violation the expansion constraint (θ) by dimension Y. In contrast, according to Eq. 3.2, the expansion of H_1 will lead to overlap with H_2 which belongs to a different class. In this case, to solve the overlap problem, a new hyperbox is created, namely, $H_3 \in C_1$ with V_3=W_3= (0.45,1), as shown in Figure 3.1 (b).

When the next input pattern (P_2) is provided, Eq. (2.1) is triggered to specify the nearest hyperbox that belong to the same P_2 class as a winning hyperbox; according to that, H_1 is selected to include $P_{1,}$ as shown in Figure 3.1 (c). both Eqs. 3.1 and Eq. 3.2 are used to evaluate the ability of H_1 to include P_2. According to Eq. 3.1, $H_{2,}$ H_1 is met the expansion constraint, however a new hyperbox is created ($H_4 \in C_1$) with V_4=W_4= (0.45,0.45) according to Eq. 3.2, because the expansion of H_1 will lead to overlap with H_2 which belong to different classes, as shown in Figure 3.1 (d).

In general, This example demonstrates how the new expansion process can overcome both ***overlap leniency*** and ***irregularity of the hyperbox expansion***, thereby improving the network performance.

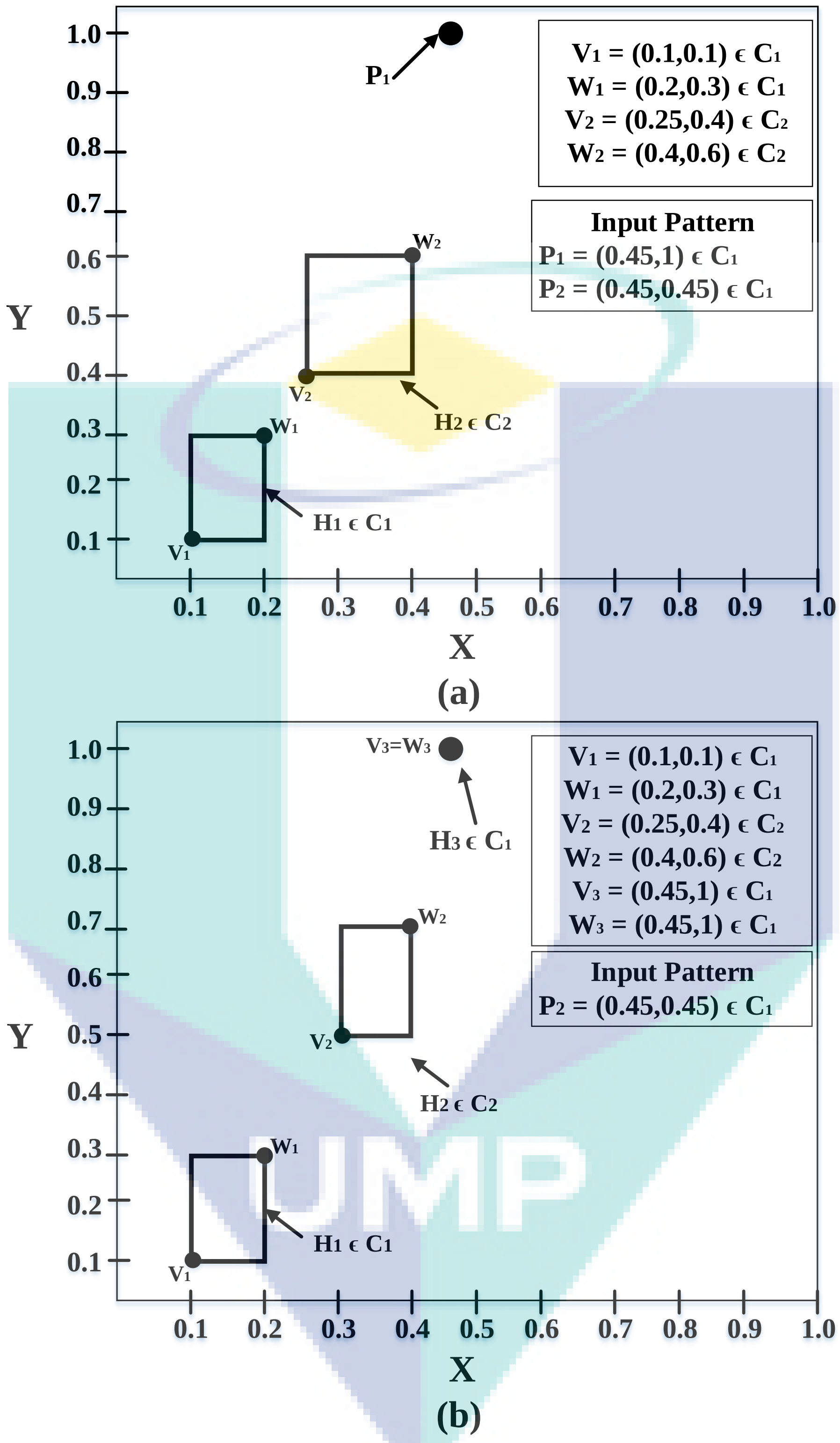
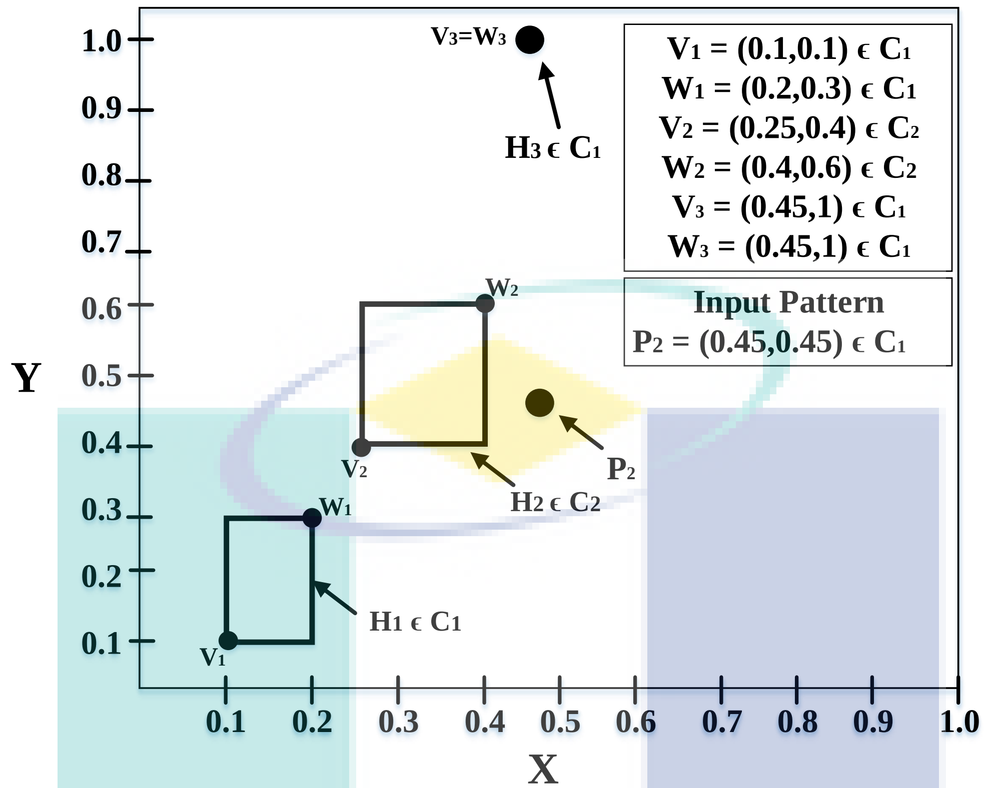
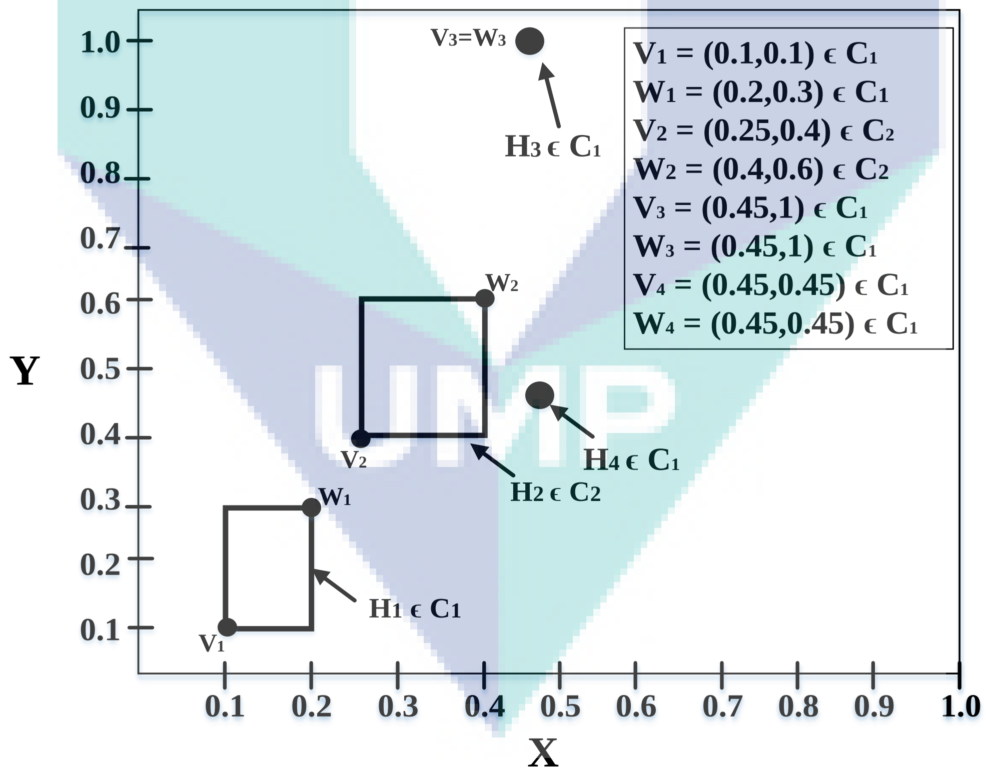


Figure 3.1 New Expansion Process



(c)



(d)

Figure 3.1 Continued

3.2.1.2 MDFMM Overlap Test Rule

The overlap test is used to determine whether there is any overlap between hyperboxes that belong to different classes. As discussed in Section 2.5 (b), the original FMM overlap test rules have two limitations: missing overlap rules and the boundary region between hyperboxes that belong to different classes. The EFMM network dealt with the missing overlap rules by adding new rules to cover all overlap cases; however, that generated a more complex process. Therefore, a new formula is proposed for simplifying the process while covering all overlap cases, which can be realized by replacing the current rules with Eq. 3.2. The overlap test process (Eq. 3.2) is triggered to identify whether the expanded or the created hyperbox overlaps with a hyperbox that belongs to a different class. The network will initiate or omit contraction process according to the overlap detection results. The second limitation, as highlighted in Section 2.5 (b), is eliminated by treating the shared boundary region that belongs to different hyperboxes from different classes as an overlapped region. The contraction process in the MDFMM model is applied to the next dimension with a minimal overlapped area if the selected dimension belongs to a boundary point. This approach improves MDFMM performance by creating a more accurate decision boundary with at least one pure dimension. Furthermore, it avoids the network using random selection to determine the winning hyperbox during the test phase, as highlighted in Section 2.5 (b). Eq.3.3 is used in MDFMM to determine the minimum overlap value for the i^{th} dimension. If the selected dimension with minimal overlap value belongs to a boundary region, where the overlap value equals =0, the contraction process is applied on the next n_{th} dimension with a minimal overlap value with that is greater than 0.

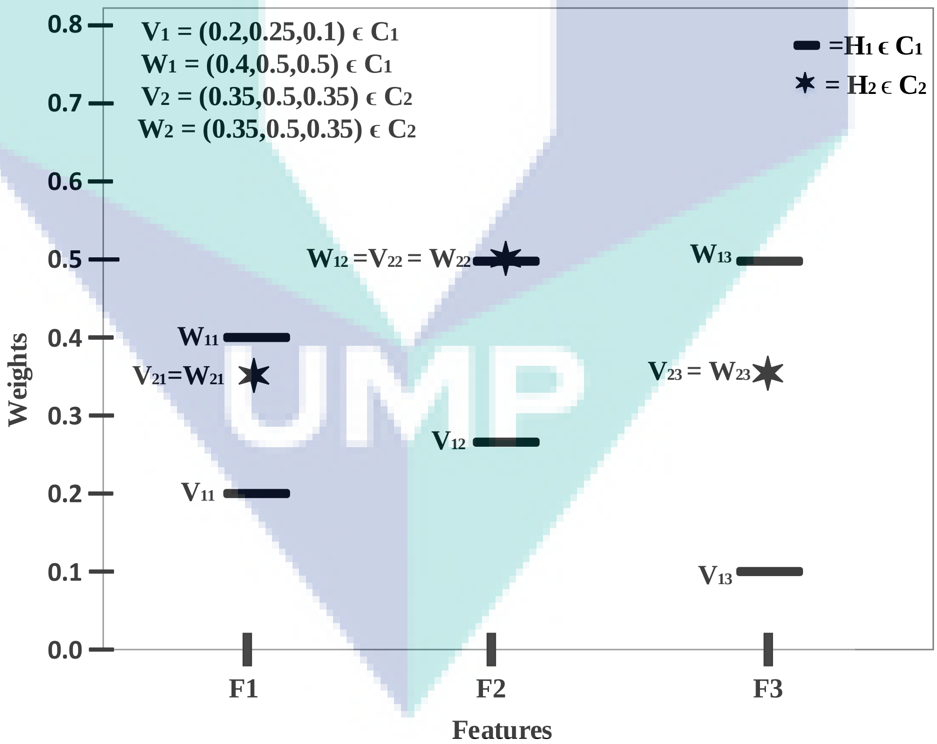
$$Overlap\ value = min((W_{ki} - V_{ji}), (W_{ji} - V_{ki})) > 0 \qquad 3.3$$

3.2.1.3 MDFMM Contraction Technique

As discussed in Section 2.5 (c), the FMM contraction process has two limitations: missing contraction rules and data distortion. The problem of missing contraction rules was overcome by the EFMM model via extension of the original FMM contraction rules; however, it requires a complex process. Furthermore, the data distortion problem remained in the previous models. Even though the MDFMM expansion process prevents any overlap case between hyperboxes that belong to different classes during the expansion process, there is a containment overlap case that occurs when input sample P falls inside a created hyperbox (H_j) that is from a different class during the learning stage cannot be avoided. In this case, a new hyperbox (H_k) will be created based on the input sample P with the same minimum and maximum points ($V_k = W_k$) inside the created hyperbox (H_j). MDFMM uses Eq.3.3 to determine the minimal overlap dimension, then, Eq.3.4 is applied to eliminate the specified overlapped area. The contraction process is applied by creating an insulated area in the selected overlap dimension. The insulated area can be created by adding new minimum and maximum points using a scale value, as expressed in Eq.3.4. The purpose of the scale value is to prevent the generation of an overlapped boundary and its value (0.001) is selected based on (Shinde and Kulkarni 2016). This process generates a more accurate boundary description for the overlapped area; nevertheless, it led to an increase number of hyperboxes by dividing H_j into two hyperboxes, as expressed in Eq.3.4. Three hyperboxes will exist: H_{j1} with V_{j1_old} and W_{j1_new}, H_{j2} with V_{j2_new} and W_{j2_old}, and H_k with the same minimum and maximum points ($V_k = W_k$). Even though the new contraction process generates more hyperboxes, however, it overcomes the data distortion problem and generates a more accurate decision boundary, which improves the classifier performance.

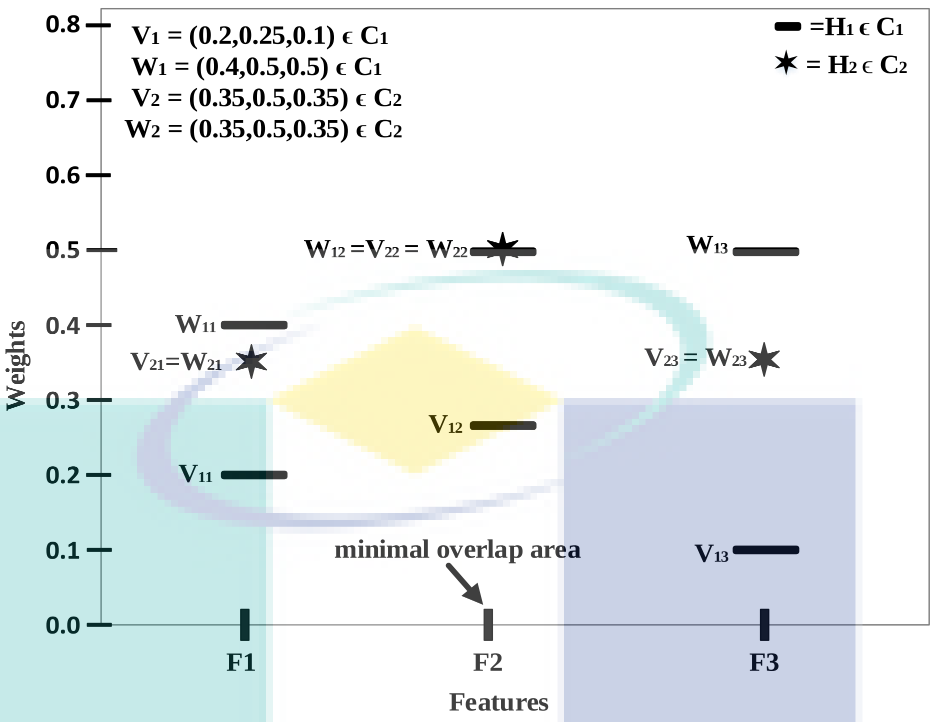
$$
\begin{aligned}
& if \quad V_{ji} < V_{ki} == W_{ki} < W_{ji} \quad then \\
& \qquad V_{j1\_old} = V_j \\
& \qquad W_{j1\_new} = V_k - scale \\
& \qquad V_{j2\_new} = W_k + scale \\
& \qquad W_{j2\_old} = W_j
\end{aligned} \tag{3.4}
$$

Where V_{j1new} and W_{j2new} represent the newly added points. The contraction process can be demonstrated using an example, as shown in Figure 3.2. Suppose there are two hyperboxes that belong to different classes in 2D space, i.e., $H_1 \epsilon C_1$ with minimum V_1= (0.2,0.25,0.1) and maximum point W_1= (0.4,0.5,0.5), and $H_2 \in C_2$ with same minimum and maximum points $V_2 = W_2$= (0.35,0.5,0.35). The hyperbox size is θ=0.4, and the x-axis and y-axis represent features and min-max point weights, respectively. All dimensions for both hyperboxes (H_1 and H_2) are overlapped, as shown in Figure 3.2 (a). Eq. 3.3 is utilized to identify the dimensions with that have minimal overlap values; the overlap values were F_1(0.05), F_2(0), and F_3(0.15). Even though the minimal overlap value is located in F_2, as shown in Figure 3.2 (b), the contraction process is applied on F_1 since F_2 is a boundary point, as shown in Figure 3.2 (c). Next, Eq.3.4 is applied to eliminate the overlap area (F_1). According to Eq.3.4, an insulated area is created in the selected minimal overlap dimension (F_1). The insulated area is created by adding new points (V_{31}, W_{31}), and the H_1 minimum and maximum points are updated via Eq.3.4, where V_{11} has the same value (0.2), while the new value for maximum point W_{11} is (0.349). According to Eq.3.4, the new minimum (V_{31}) and maximum point values (W_{31}) for H_3 are V_{31} = 0.351 and W_{31} = 0.4, where V_{31} is equal to $W_{21} - scale$, while W_{31} is equal to old value of W_{11}. As a result, three hyperboxes will exist, namely, H_1 with V_1= (0.2,0.25,0.1) and W_1= (0.349,0.5,0.5), H_2 with V_2= W_2= (0.35,0.5,0.35), and H_3 with V_3= (0.351,0.25,0.1) and W_3= (0.4,0.5,0.5), as shown in Figure 3.2 (d). Overall, this new contraction process leads to the creation of at least one pure region between H_1 and H_2. Therefore, it helps create a more accurate decision boundary, thereby improving the classifier performance. The effectiveness of MDFMM contraction process is shown in Section 3.3 experiment 2.

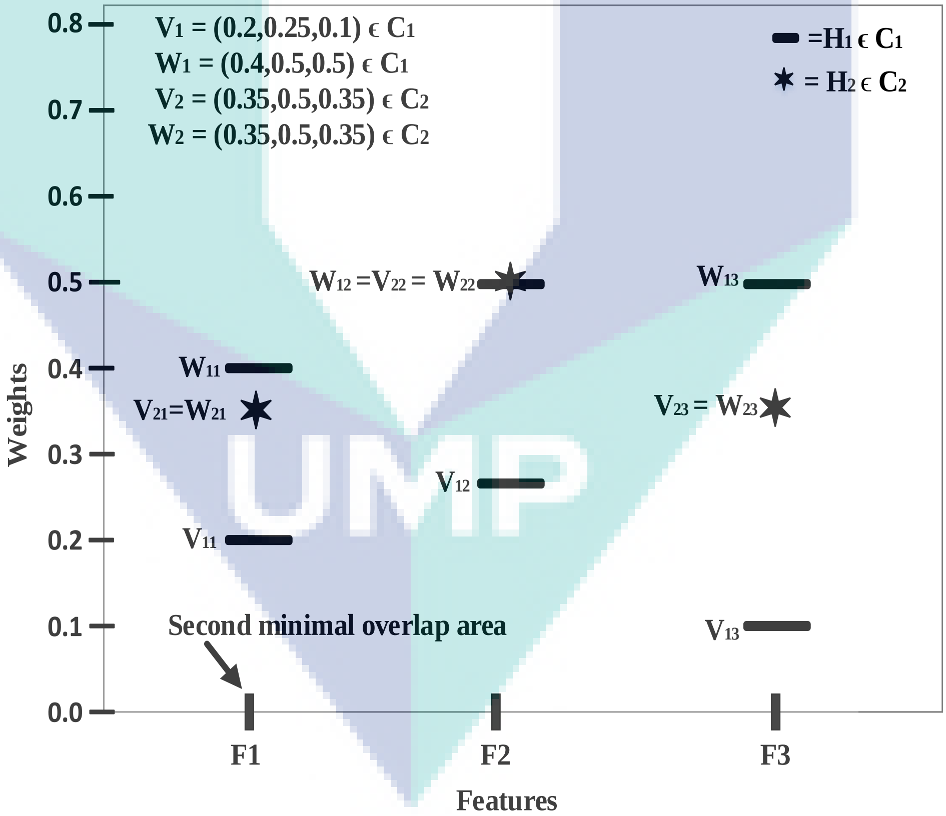


(a)

Figure 3.2 New contraction process



(b)



(c)

Figure 3.2 Continued

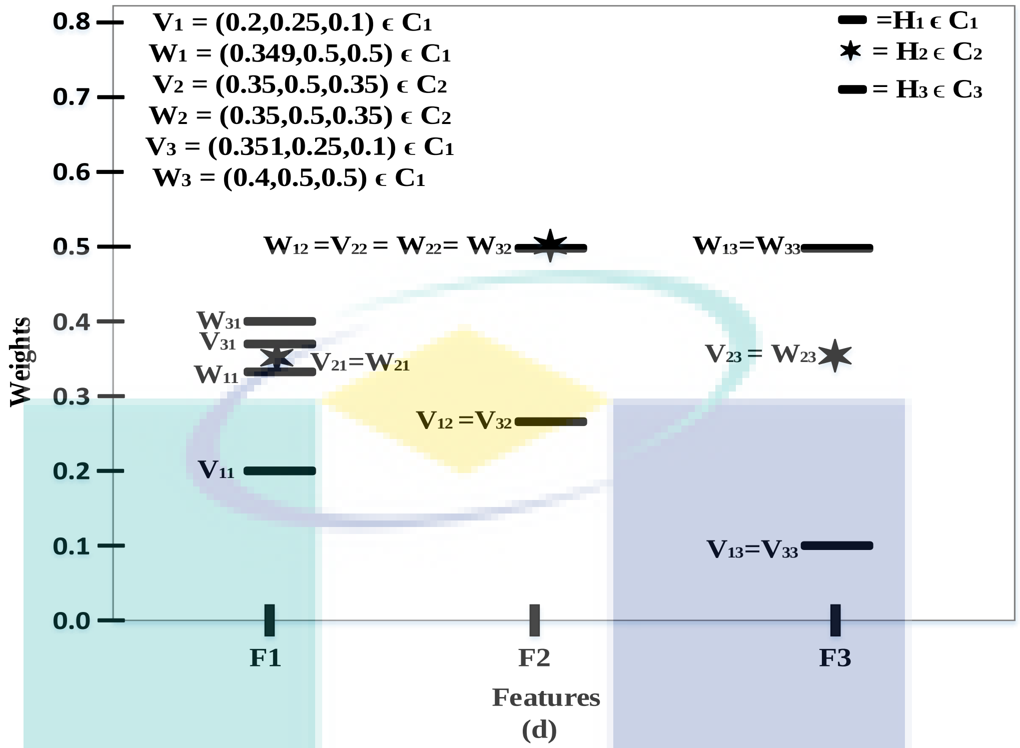


Figure 3.2 Continued

3.2.2 Testing Based on a New Decision Making Process

As discussed in Section 2.5 (d), there are a few cases in which the input pattern has the same membership degree for two or more hyperboxes that belong to different classes. In this scenario, FMM will randomly select one of the winning hyperboxes. This type of prediction could negatively affect the classification performance. To overcome this limitation, this study proposed a new process that combines a membership function (Eq.2.1) and a distance formula (Eq. 3.5). When a test sample is provided, MDFMM uses the membership function to find the fitness value for the sample according to the created hyperboxes. If there are two or more winning hyperboxes that belong to different classes and have the same fitness values, MDFMM uses the distance function (Eq. 3.5) to identify the winning class. Otherwise, the hyperbox with the highest fitness value will be selected as a winner.

$$D_j = \frac{1}{2n}\sum_{i=1}^{n}\left|a_{hi} - v_{ji}\right| + \left|a_{hi} - w_{ji}\right| \qquad 3.5$$

Where D_j is the distance for the j^{th} hyperbox, n represents the number of dimension, and a_{hi} , v_{ji},and w_{ji}, are the input sample, minimum and maximum points, respectively. The process for determining the winning class is illustrated in the following flowchart.

The proposed modification can be further clarified by revisiting the example in Section 2.5 (d). When input sample P_1 = (0.2,0.25,0.2) is provided. The membership function is utilized to measure the fitness of P_1 relative to both hyperboxes: H_1 and H_2. According to Eq. 2.1, P_1 has same fitness value (0.9750) relative to H_1 and H_2. Therefore, Eq.3 5 is utilized to calculate the distances between the input sample and the selected hyperboxes. According to Eq. 3.5, the distance between P_1 and H_1 is (0.133), while it is (0.125) between P_1 and H_2. Therefore, class H_2 is selected as the winning class. This modification helps to avoid random the decision-making, thereby reducing the misclassification error.

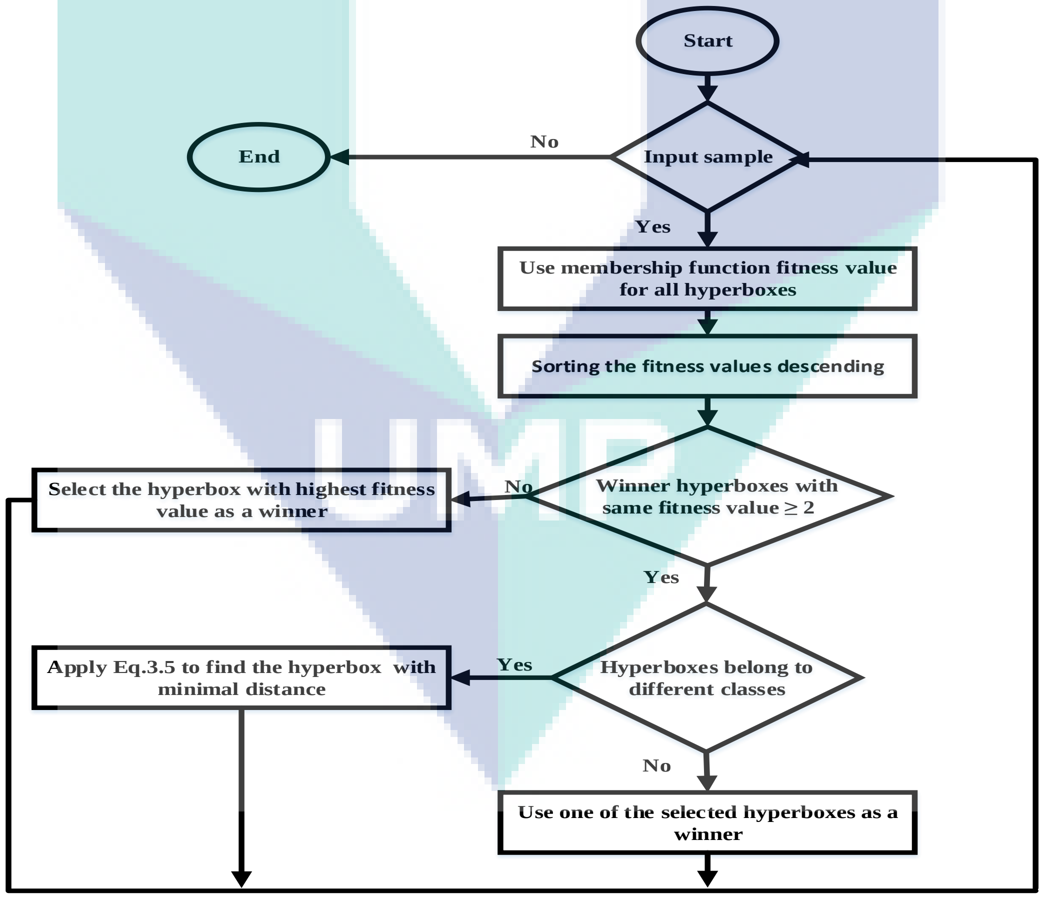


Figure 3.3 Flow chart of testing phase

3.3 Performance Evaluation

To evaluate the effectiveness of the MDFMM, four experiments were conducted. The evaluation process was based on the classification accuracy. In the evaluation performance, 11 UCI benchmark datasets have been selected from UCI Machine learning repository (Bache and Lichman 2013), to assess the performance of MDFMM. The chosen datasets cover examples of different levels of difficulties, input features, output classes, and number of instances(Mohammed and Lim 2017b). Among these data, Glass, Thyroid, WBC (origin), and Liver are most difficult problem in which this is a highly imbalanced and overlapping data. Other datasets such as Wine, Seed, and Zoo possess moderate imbalanced characteristic. Overlapping dataset such as PID, Heart (Stalog) are also chosen to cover different aspect of MDFMM performance. In addition, balanced dataset, such as Ionosphere, and less overlapping datasets, such as Iris, are also included in the experiments. Furthermore, most of the existing models used these datasets for purpose of assessing. Thus, using these datasets will assist to compare the proposed models with others to achieve a fair evaluation. Table 3.1 lists the statistical information on all UCI benchmarks datasets that are utilized in these experiments. While the description of each data set is as follows.

1. ***Iris***: Iris represents a plants data set. The data contains 150 instances, each with four features i.e., sepal length, sepal width, petal length, and petal width, from three classes. The classes refer to three kinds of Iris plants, namely, Iris Virginica, Iris Setosa, and Iris Versicolor. The dataset includes fifty instances for each of the three classes.

2. ***Wine***: Wine dataset consist of 178 instances in thirteen features distributed in three classes. This dataset represent three kind of wine gathered from the same place in Italy, in which 59 instances (33%) were from class 1, 71 instances (40%) were for class 2, and 48 instances (27%) were for class 3.

3. ***Glass***: The Glass identification dataset determines the type of glass based on its component, it consist of 214 instances, each with nine features, from six classes. They are building non-float processed 76 instances, headlamps 29 instances, building window float processed 70 instances, tableware 9 instances, and vehicle windows float processed 17 instances.

4. ***Heart (Statlog)***: The Heart (Statlog) consists of 270 samples of patients with heart issue and healthy persons, each sample has 13 features (age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise induce angina, oldpeak = ST depression induced by exercise relative to rest, the slope of the peak exercise ST segment, number of major vessels (0-3) coloured by fluoroscopy, thal: 3 = normal; 6 = fixed defect; 7 = reversable defect). It contains two classes, namely, present or absent regarding the presence and absence of heart diseases, respectively.

5. ***Liver***: The Liver disorders dataset was prepared by BUPA medical research company contains 345 samples of liver patient where and two classes, where class 1 has 145 instances, and class1 has 200 instances. While the number of features equals six, namely, mean corpuscular volume, alkaline phosphatase, alamine aminotransferase, gamma- glutamyl transpeptidase, and number of half-pint equivalents of alcoholic beverages. The first five features are blood test that were thought to be sensitive that might arise from excessive consumption alcohol.

6. ***Ionosphere***: Ionosphere dataset contains radar data gathered by a system in Goose Bay, Labrador. The goals were free electrons in the ionosphere, this dataset consist of 351 instances, each with 34 features, from two classes. Where the class 1 represents "good" radar data instances which have evidence of some kind of structure in the ionosphere, while class 2 represents "bad" instances which do not include that evidence.

7. ***WBC***: the WBC dataset is collected from Dr. Wolberg's clinical cases at university of Wisconsin, and the problem is to specify whether the tumors were benign or malignant based on data for each cancer patients. This dataset contains 699 instances, each with nine features from two classes.

8. ***PID***: PID dataset consists of 768 instances from National Institute of Diabetes and Digestive and Kidney Diseases, the aim is to specify whether the patient shows signs of diabetes based on World Health Organisation criteria. Where the number of features in this dataset is 8 and class is two. In which 268 instances (35%) of patients diagnosed with diabetes, and 500 instances (65%) diagnosed as healthy.

9. ***Zoo***: Zoo dataset uses to classify animals into seven classes, where class 1 has 41 instances, class 2 has 20 instances, class 3 has 5 instances, class 4 has 13 instances, class 5 has 4 instances, class 6 has 8 instances, and class 6 has 10 instances. It contains 101 samples, each sample has 17 features namely, hair, feathers, eggs, milk, airborne, aquatic, predator, toothed, backbone, breathes, venomous, fins, legs, tail, domestic, catsize, and type.

10. ***Seed***: Wheat Seed dataset contains 210 instances, each with seven features (Area, perimeter, compactness, length of kernel, width of kernel, asymmetry coefficient, length of kernel groove, and length of kernel groove) from three classes, namely, Rosa, Kama, and Canadian.

11. ***Thyroid***: Thyroid dataset consists of 7200 instances, each with 21 features, from three classes. In which 166 (2.3%) instances were from class hyperthyroid, 368 (5.1%) instances were from class normal, and 6666 (92.5%) instances from class hypothyroid.

Table 3.1 UCI benchmarks datasets

Number	Benchmark data	Instances	Features	Classes
1	Iris	150	4	3
2	Wine	178	13	3
3	Glass	214	9	7
4	Heart (Statlog)	270	13	2
5	Liver	345	6	2
6	Ionosphere	351	34	2
7	WBC (Origin)	699	9	2
8	PID	768	8	2
9	ZOO	101	17	7
10	Seed	210	7	3
11	Thyroid	7200	21	3

The first three experiments each consist of two sub-experiments, in which the original FMM and the proposed MDFMM model are compared on various UCI benchmark datasets. In the second sub-experiment, the proposed model and other FMM variants, i.e., GFMM, DCFMM, FMCN, MLF, MFMMN, and EGFM, were compared on various UCI benchmark datasets. In the fourth experiment, MDFMM was compared in term of performance with several other non-FMM- related classifiers: naïve bayes, C4.5, sequential minimal optimization (SMO), fuzzy gain measure, and hybrid higher neural classifier (HHONC). The aim of these experiments are as follows:

- **Experiment 1**: aims at evaluating the effect of the expansion coefficient on the performance of MDFMM. Six UCI benchmarks data set were used: Iris, Heart, WBC, PID, Seed, and Zoo. The hyperbox size ranged from (0.01-0.9);

- **Experiment 2**: aims at evaluating the effectiveness of the new contraction process and how solving the data distortion problem could affect the performance of MDFMM. In this experiment, the hyperbox size ranged from (0.1-0.9) and six UCI benchmark datasets were used: Iris, Heart, PID, Thyroid, Wine, and Liver, Ionosphere, Seed;

- **Experiment 3**: aims at comparing the performance of MDFMM to those of other FMM variants using various training size in the first sub experiment and 3k-folds cross -validation in second sub experiment. Five UCI benchmark datasets were used, i.e., Iris, Wine, Glass, and Ionosphere, Thyroid. The hyperbox size varied from 0.0 to 0.1;

- **Experiment 4**: aims at evaluating the effectiveness of MDFMM comparing to other non-related classifiers. Four UCI benchmark datasets were used: Iris, Wine, WBC, and Glass;

Furthermore, the overall percentage performance improvements of MDFMM compared to the other model are presented in Section 3.3.5. In these experiments, three statistical analysis techniques were utilized: The first was k-fold cross validation method, which was used to estimate the generalization error of the MDFMM model. In this research, the data were divided into k mutually exclusive sets, where k=3 and k=5 for 3-fold and 5-fold cross validation, respectively, and one of the *K* subsets is utilized for testing and the other subsets are combined to form a training set. Then, the average error across all *K* trials is computed, Figure 3.4 shows the two types of cross validation 5 and 3 folds. Second, the bootstrap method was used as a second statistical technique. The bootstrap method carries out re-sampling based on the observations from an experiment. Re-sampling is repeated multiple times to establish a bootstrap distribution. The parameter of interest, namely, the mean/confidence interval, is estimated based on the bootstrap distribution. In this research, 3000 resampling were used to estimate the performance indicators, such as the classification rates, standard deviations, and 95% confidence intervals. Third, the Wilcoxon signed rank test with 5% level of significant was utilized as a third statistical technique. The Wilcoxon test is a non-parametric test compares the results on a case-by-case basis (Smola and Vishwanathan 2008). These different statistical analysis methods have been used in order to obtain a fair comparison performance evaluation and comparison of the proposed model (MDFMM and MDFMM-Kn) with those compared techniques. The alternative techniques have utilised the same used statistical analysis methods as the present research in the evaluation process. Datasets with various numbers of instances, features, and classes were used for the experiments to evaluate the strength of the proposed MDFMM model

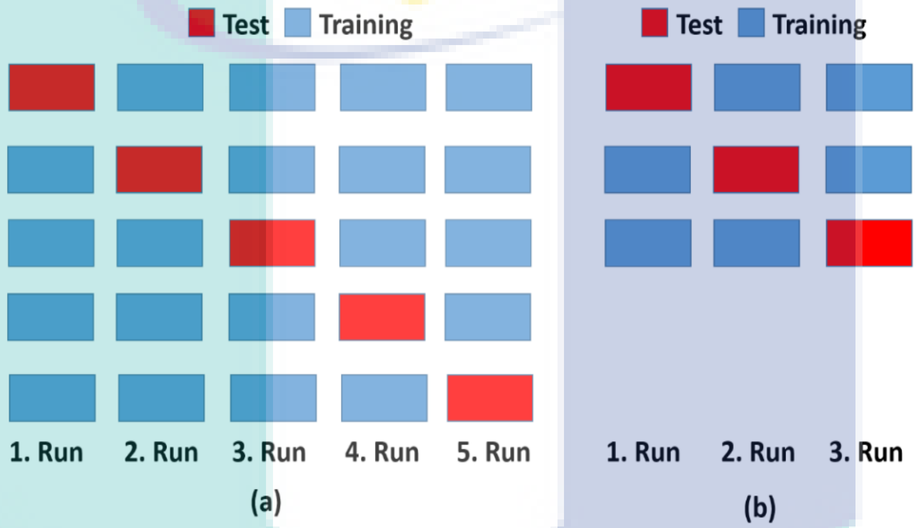


Figure 3.4 k-fold cross validation method (a) 5k-fold (b) 3k-fold

As for accuracy, in this research we will use the confusion matrix for calculation the classification accuracy and classification error (misclassification). The confusion matrix includes the numbers of correctly and incorrectly classified instances for each class. Table 3.2 shows the confusion matrix. As shown in Table 3.2, there are number of term associated with confusion matrix such as:

- True Positive (TP): the number of positive case that were correctly identified.
- False Positive (FP): the number of negative cases that were incorrectly classified as positive.
- True Negative (TN): the number of negative cases that were classified correctly.

- False Negative (FN): the number of positive cases that were incorrectly classified as negative.

The classification accuracy (AC) and the classification error/misclassification are calculated by using the following equation:

$$AC = \frac{TP+TN}{TP+TN+FP+FN} \quad 3.6$$

$$Classifcation\ error = \frac{FP+FN}{TP+TN+FP+FN} \quad 3.7$$

Table 3.2 Cofusion matrix

	CLASS	Actual Class	
		True	False
Predicted Class	True	True Positive (TP)	False Positive (FP)
	False	False Negative (FN)	True Negative (TN)

3.3.1 Experiment 1

The main objective of this experiment is to evaluate the effect of the expansion coefficient on the performance of the proposed model, namely, MDFMM, and on four other classifiers: FMM, GFMM, DCFMM, and FMCN. In the first sub-experiment, six benchmark datasets from the UCI machine learning repository were used: Heart, WBC, Seed, Iris, PID, and Zoo. Figure 3.5 shows the classification rates along with 95% confidence intervals of MDFMM and FMM. Five-fold cross-validation was utilized and the hyperbox size (θ) was varied from 0.1 to 0.9 in steps of 0.1, where each step was repeated 10 times. As shown in Figure 3.5, the MDFMM model outperformed FMM. On various datasets, the average (bootstrap) test accuracy rates of MDFMM were comparable to those FMM when the hyperbox size ranged from 0.1 to 0.4, as shown in Figure 3.5 (a,b,c,d), where the error bars indicate the 95% confidence intervals. However, MDFMM outperformed FMM statistically (at the 95% confidence level) when the hyperbox size ranged from 0.5 to 0.9. this result demonstrates the ability of the MDFMM model in overcoming the irregular dimension shape and the overlap leniency problem by incorporating the proposed modifications. In other words, MDFMM contributed to reducing the overlap between hyperboxes that belong to different classes by preventing the creation of hyperboxes of large size. MDFMM performed more stably and obtained better average accuracy values compared to the original FMM for all hyperbox size, as shown in Figure 3.5.

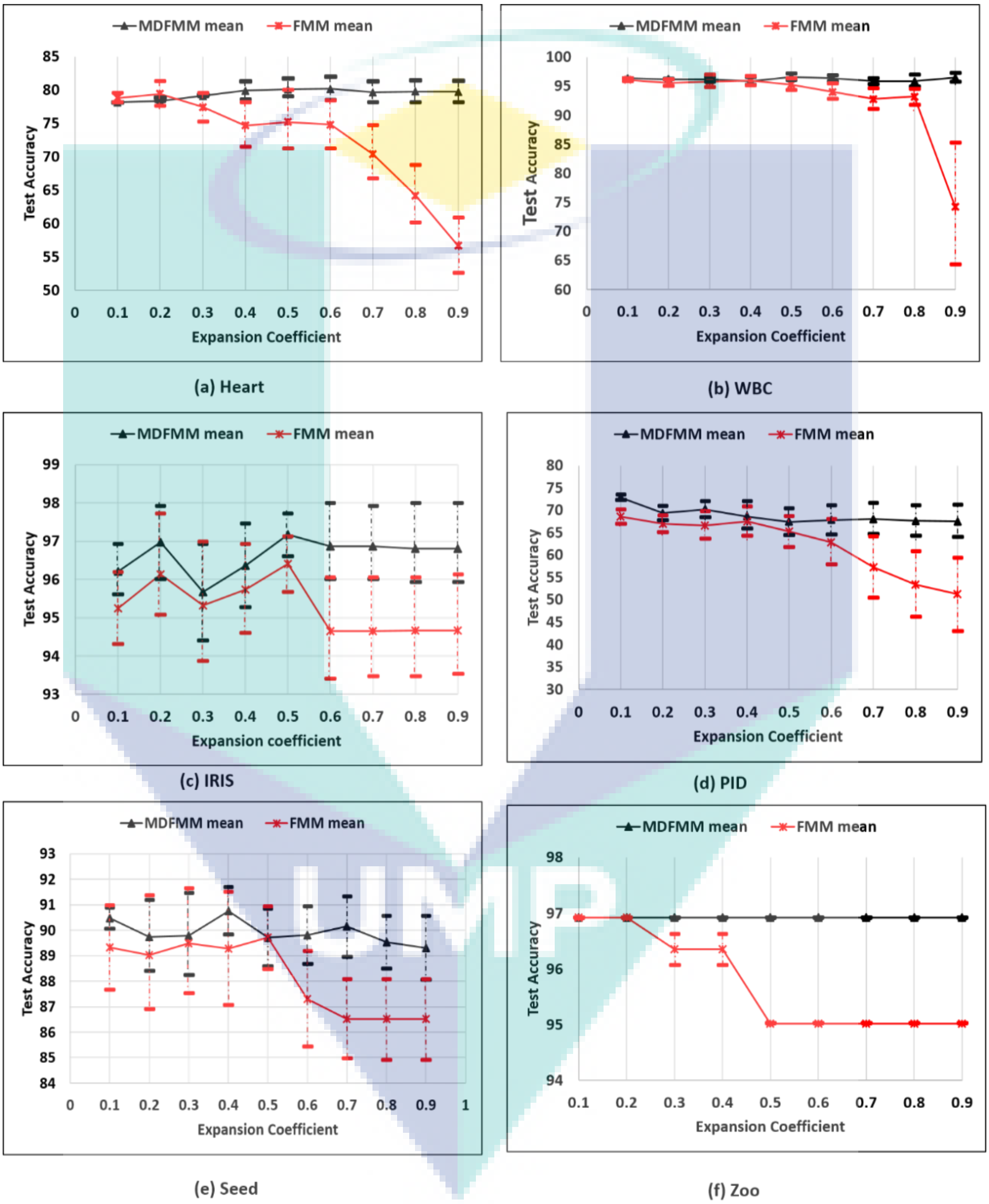


Figure 3.5 Average (bootstrap) test accuracies of MDFMM and FMM on six UCI benchmark datasets. The error bars indicate the 95% confidence intervals.

When compared between the FMM and MDFMM in term complexity, the number of hyperboxes generated by FMM is less than MDFMM, as shown in Figure 3.6. due to the modifications that MDFMM introduced in order to overcome the FMM limitations, where this modification focus on building more accurate decision boundary, for this reason the MDFMM more complex structurer than FMM network.

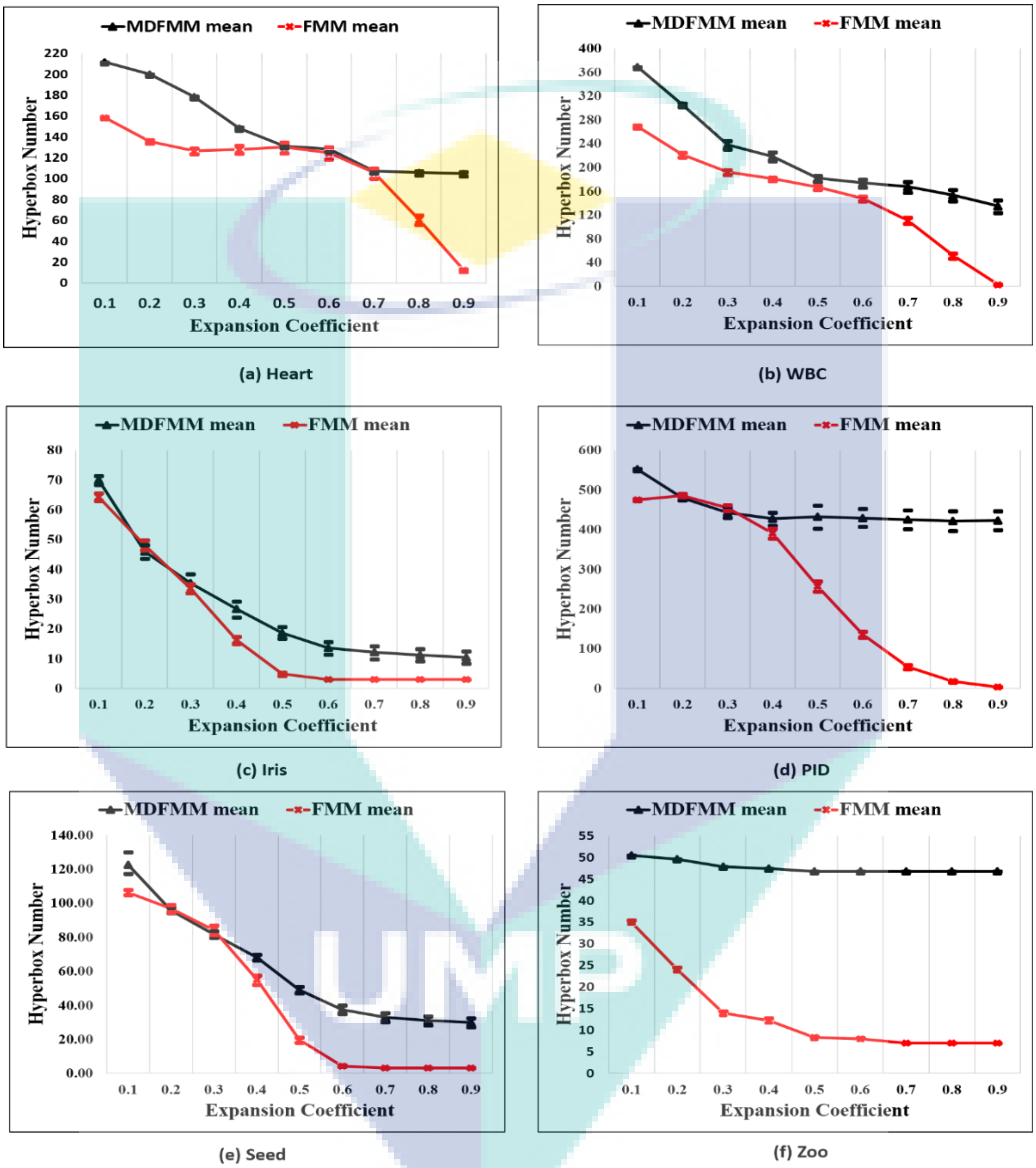


Figure 3.6 The average 5k folds cross validation number of hyperboxes of MDFMM and FMM for different data set

The Wilcoxon signed-rank test is conducted in this experiment to further assess the performances statistically; this test is a nonparametric test utilized to specify differences between two methods over each dataset (Smola and Vishwanathan 2008). The null hypothesis is rejected when the p-value is lower than the significant level ($\alpha = 0.05$

or 95% confidence interval). In other words, the performances statistically differ. In this experiment, the null hypothesis claims that MDFMM and FMM models both exhibit similar performances under testing. For comparison, Table 3.3 shows the average *p*-value in each hyperbox size (θ).

Table 3.3 The p-value of the Wilcoxon signed rank test

Θ	**FMM VS MDFMM**					
	Zoo	PID	Iris	Heart	Seed	WBC
0.1	0.3681	**0.000182**	0.065584	**0.000062**	**0.0252**	**0.001414**
0.2	0.7672	**0.003886**	0.181047	0.260676	0.226	**0.01934**
0.3	0.0289	**0.000725**	0.674569	0.014005	0.5693	0.5197
0.4	**0.000024**	0.217563	0.908830	**0.000182**	0.519	0.9698
0.5	**0.000024**	0.105122	0.907222	**0.000575**	**0.0275**	**0.0017**
0.6	**0.000024**	**0.023231**	**0.000459**	**0.000245**	0.5273	**0.00024**
0.7	**0.000024**	**0.001505**	**0.000459**	**0.000182**	**0.004**	**0.000328**
0.8	**0.000024**	**0.000130**	**0.000459**	**0.000180**	**0.0076**	**0.000725**
0.9	**0.000024**	**0.000328**	**0.000459**	**0.000181**	**0.0250**	**0.000011**

In Table 3.3, the highlighted cells are those with *p*-values lower than the significant level ($\alpha = 0.05$), rejecting the null hypothesis. In other words, their performances are statistically different (at the 95% confidence interval). The MDFMM performed better (statistically) than FMM (*i*) Zoo and heart when hyperbox size θ> 0.4; (*ii*) PID, Iris, and WBC when hyperbox size θ >0.6. The MDFMM shows significant results compared with FMM as highlighted in bold text. The results positively indicate the usefulness of the proposed model in increasing classification rates.

In the second sub-experiment, MDFMM is compared to other FMM variants in terms of performance. By following the same procedure as in (Zhang et al. 2011), 50% of the iris data set is used for training, while all the data (100%) are used for testing. The hyperbox size is varied from 0.01 to 1 in steps 0.02. Figure 3.7 shows the effects of increasing the hyperbox size on the performances of MDFMM and other FMM variants. The MDFMM network obtained the lowest misclassification ratio comparing to other networks, where the minimum error rate was obtained when the hyperbox size was set to (0-0.4), while the maximum error rate was obtained when hyperbox size was set to 0.5. Overall, MDFMM outperforms other classifiers in term of the error rate. The DCFMN model outperformed MDFMM when the hyperbox size was in the range (0.4-0.5); however, increasing the hyperbox size to Θ>0.5 led to the collapse of the learning efficiency for all classifiers except MDFMM. As the hyperbox size increases, the number of overlapping areas between hyperboxes that are from different classes increases. Hence, FMM and its variants show unstable performance; however, MDFMM overcomes this problem by controlling the expansion process and avoiding the generation of overlap during this process.

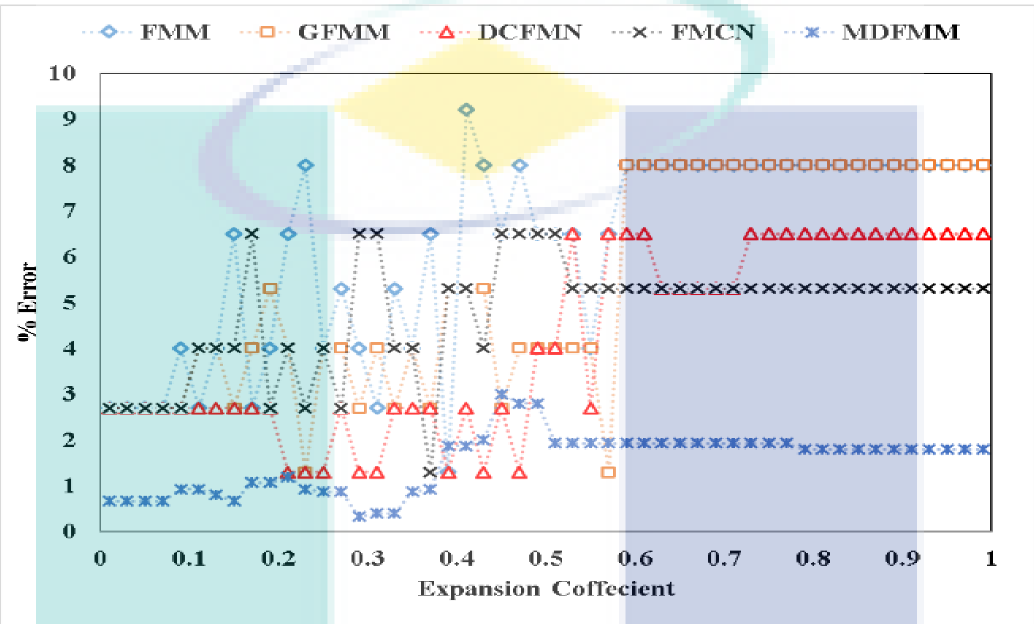


Figure 3.7 Performance on the Iris data using various hyperbox sizes

In general, it is obvious that MDFMM is able to produce better classification rate than FMM and it variants. This is proved the effectiveness of the modifications in increasing the classification rate and making the MDFMM more robustness toward increasing the hyperbox size.

3.3.2 Experiment 2

This experiment aims to ascertain the learning stability of the proposed MDFMM model and evaluate the new contraction procedure and its effect on overcoming the data distortion problem. In this experiment, the learning stability of MDFMM is analysed by evaluating whether previously trained patterns can directly access their associated class hyperboxes accurately during repeat presentation (Mohammed and Lim 2015). Two sub-experiments are conducted based on this experiment. In both sub-experiments, 100% of the data samples are used for training, and the same order of training patterns is presented again. Thus, if learning is stable in MDFMM, then this method is expected to achieve 100% accuracy.

In the first sub experiment, six benchmarks datasets were used: Iris, Heart, Thyroid, Liver, Wine, and PID. For all data sets, 100% of the data instances were used for training and the same amount used for testing. A series of systematic evaluations were performed by increasing the hyperbox size (θ) from 0.1 to 0.9 in steps of 0.1 was made; the test was repeated 10 times for each step. On all six datasets, the MDFMM model outperformed FMM. Even though the hyperbox size increased, the MDFMM model yielded an error rate 0, in contrast to FMM, which generated more misclassified cases, as shown in Figure 4.8. Increasing the hyperbox size created more overlapped cases, as a result, increased the number of contraction processes which caused more distortion of the data presentation in the overlapped hyperboxes. Unlike FMM, the performance of MDFMM was very stable with an error of 0, which is attributed to the efficiency of the new proposed methods, especially, the new contraction process in solving the data distortion problem and generating more accurate hyperbox structures.

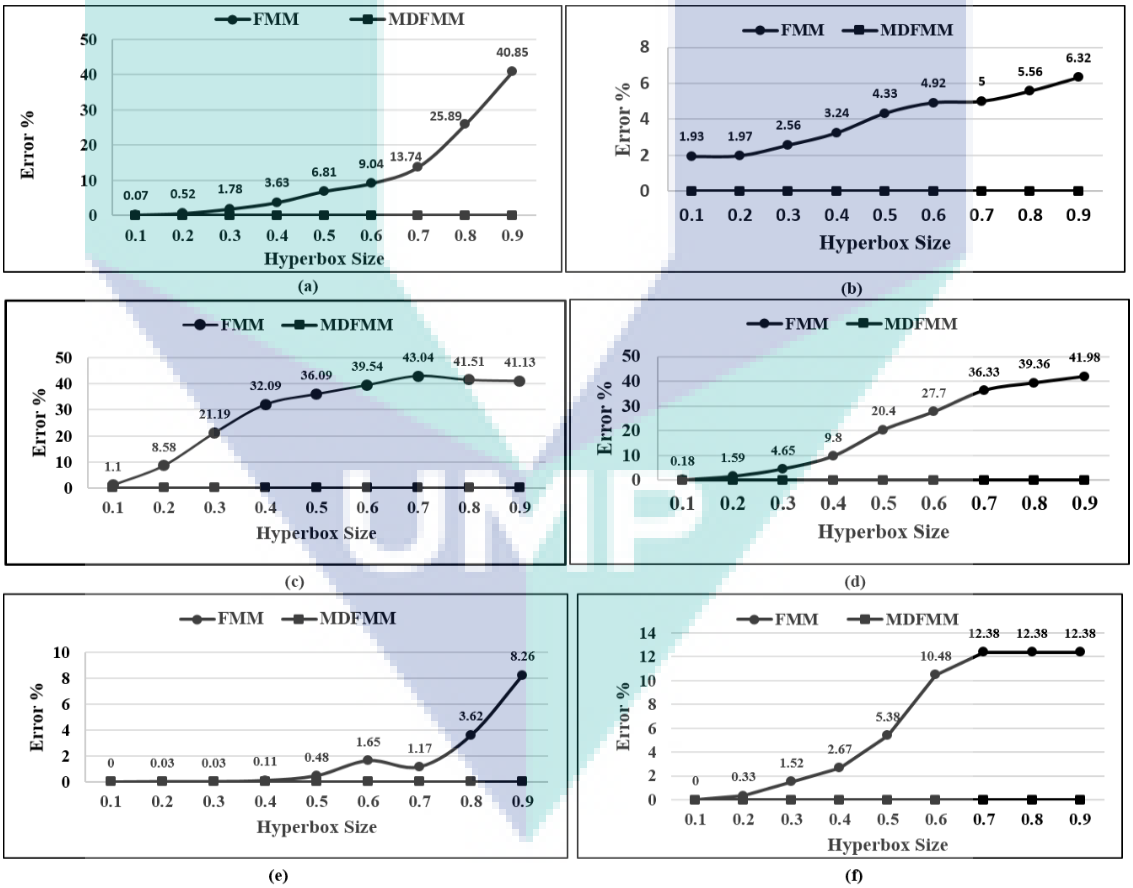


Figure 3.8 Performance comparison using different UCI benchmark datasets with different hyperbox size. (a) Heart dataset. (b) Thyroid dataset. (c) Liver datasets. (d) PID datasets. (e) Ionospher. (f) Seed.

In the second sub-experiment, the MDFMM model was compared with other FMM variants, as in (Donglikar and Waghmare 2017). For a fair comparison, the test data set in this experiment was the same as the reported in (Donglikar and Waghmare 2017). The proportions of the training and testing datasets (Wine, Iris) were both 100% and the size of the hyperbox varied from 0.1 to 0.9 with steps 0.1, the test was repeated 5 times for each step. Table 3.4 summarizes the results on the both datasets (Wine, Iris). MDFMM outperformed the other models in term of accuracy of 100% for various hyperbox sizes. The accuracy rates of other classifiers were changed substantially as the hyperbox size increased comparing to MDFMM, which yielded a stable accuracy rate.

In these models (FMM, GFMM and GEFMM), when hyperbox size increases, the accuracy rate decreases due to two main reasons: overlap test and contraction procedure. In overlap test, the FMM and GFMM suffer from missing overlap rules, resulting in the incapability of the overlap procedure to cover all cases of overlaps between hyperboxes of different classes. Accordingly, this phenomenon will negatively affect network classification with respect to accuracy rate. Concerning the contraction procedure, three models (FMM, GFMM, EGFMM, EFMM, FMM-Kn, and EFMMII) use the same contraction procedure and suffer from data distortion problem, as shown in Section 3.2.1.3.

The contraction procedure causes the removal of unambiguous areas between overlapping hyperboxes, yielding errors during the training procedure. Consequently, inaccurate decision boundary is generated, thus decreasing classification accuracy. When the size of overlap area increases, the amount of distorted area generally increases. This condition explains the increasing misclassification rate of these models when the size of hyperbox increases. By contrast, the proposed MDFMM generates 100% classification accuracy at every hyperbox size due to the new contraction procedure, leading to the creation of at least one pure region between overlapping hyperboxes. Consequently, the MDFMM succeeds in generating additional accurate decision boundaries, producing higher classification rate compared with that of alternative variants.

Table 3.4 Comparison between MDFMM and other FMM variants on two benchmark datasests. The results (percntage of classification rate) of GFMM, FMM and EGFM

Dataset	Classifier	Hyperbox size (θ) 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Avg 1
Iris	**FMM**	99.67	99.67	99.47	98.27	97.93	96.47	96.47	96.47	96.47	97.88
	GFMM	100	92	88	98	88.67	95.33	93.33	92.67	95.33	93.70
	EFMM	99.93	98.66	96.26	94.13	90.86	89.46	89.13	88.81	88.46	92.86
	EGFMM	100	98	97.33	97.33	97.33	96.67	96.67	96.67	96.67	97.41
	FMM-Kn	99.67	99.07	99	97.53	97.93	96.74	96.47	96.47	96.47	97.71
	EFMMII	99.73	98.07	96.47	93.67	90.40	88.33	87.67	89.20	89.27	92.53
	MDFMM	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Wine	**FMM**	100	100	100	100	99.94	99.49	98.48	97.47	95.84	99.02
	GFMM	100	99.66	99.44	94.38	94.94	97.19	89.33	96.07	94.94	96.22
	EFMM	100	100	99.89	99.60	98.98	95.93	95.03	94.46	93.39	97.48
	EGFMM	100	100	100	99.44	98.88	97.75	97.19	98.31	98.88	98.94
	FMM-Kn	100	100	100	100	99.77	99.10	98.31	97.85	96.89	99.10
	EFMMII	100	100	99.44	96.50	94.35	94.18	89.15	84.15	82.15	93.32
	MDFMM	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

3.3.3 Experiment 3

In this experiment, the aim is to assess the effect of the training size on the performance of MDFMM and compare the results with FMM classifiers and other variants. Two sub experiments were conducted to evaluate the performance of MDFMM on five different datasets: Iris, Wine, Glass, Thyroid, and Ionosphere. To evaluate the effect of changing the training size on the network performance, various training sizes for the Iris dataset were applied in first sub-experiment. MDFMM was compared in term of performance with four FMM variants: FMM, GFMM, DCFMN, and FMCN. To follow the same experimental procedures as in (Zhang et al. 2011), the training dataset percentage was varied from 30% to 70%, while 100% of the data set was used for testing. The expansion coefficient value (θ) was varied from 0.01 to 0.4 in increments of 0.02, and the experiment was repeated 10 times for each hyperbox size (θ). The Table 3.5 lists the standard deviation, average misclassification rate, and minimum, and maximum result values for each training size. According to Table 3.5, the MDFMM model outperformed the other classifiers: it achieved the minimum misclassification rate for all experiments. A misclassification rate of 0 was achieved by MDFMM when the training size was 60% and 70%. The error rate increased when training size was reduced because the reduction of the training size leads to the creation of a few hyperboxes unable to adequately represent the knowledge; thereby causing the misclassification rate to increase. In the standard deviation analysis, MDFMM obtained the lowest value for all cases comparing to other classifiers, which demonstrates the efficiency of the proposed techniques in supporting a stable performance of the MDFMM model.

Table 3.5 Performance comparison of MDFMM with other variants on the Iris dataset with with different training size

Method	Training Size %	Misclassification rate			
		Min	Max	Avg.	Std.
MDFMM	30	0.2	4	1.26	1.38
	40	0.22	2	0.64	0.66
	50	**0.27**	**3.9**	**1.38**	**0.92**
	60	**0**	**0**	**0**	**0**
	70	**0**	**0**	**0**	**0**
DCFMN	30	0.67	4	1.99	1.67
	40	0	2.67	1.49	1.21
	50	0	2.67	1.09	1.01
	60	0	2	0.8	0.65
	70	0	2	0.51	0.61
FMCN	30	0.67	4.67	2.43	1.95
	40	0.67	4	1.87	1.44
	50	0	2.67	1.39	1.14
	60	0	2.67	1.01	0.72
	70	0	2	0.73	0.91
GFMN	30	0.67	6	2.62	2.49
	40	0	4	2.06	2.06
	50	0	3.33	1.46	1.46
	60	0	2.67	1.15	1.15
	70	0	2.67	0.84	0.84
FMM	30	0.67	4.67	2.45	2.75
	40	0.67	4	1.87	1.86
	50	0	3.33	1.47	1.54
	60	0	2.67	1.11	1.24
	70	0	2	0.8	1.06

In the second sub-experiment, the K-fold cross validation method was used to evaluate the effectiveness of using a variety of training samples with the MDFMM model comparing to six other classifiers: GFMMN, EFC, FMCN, DCFMN, MLF, and MFMMN. By following the same experimental procedure as in (Shinde and Kulkarni 2016), 3-fold cross-validation was conducted for all datasets. The expansion coefficient value (θ) was varied from 0.0 to 0.1 in steps of 0.02. Table 3.6 lists the average misclassification rate for various coefficient values and the average number of generated hyperboxes. Comparing with other classifiers, the MDFMM model showed better accuracy performance on all datasets. The number of generated hyperboxes was another criterion that was used to assess the performance of MDFMM. The numbers of hyperboxes generated by GFMMN, EFC, FMCN, DCFMN, and MLF were smaller comparing to the MDFMM, which affected the knowledge presentation in the network structure and increased the misclassification rates for the other classifiers.

Table 3.6 Comparison between MDFMM and other FMM variants on four benchmark classification problems

Dataset	Classifier	Misclassification			Number of hyperboxes
		Max	Min	Avg.	
Wine	GFMMN	15	2.22	7.09	129
	EFC	7.78	2.22	5.53	133
	FMCN	7.78	2.22	5.53	183
	DCFMN	7.78	2.22	5.47	124
	MLF	7.78	2.22	5.39	133
	MFMMN	20	5	11.01	**34.5**
	MDFMM	**5.22**	**1.67**	**3.42**	118
Glass	GFMMN	56.52	33.33	50.21	**56**
	EFC	64.73	32.37	57.47	89
	FMCN	70.53	32.37	58.75	394
	DCFMN	70.53	33.33	60.03	77
	MLF	43.96	31.88	39.68	227
	MFMMN	42.25	28.17	36.44	97
	MDFMM	**40.8**	**29.2**	**34.2**	118.72
Ionosphere	GFMMN	57.26	10.54	31.07	120
	EFC	66.67	12.54	37.88	150
	FMCN	53.85	10.54	21.57	271
	DCFMN	20.46	6.55	12.23	132
	MLF	16.24	6.55	10.70	184
	MFMMN	25.64	**3.42**	16.24	**89**
	MDFMM	**14.5**	4.1	**10.4**	220.46
Thyroid	GFMMN	69.41	14.61	38.63	57
	EFC	25.11	9.13	14.7	66
	FMCN	25.11	9.13	14.22	116
	DCFMN	25.11	6.83	10.24	66
	MLF	24.66	5.48	9.12	72
	MFMMN	10.8	5.39	8.06	**8.3**
	MDFMM	**6.29**	**4.85**	**5.59**	2018.79

3.3.4 Experiment 4

This experiment was conducted to assess the effectiveness of MDFMM. In this experiment, MDFMM was compared with several non-FMM-related classifiers, i.e., Naïve Bayes, C4.5, SMO, Fuzzy gain measure, and HHONC, which was reported in (Fallahnezhad et al. 2011). Four benchmarks datasets were utilized, namely, Iris, Wine, WBC, and Glass, as listed in Table 3.7. In this comparison, each dataset is divided randomly as follows: 75% of the data are used for training and the remaining data are reserved for testing. This experiment followed the same procedure as that reported in (Fallahnezhad et al. 2011). Table 3.7 lists the classification accuracy and standard deviation results. The results that are presented in Table 3.7 demonstrate that MDFMM yields better results than the other classifiers. The accuracy advantages of MDFMM ranged from 0.5 to 2% on Iris, Wine, and WBC. This advantage was substantial on the Glass dataset and reached more than 10% to the closest classifier.

Table 3.7 Accuracy comparison of MDFMM and various classification on several UCI datasets

Method	Iris	WBC	Wine	Glass
Naïve Bayes	96.0±0.3	95.9±0.2	96.75±2.32	42.9±1.7
C4.5	95.13±0.2	94.71±0.09	91.14±5.12	67.9±0.5
SMO	96.69±2.58	97.51±0.97	97.87±2.11	58.85±6.58
Fuzzy gain measure	96.88±2.4	98.14±0.9	98.36±1.26	69.14±4.69
HHONC	97.46±2.31	97.17±1.17	97.88±2.29	56.5±7.58
MDFMM	98.28±1.41	98.69±0.77	98.83±1.89	78.87±0.76

3.3.5 Overall comparative performance analysis

This section aims to measure the overall classification rate improvements percentage of the MDFMM comparative to other selected models, as described in experimental results of experiment 3. To calculate the improvements percentage of each dataset, the difference in the average classification (obtained Table 3.6) between MDFMM and each model is calculated. The improvement percentage is calculated using following equation:

$$A_i = \frac{Avg.classification\ of\ R - Avg.classification\ of\ MDFMM}{Avg.classification\ of\ R} * 100 \qquad 3.8$$

Where Ai represents the improvement percentage of MDFMM relative to the Rth model for ith dataset. Next, the overall average improvement of MDFMM for each dataset is computed using the following equation:

$$Avg_k = \frac{\sum_{i=1}^{n} A}{n} \qquad 3.9$$

Where Avg_k represents the overall improvement percentage of MDFMM respected to the kth model on all datasets; A is the improvement percentage, which is computed using Eq. 3.8 for the all datasets i= (1,2,3,…,n); n indicate to the total number of datasets.

Figure 3.9 plots the MDFMM improvement percentage compared to other models from Table 3.6 in experiment 3. MDFMM achieves higher improvements percentages by generating the highest classification percentage (58.93, 53.29, 48.10, 35.22, 22.97, 35.42) for all datasets compared with six models: GFMM, EFC, FMCN, DCFMN, MLF, and MFMMN.

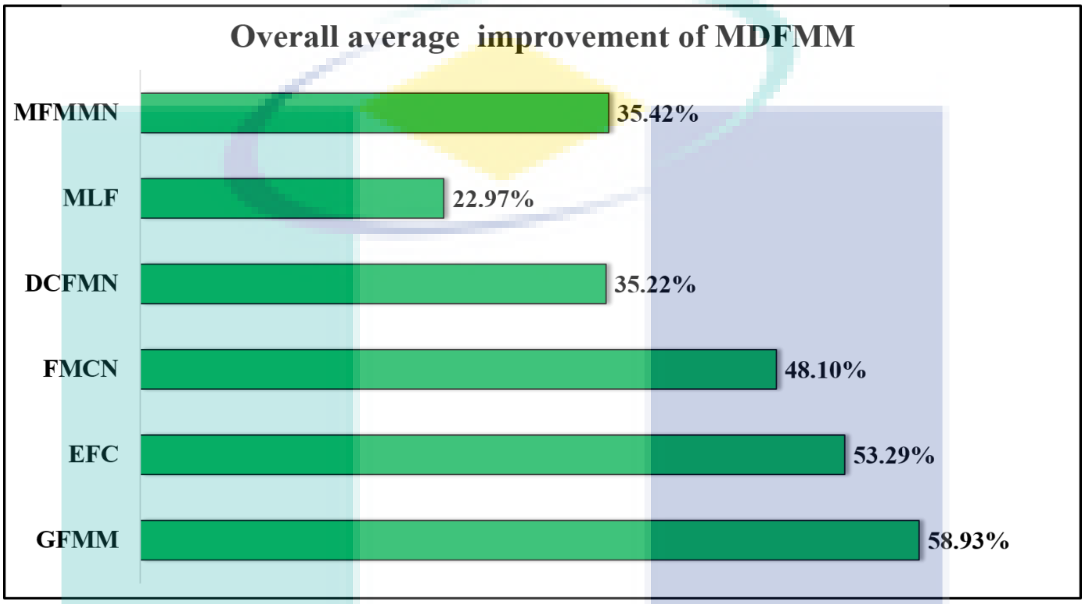


Figure 3.9 Overall improvement percentage

3.4 Discussion

This section demonstrates the main findings of this work. Based on the findings of experimental case studies, the MDFMM has evidently succeeded in addressing the core limitations in FMM with respect to expansion procedure (leniency overlap and irregularity of hyperbox expansion), overlap test (boundary overlap and missing overlap rules) and contraction procedure (missing rules and data distortion). By contrast, FMM and its variants still suffer from various limitations in the learning phase, as shown in Table 2.3 , Section 2.5, Chapter 2.

As highlighted in Section 2.5, the EFMM, EFMMII, EGFMM, EFMM-ACO and MFMMN exhibit a fewer number of limitations compared with other FMM variants because they overcome at least three out of the six limitations that remain in the original FMM. Models EFMM, EFMMII, EGFMM and EFMM-ACO overcome the irregularity expansion problem, missing overlap rules and missing contraction rule by introducing a new expansion rule and extending overlap and contraction rules. However, all these models overlook issues on data distortion, leniency overlap and boundary overlap, which can negatively affect the performance of neural networks. Although EFMM, EFMMII, EGFMM, EFMM-ACO and MFMMN handle the missing overlap and contraction rules, their extension leads to increased procedure complexity due to a large number of overlap and contraction rules. As various limitations still exist in different FMM variants, this paper introduces the MDFMM model for pattern classification.

The MDFMM is proposed to overcome all the limitations highlighted in Table 2.3; these limitations include leniency overlap, irregularity expansion, boundary overlap, missing overlap rules, data distortion and missing contraction rules. This study has introduced four new contributions. Firstly, unlike the EFMM, EFMM-ACO, EGFMM and EFMMII, the MDFMM provides a new expansion procedure to overcome the leniency and irregularity problems by preventing the use of expansion procedure if leads to create overlap with other hyperboxes of different classes (Eqs. 3.1 and 3.2). Secondly, missing overlap cases and boundary overlap have been addressed by employing a new overlap test formula (Eq 3.2 and Eq. 3.3) that replaces the overlap test procedure from the alternative FMM models (Eqs. 2.8–2.11). Thirdly, unlike EFMM, EFMM-ACO, EGFMM, EFMMII, GFMM, MFMM, FMM-Kn and MFMMN, the MDFMM introduces a new contraction procedure that can handle data distortion and missing contraction rules problem by replacing (Eqs. 2.11–2.17) with a new formula (Eq. 3.4). Lastly, a new prediction strategy that integrates the membership function (Eq. 2.1) and distance formula (Eq. 3.5) is proposed to improve decision making during the testing stage. Results of performance evaluation indicate that MDFMM performs better than the existing models and can be useful in real pattern classification problems. Furthermore, Table 3.8 depicts the new equations that have been proposed, in which each equation is discussed with respect to its usage impact.

Table 3.8 Equations usage impact.

Equations	Usage description
Eq. 3.1	Used to overcome the irregularity problem by checking the expansion coefficient dimension by dimension.
Eq. 3.2	Used to overcome the leniency and missed overlap test rules by terminating the expansion process in case it leads to overlap other hyperboxes from different classes.
Eq. 3.4	Used to avoid the data distortion by creating an insulated area in the selected overlap dimension.
Eq. 3.5	Used to overcome the random selection winning hyperbox by using distance function as final decision process for selecting the winning hyperbox.

3.5 Summary

In this chapter, we proposed a modern fuzzy min-max neural network (MDFMM) that can deliver a high classification performance while maintaining the salient features of FMM. Although the FMM neural network has several remarkable characteristics, such as single one-pass operation, online learning, overlapping classes, and hidden layer construction in a dynamic manner, it has limitations in its learning process: First, the expansion process tends to generate unnecessary overlapped regions between hyperboxes from different classes due to two problems: the leniency and the irregularity of the hyperbox expansion. Second, the existing overlap test rule suffers from two main limitations: a complex overlap test process and the randomization of decision-making when dealing with boundary regions that belong to different classes as a non-overlapped case. Third, the existing contraction process is affected by the complexity of the overlap test process and tends to generate data distortion problems due to the loss of part of the hyperbox information during the contraction process. Fourth, in some cases, the membership function leads to random decision-making. Based on that, we analysed the efficacy of FMM/EFMM models in dealing with overlapped regions that belong to different classes and we introduced MDFMM with the aim of overcoming the current limitations and improving the classification performance.

The main contributions of this chapter are as follows: First, a new expansion process was utilized that overcomes both the overlap leniency problem and the irregularity of the hyperbox expansion problem. That led to simplification of the training process and prevention of overlap cases between hyperboxes from different classes being generated during the expansion process. Therefore, the only possible overlap case that can occur in MDFMM during the learning stage belongs to containment case. Second, a new overlap test formula is proposed to simplify the previous overlap test process. It can identify all possible overlap cases. This process in necessary for terminating the expansion process if it leads to an overlap case and activating the contraction process by identifying the containment overlap case. Third, a new contraction technique is proposed that overcomes the data distortion problem and generates more accurate hyperbox decision boundaries. Fourth, a new strategy is introduced into the testing phase by integrating the distance equation with the membership function to solve the randomization of the decision-making problem. To assess the performance of the

MDFMM model and quantify the performance statistically, a series of experiments on various UCI benchmark datasets were performed using the bootstrap method and k-fold cross-validation. Via the experimental studies, the MDFMM accuracy as a function of the number of generated hyperboxes is compared with those of FMM variants and other models from the literature. Overall, the results demonstrate the superior performance of MDFMM in various aspects compared with other classifiers and the effectiveness of the proposed modifications in improving the classification performance. The MDFMM network tends to generate more hyperboxes compared to the FMM network, which is necessary for providing a more accurate description of the network structure and generate more accurate decision boundaries. Therefore, further enhancement to MDFMM is necessary, and this is the subject of the chapter 4.

CHAPTER 4

A MODERN FUZZY MIN MAX NEURAL NETWORK BASED ON K-NEAREST TECHNIQUE

4.1 Introduction

As described in Chapter 3, the FMM network is extensively analysed to clarify the existing limitations. Different problems (i.e. expansion, overlap test and contraction processes) are observed in the learning stage, and the decision-making process is highlighted in the testing stage. All these limitations negatively affect the performance of the FMM network. Therefore, the modern FMM (MDFMM) is put forward and introduced in Chapter 3. The MDFMM consists of four new contributions to address the existing limitations of the FMM network and its variants. Firstly, the MDFMM involves a new expansion process that can create a highly accurate knowledge structure by overcoming overlap leniency and the irregularity of hyperbox expansion problems. Secondly, a new overlap test formula is presented to diminish the complexity of the overlap test rules for discovering all possible overlap cases. Thirdly, a new contraction process is provided to overcome data distortion and the complexity problems of the original contraction process in the FMM network. Fourthly, a new prediction strategy is used to overcome the decision-making randomisation problem during the testing stage.

The MDFMM was evaluated in the current work using different experiments conducted on the basis of diverse UCI benchmark datasets. The results indicate the effectiveness of the proposed modifications in improving the MDFMM classification performance and providing diminished error rates, highly accurate knowledge structure and particularly stable performance against parameter changes (expansion coefficient value (θ) and training size) relative to other classifiers. However, the number of generated hyperboxes in the network structure increases as a result of the creation of pure decision boundaries (Section 3.5.1).

In the FMM network, the structure complexity term denotes the number of generated nodes (hyperboxes) in the hidden layer. The degree of computation cost increases as the number of hyperboxes increases. Hence, using few hyperboxes can decrease the computation cost (Ramos et al. 2008). In the last decade, several models (MFMM, FMM-GA, AFMN, EFMMII and FMM-K-nearest (Kn)) have been proposed to overcome the complexity problem in the FMM network (Section 2.3). Most of these models, excluding FMM-Kn, entailed the use of pruning technique.

Using a pruning technique can generally reduce the network structure whilst providing an accurate performance that is comparable to that of other non-pruned models. The dynamic process of the pruning technique focuses on removing the created hyperboxes with low knowledge structure, which identified as a hyperboxes that have confidence value below the user-defined threshold factor (Augasta and Kathirvalavakumar 2013; Mohammed and Lim 2017b). Despite the effectiveness of pruning technique in solving problem complexity, it adversely affects network structure by pruning a part of network knowledge and influences the quality of the learning process through the use of a part of a learning sample for prediction. It also increases the number of user-defined parameters with random value initialisation and makes the network barely adaptive. Mohammed and Lim proposed the FMM-Kn model to overcome the problems in the use of pruning technique (Mohammed and Lim 2017a).

The FMM-Kn model is able to reduce the network structure by modifying the original expansion process and enhancing the approach to specifying and selecting winning hyperbox during the learning stage. This approach creates a few accurate hyperboxes and thus improves the FMM network's classification accuracy. The proposed modifications can help address the complexity problem and create a highly accurate knowledge structure for the FMM classifier. However, the backbone model used in these modifications is the FMM classifier, which has various limitations (Section 3.3). In this regard, the FMM-Kn model inherits the limitations of the original FMM network. In this chapter, new modifications for the MDFMM classifier are introduced with the goal of improving its performance against complexity and noise tolerance problems. These modifications generate a Modern Fuzzy Min-Max model on the basis of the Kn technique (MDFMM-Kn) by integrating the advantages of the Kn technique and the MDFMM model.

The remainder of this chapter is organised as follows. Section 4.2 presents an analysis of the MDFMM network. Section 4.3 discusses the Kn technique and its usability for the MDFMM network. Section 4.4 explains the integration of the Kn technique with MDFMM model. Section 4.5 describes the experimental results. Section 4.6 provides a summary of the research findings.

4.2 Analysis of the MDFMM Classifier

The classification accuracy rate of the MDFMM network is better than that of the FMM network and its variants. However, it tends to generate more hyperboxes than the FMM network does (Section 3.5.1). The increase in the number of hyperboxes helps overcome the limitations of the FMM network and generate a highly accurate knowledge structure. However, the MDFMM network inherits another limitation from the original FMM, that is, the expansion rule. The expansion rule focuses on winning hyperbox and thus leads to the creation of many small hyperboxes within the vicinity. As a result, the network structure becomes increasingly complex. Although many researchers have sought to produce a less complex structure classifier and maintain an acceptable classification accuracy (Liu et al. 2012; Mohammed and Lim 2017a, 2017b; Quteishat et al. 2010; Shinde and Kulkarni 2016), further improvements must be made to minimise the complexity of the MDFMM network structure and maintain its classification performance. In this section, the complexity problem of the MDFMM network is investigated, and additional details are illustrated.

The MDFMM network is limited by the need to specify a winning hyperbox that is supposed to expand and include the new input sample. This requirement affects the expansion process and causes the creation of many unnecessary small hyperboxes during the learning stage. The MDFMM classifier generally uses the membership function shown in Eq. 3.1 to identify a set of hyperboxes that belong to the same class of input samples (Section 3.2). Subsequently, the hyperbox with a high fitness function is selected as a winner to expand and contain the new input sample if the expansion process does not violate equations. (Eq. 3.19) and (Eq. 3.20); otherwise, a new hyperbox that encodes the current input sample is created. Hence, adopting the expansion process for the winning hyperbox alone generates numerous small hyperboxes within the vicinity of the winning hyperbox. In the case of the MDFMM network, many new small hyperboxes are created whenever the winning hyperbox does not meet the expansion criteria. This condition increases network complexity. The FMM network and most of its variants suffer from the same problem, as described by (Mohammed and Lim 2017a).

The problem can be further clarified using a 2D example (Figure 4.1). Suppose that three hyperboxes belong to different classes (i.e. H_1 and H_3 belong to class (C_1), and H_2 belongs to class (C_2) with minimum and maximum points of V_1 = (0.1,0.1) and W_1 = (0.2,0.2), V_3 = (0.1,0.5) and W_3 = (0.3,0.7) and V_2 = (0.32,0.27) and W_2 = (0.5,0.4), respectively, as shown in Figure 4.1(a)). The hyperbox size is $\theta = 0.3$. When the first input sample (i.e. $P_1 = (0.25,0.35) \in C1$) is provided, the MDFMM network uses Eq. 3.1 to specify the closest hyperbox that belongs to the P_1 class as a winner. On the basis of the specification, H_3 is selected as a winner hyperbox to contain P_1 owing to its high fitness value. Then, Eq. 3.19 and Eq. 3.20 are triggered to check the ability of H_3 to include P_1. Although H_3 satisfies Eq. 3.20, it violates the expansion constraint (θ) according to Eq. 3.19. Therefore, a new hyperbox (i.e. $H_4 \in C_1$) is created with $V_4 = W_4$ = (0.25,0.35), as shown in Figure 4.1(b). When the second input sample (P_2 = (0.35,0.26) $\in C_1$) is applied, H_4 is selected to be expanded and contain P_2. However, H_4 cannot be expanded because it violates Eq. 3.20; here, the expansion of H_4 causes it to overlap with H_2, which belongs to a different class. Hence, a new hyperbox (i.e. $H_5 \in C_1$) is created with $V_5 = W_5$ = (0.35,0.26), as shown in Figure 4.1(c).

This scenario evidently shows the limitation of the original selection process and the manner by which the selection of the hyperbox with a high fitness value as the sole winner from a set of hyperboxes could increase structure complexity through the generation of unnecessary small hyperboxes. This limitation could adversely affect the performance of the MDFMM network. Therefore, a new modification is proposed in the current work to overcome the complexity problem. This modification integrates the Kn expansion technique with the MDFMM model. This approach could improve the MDFMM network structure and provide highly accurate classification performance with diminished noise tolerance effect when noise datasets are used. In the following section, the Kn concept will be discussed and analysed to explain and understand its working mechanism.

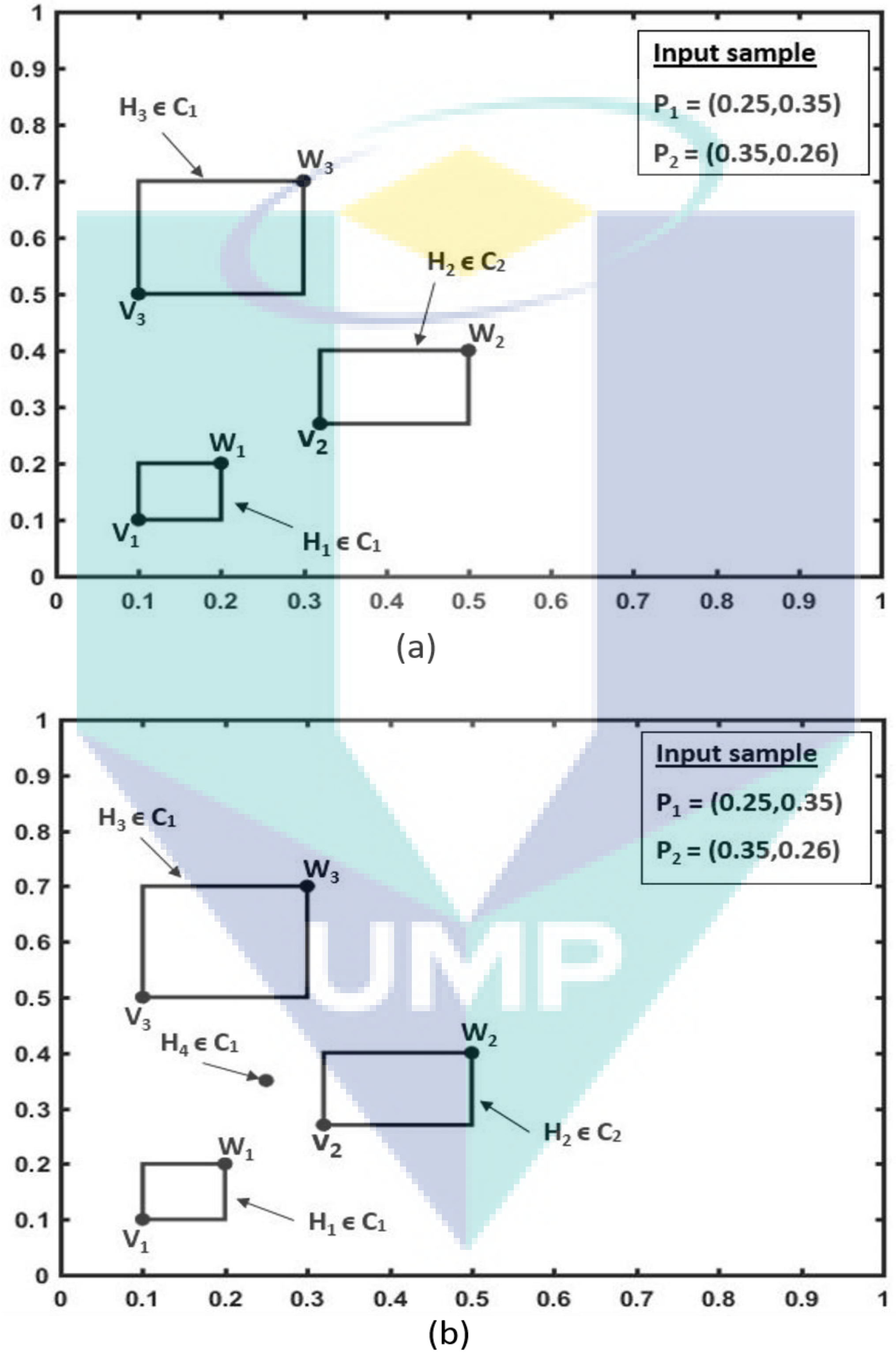


Figure 4.1 MDFMM expansion process

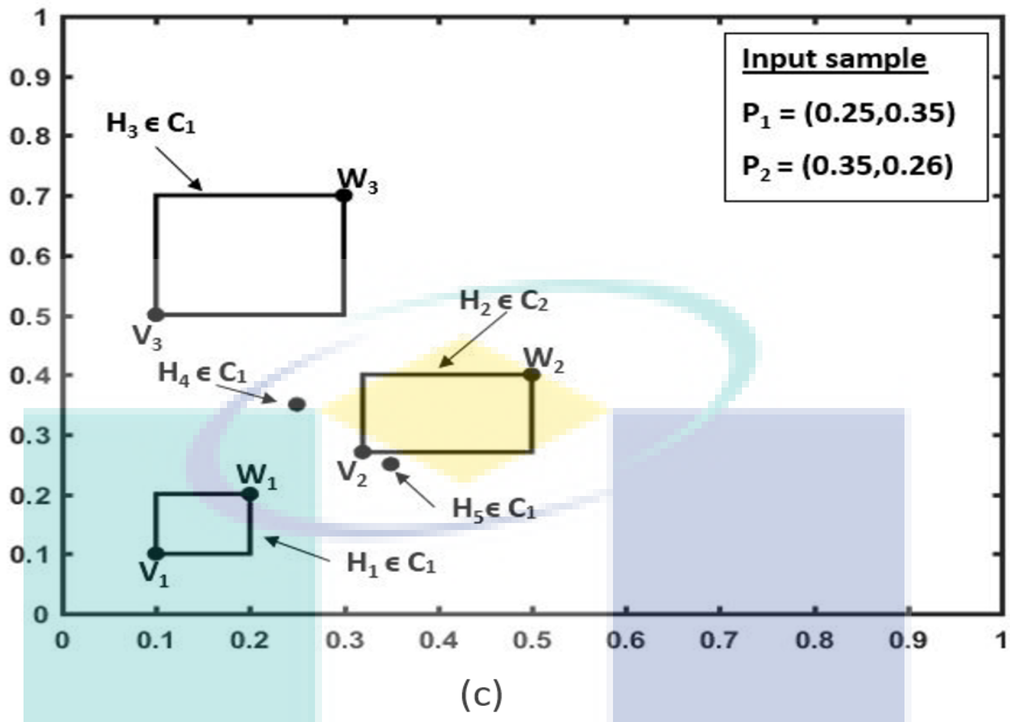


Figure 4.1 Continued

4.3 The K-Nearest Technique and Usability for the MDFMM

The Kn technique that uses the Kn hyperbox expansion rule is introduced to overcome the complexity of the FMM network (Section 4.1). To reduce the network structure, the Kn hyperbox expansion rule averts the creation of many small hyperboxes within the vicinity of the winning hyperbox (Mohammed and Lim 2017a). The FMM-Kn classifier utilises the membership function to define a set of hyperboxes that belongs to the same class of the input sample. Subsequently, the hyperbox with a high fitness function is selected as the winner. Then, Eq. 3.5 is utilised to evaluate the ability of the selected hyperbox to contain the new input sample. If the selected hyperbox does not satisfy Eq. 3.5, the expansion process is terminated, and a step-by-step verification for the set of hyperboxes is applied until the K^{th} hyperbox that meets Eq. 3.5 is found. Otherwise, a new hyperbox is created from the input sample with the same minimum and maximum values ($V_i = W_i$). Relative to the original FMM classifier, the proposed model is evidently able to generate a less complex network structure and maintain a satisfactory accuracy rate (Mohammed and Lim 2017a).

The Kn expansion rule can be clarified by using a 2D example, as shown in Figure 4.2. Suppose that three hyperboxes (i.e. H_1, H_2 and H_3) belong to different classes, where H_1 and $H_3 \in C_1$ and $H_2 \in C_2$ with minimum and maximum points of V_1 = (0.1,0.1) and W_1 = (0.2,0.2), V_3 = (0.1,0.5) and W_3 = (0.3,0.7) and V_2 = (0.32,0.27) and W_2= (0.5,0.4), respectively. Assume that two input samples (i.e. P_1 = (0.25,0.35) and P_2 = (0.35,0.26) $\in C_1$) exist, as shown in Figure 4.2(a). When the input sample P_1 is applied, according to Eq. (3.1) (Section 3.2), H_3 is selected as the winning hyperbox, as shown in Figure 4.2(b). The expansion coefficient (Eq. 3.5) is used to evaluate the ability of H_3 to include P_1. According to Eq. 3.5, H_3 cannot be expanded because the expansion constraint (θ) is violated. Thus, the next nearest hyperbox is selected as the winning one (H_1). Meanwhile, the constraint of Eq. 3.5 is satisfied. Hence, H_1 is expanded to contain P_1, as shown in Figure 4.2(c). Moreover, Eq. 3.6 and Eq. 3.7 are applied to update the minimum and maximum points for H_1, where V_1 = (0.1,0.1) and W_1 = (0.35,0.35).

When the next input sample (P_2) is provided, Eq. (3.1) is triggered because the nearest hyperbox that belongs to the same $P_2 \in C_1$ class is selected as the winning one. On the basis of this outcome, H_1 is selected to include P_2, as shown in Figure 4.2(d). H_1 is expanded and contains P_1 considering that expansion coefficient Eq. 3.5 is met through the expansion process. Accordingly, H_1 now exhibits the same minimum point, that is, V_1 = (0.1,0.1), whilst its maximum point is modified to W_1 = (0.35,0.26), as shown in Figure 4.2(e). In this example, the Kn technique is able to improve the general FMM performance by decreasing the number of unnecessary hyperboxes generated and thereby limiting the number of overlapped cases between hyperboxes from different classes. With these results, the overlap and contraction processes are shortened, and the effect of data distortion on knowledge structure is minimised. This technique can diminish noise tolerance capability, as explained by (Mohammed and Lim 2017b).

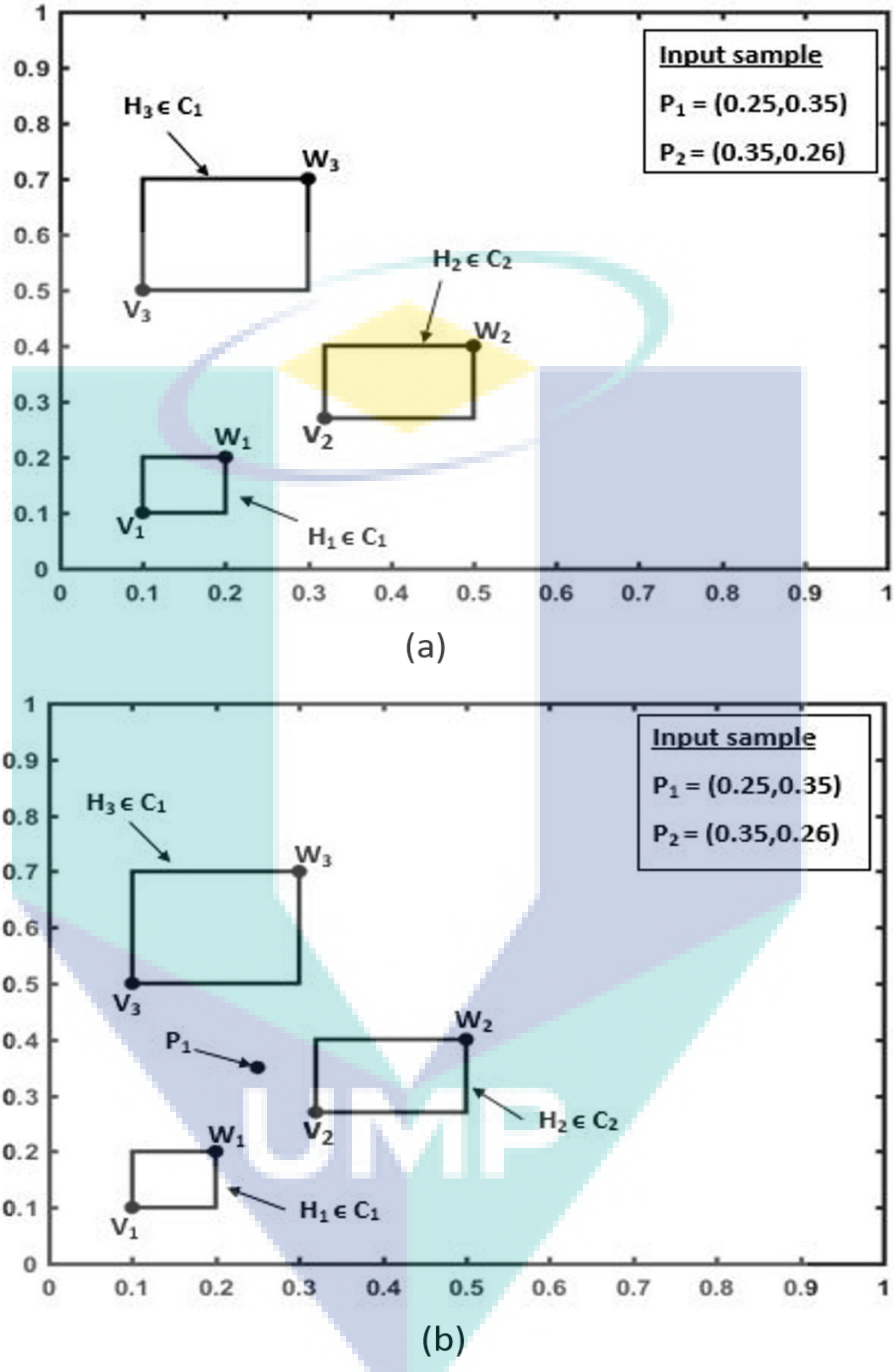
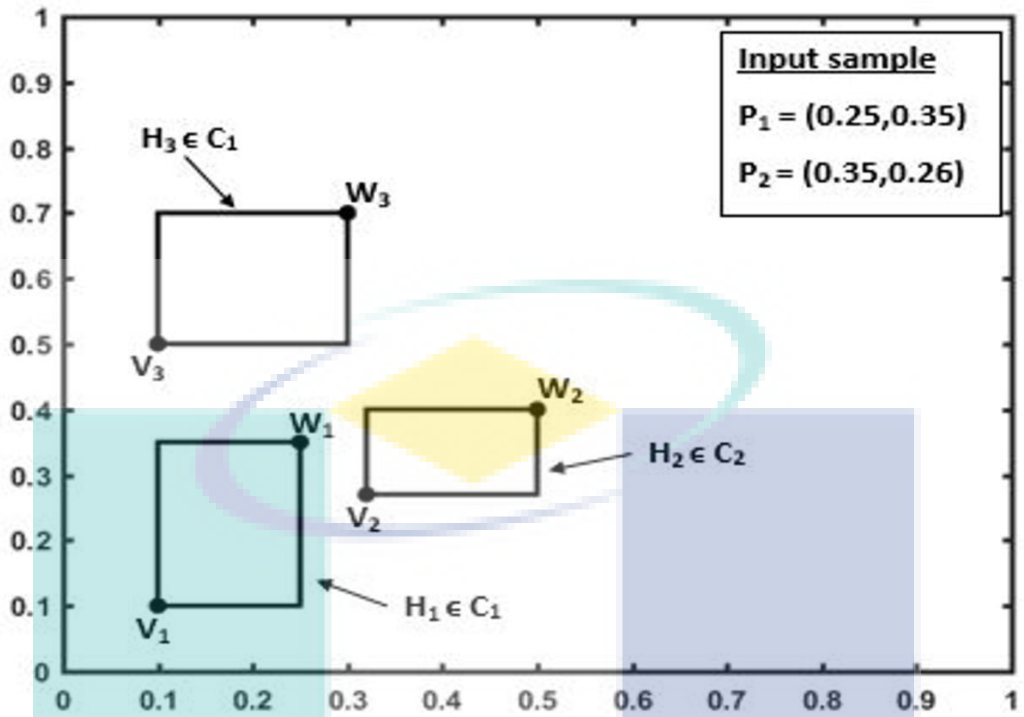
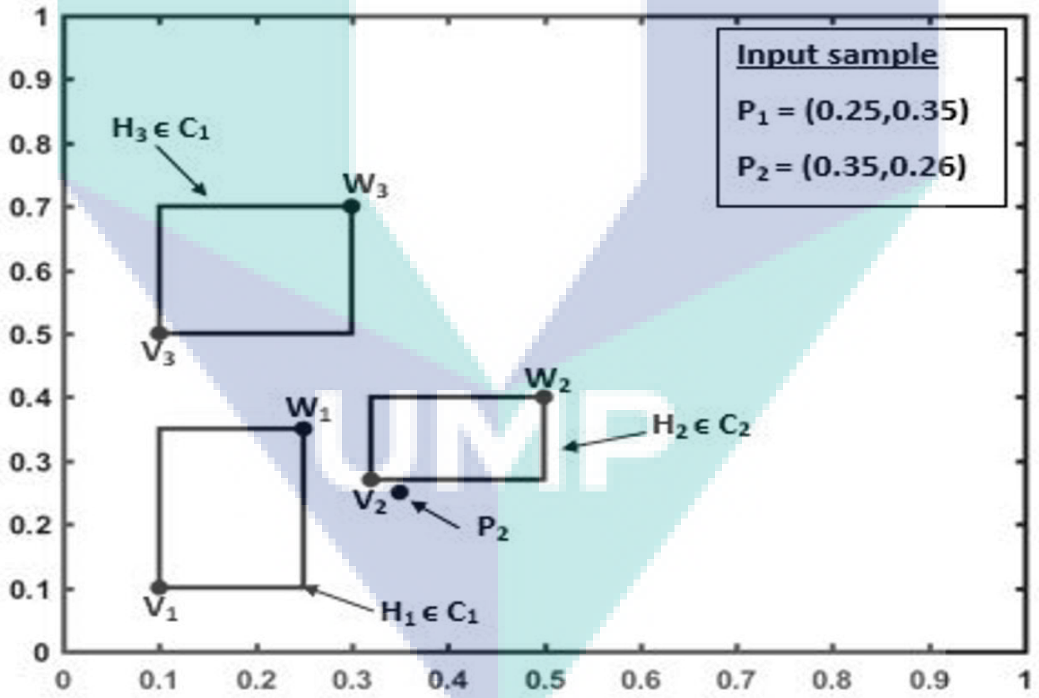


Figure 4.2 FMM k-nearest expansion process

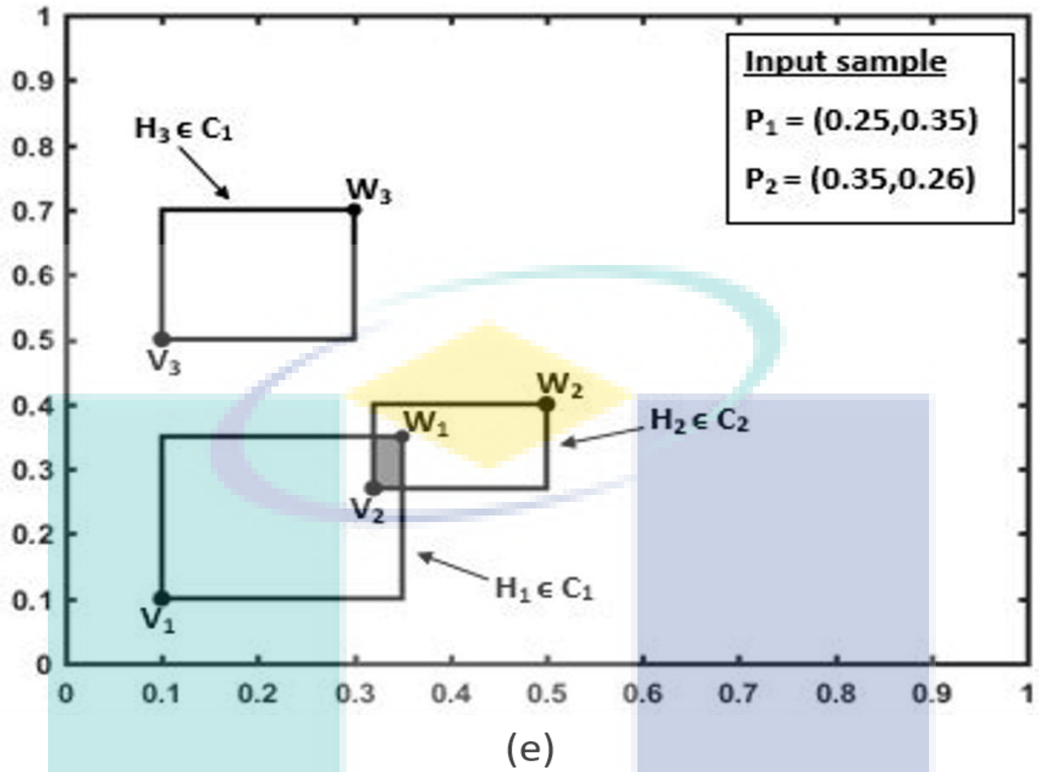


(c)



(d)

Figure 4.2 Continued



(e)

Figure 4.2 Continued

In the Kn expansion rule, all hyperboxes that exhibit the same class label as the winning one are selected, and expansion coefficient (Eq. 3.5) is applied to all selected hyperboxes. Firstly, the hyperbox with a high fitness value is selected, and the expansion coefficient (Eq. 3.5) is utilised to evaluate the ability of the winning hyperbox to include the new input pattern. If the winning hyperbox is prohibited from expanding because expansion coefficient (Eq. 3.5) is violated, then the next nearest hyperbox is selected as the winning hyperbox. In general, the first Kn hyperboxes that can satisfy Eq. 3.5 are selected for expansion. If all selected hyperboxes cannot meet expansion coefficient (Eq. 3.5), then a new hyperbox is created.

Fundamentally, using the Kn technique with the MDFMM model leads to effective results. It is needed to use equations (Eq. 3.19) and (Eq. 3.20) to integrate the MDFMM model and Kn technique, which is introduced by the MDFMM model instead of Eq. 3.5. The integration of the Kn technique with the MDFMM model is described in Section 4.4.

4.4 MDFMM Based on K-Nearest Technique (MDFMM-Kn)

The FMM complexity problem is solved by applying the Kn technique to prevent the generation of unnecessary hyperboxes in the network structure (Section 4.1) (Mohammed and Lim 2017a). Although FMM is the backbone of the MDFMM model, it tends to generate more hyperboxes than the original model does. This incremental in the MDFMM structure is related to its nature to create a highly accurate knowledge structure by solving the data distortion problem. The Kn technique integrated with the MDFMM model is applied to reduce network complexity. It needs to use the MDFMM expansion equation (Eq. 3.19) and overlap test (Eq. 3.20), which were introduced in Chapter 3 to overcome the irregularity and leniency problems.

The MDFMM-Kn can be further clarified by revisiting the example in Section 4.4 (Figure 4.2) and comparing it with that in Figure 4.3. Assume three hyperboxes exist in a 2D space (i.e. H_1 and $H_3 \in C_1$ and $H_2 \in C_2$ with minimum and maximum points V_1 = (0.1,0.1) and W_1 = (0.2,0.2), V_3 = (0.1,0.5) and W_3 = (0.3,0.7) and V_2 = (0.32,0.27) and W_2= (0.5,0.4), respectively. Assume that two input samples (i.e. P_1 = (0.25,0.35) and P_2 = (0.35,0.26)) belong to the same class C_1, as shown in Figure 4.3(a). When P_1 is applied, MDFMM uses Eq. 3.1 to select the nearest hyperbox that belongs to the same P_1 class as a winner. Accordingly, H_3 is selected to contain P_1, as shown in Figure 4.3(b). The next nearest hyperbox (H_1) that belongs to the same class is expanded because expanding H_3 results in the violation of the expansion coefficient in Eq. 3.19. According to Eq. 3.19 and Eq. 3.20, H_1 can contain P_1, as shown in Figure 4.3(c). In this regard, the H_1 minimum point remains the same (i.e. V_1 = (0.1,0.1)) whilst the maximum point is updated to W_1 = (0.25,0.35).

When the next input sample P_2 is applied, hyperbox H_1 is selected as the winner. However, expanding H_1 causes it to overlap with other hyperboxes from different classes (H_2). Hence, a new hyperbox (H_4) is created with the same minimum and maximum points (i.e. V_4 = W_4= (0.35,0.26)), as shown in Figure 4.3(e).

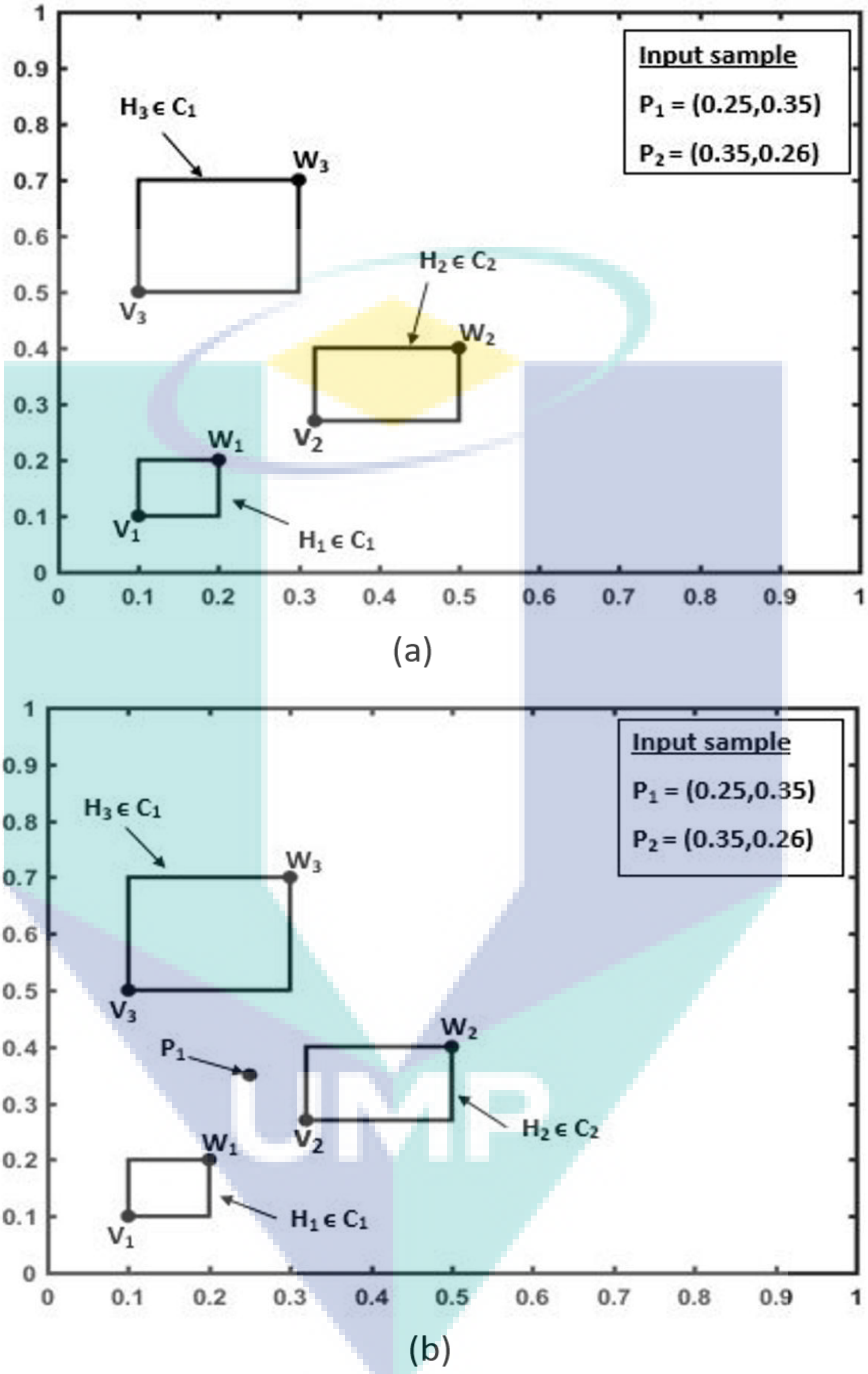
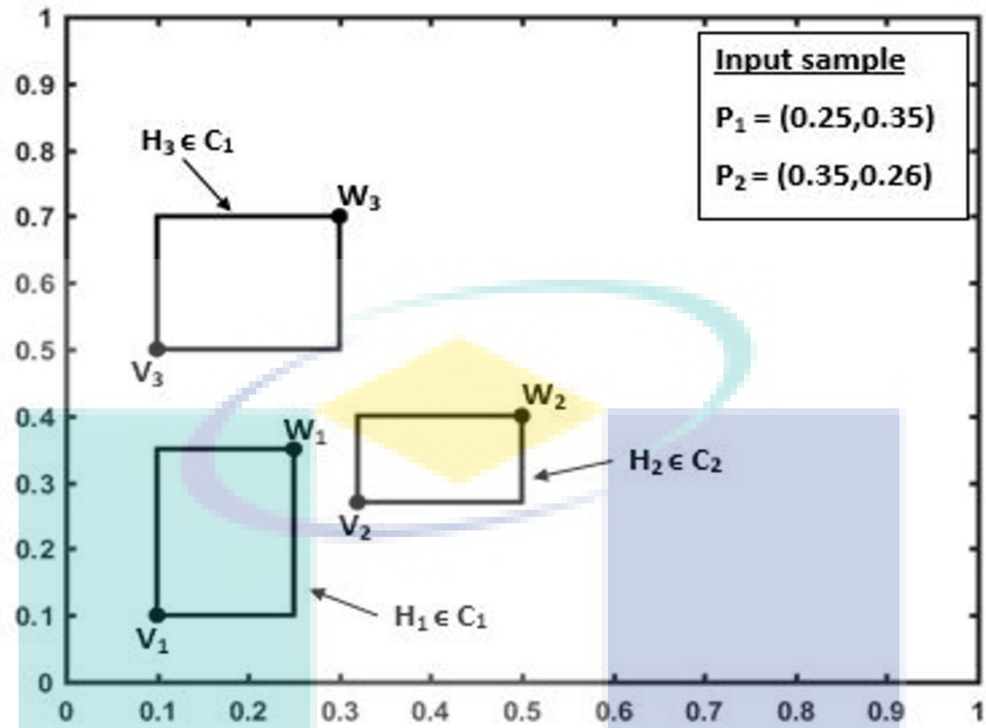
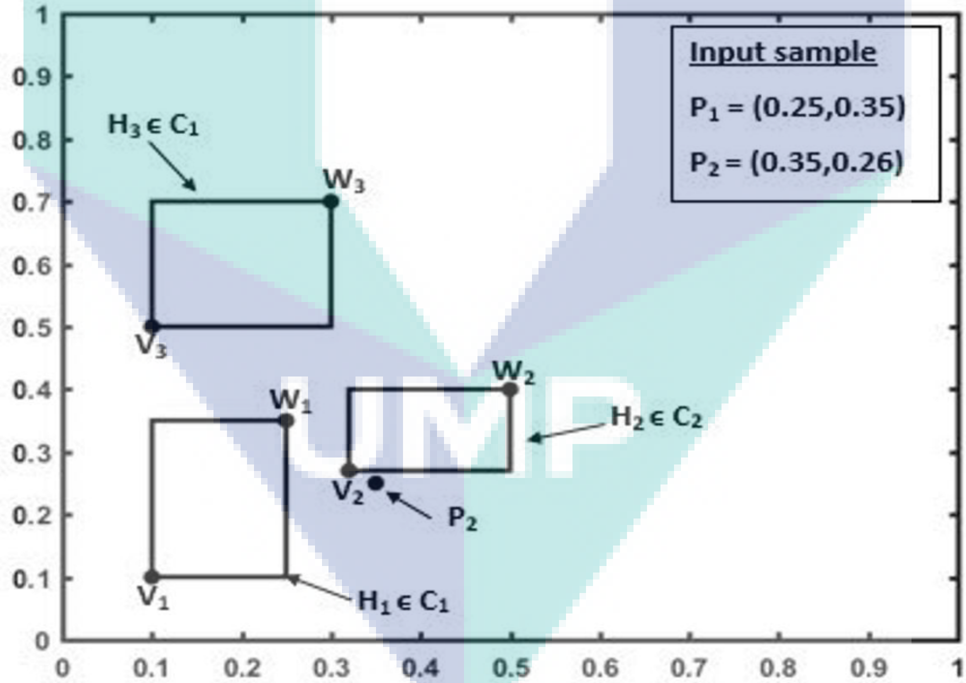


Figure 4.3 MDFMM-Kn expansion process



(c)



(d)

Figure 4.3 Continued

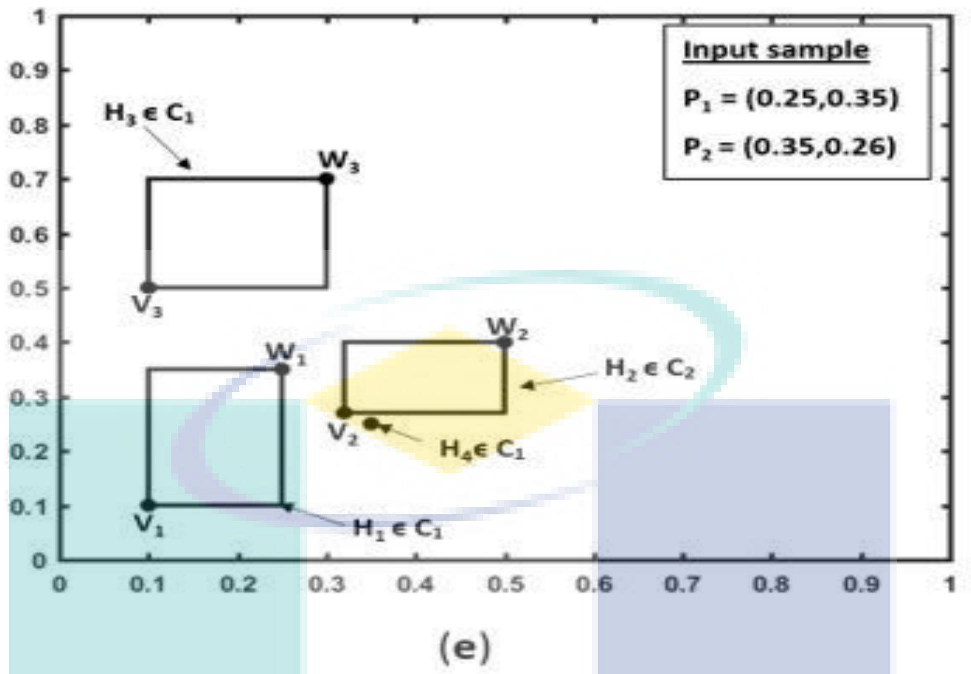


Figure 4.3 Continued

As illustrated in Figure 4.3, the MDFMM-Kn leads to the creation of four hyperboxes, which are compared with those in the MDFMM model shown in Figure 4.1. As a result, the Kn hyperbox expansion rule can reduce the network structure with marinating the classification rate.

4.5 Performance Evaluation

MDFMM-Kn was compared with the MDFMM and other popular FMM variants, such as FMM, GFMM, EFC, FMCN, DCFMN and MLF, to evaluate its performance. Three experiments were conducted to comprehensively assess the performance of MDFMM-Kn. In the first experiment, six benchmark datasets (i.e. Iris, Wine, WBC, PID, Page Blocks and Thyroid) from the UCI machine learning repository were utilised. The hyperbox size (θ) was varied from 0.1 to 0.9 in step 0.1 to show the effect of the expansion coefficient value on the network accuracy and complexity. In the second experiment, four 2D artificial datasets (i.e. Two Spirals, Corner, Half Kernel, and Outlier) with different training sizes were used to evaluate the effectiveness of changing the training size. In the third experiment, seven UCI benchmark datasets (i.e. Iris, Glass, Ionosphere, Thyroid, Parkinson, Ozone level and Spambase) were utilised to evaluate the noise effect on classifier performance. Hence, various hyperbox sizes and noise levels were adopted during the evaluation process. Table 4.1 summaries the UCI benchmark datasets, which are used in the first and third experiments.

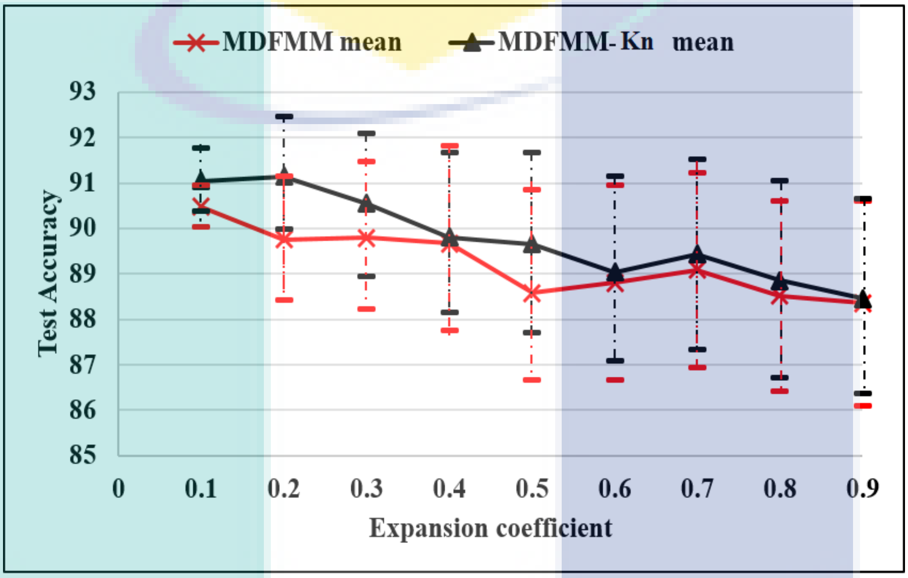
Table 4.1 Description of UCI benchmark datasets

Data set	Features	Classes	Sample size
Iris	4	3	150
Seed	7	3	210
Wine	13	3	178
WBC	9	2	699
PID	8	2	768
Page blocks	10	5	5473
Thyroid	21	3	7200
Glass	9	7	214
Ionosphere	34	2	351
Parkinson	22	2	197
Ozone level	72	2	2536
Spambase	57	2	4601

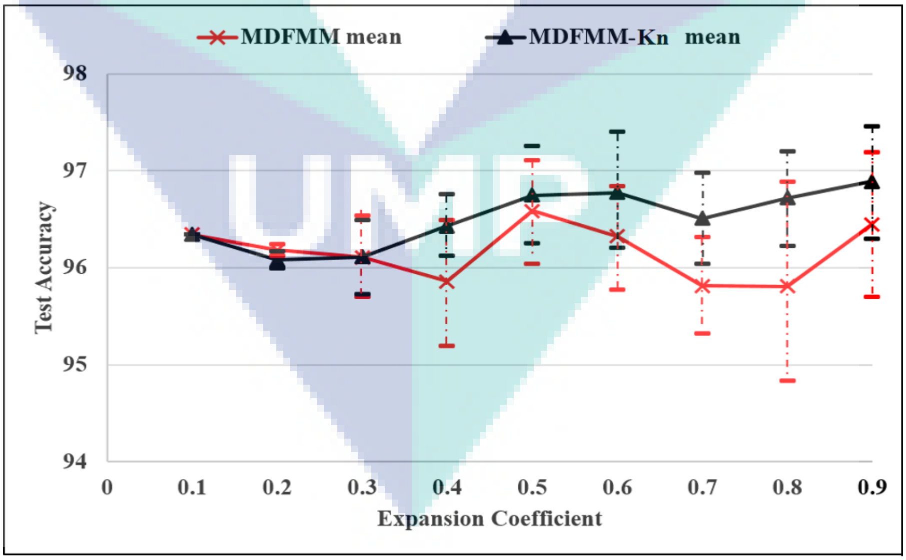
4.5.1 Experiment 1

The experiment aims to evaluate the effect of increasing the hyperbox size on the performance of MDFMM-Kn and MDFMM. Six UCI benchmarks datasets (i.e. Iris, Wine, WBC, PID, Page Blocks and Thyroid) were utilised. Figure 4.4 shows the classification rates, along with the 95% confidence intervals, of MDFMM-Kn and MDFMM. A five-fold cross-validation was performed to estimate the generalisation performance of both models. A series of systematic evaluations was conducted by increasing the hyperbox size from 0.1 to 0.9 in step 0.1. The experiment was repeated ten times for each hyperbox size. The outcomes were computed using a bootstrap method.

Figure 4.4 illustrates the accuracy rates (%) of MDFMM and MDFMM-Kn. The error bars represent the 95% confidence intervals rated using the bootstrap method. As illustrated in the figure, MDFMM and MDFMM-Kn produced almost consequences result when the hyperbox size was small, except for the Thyroid dataset. In summary, the MDFMM-Kn model outperformed MDFMM. The average (bootstrap) test accuracy rates of MDFMM-Kn for the different datasets exceeded those of MDFMM when the hyperbox size ranged from 0.1 to 0.9, as shown Figure 4.4(a–f). In Figure 4.4(d–f), MDFMM-Kn outperformed MDFMM in all hyperbox sizes, especially in Figure 4.4(e), where MDFMM-Kn statistically exceeded MDFMM (at the 95% confidence level) when the hyperbox size ranged from 0.1 to 0.9, 0.1 to 0.3, and 0.4 to 0.6 for Figure 4.4 (e) , Figure 4.4 (d), and Figure 4.4 (f), respectively. As shown in the figure, MDFMM-Kn performed more stably and obtained better average accuracy values in comparison with the MDFMM model for all hyperbox sizes.

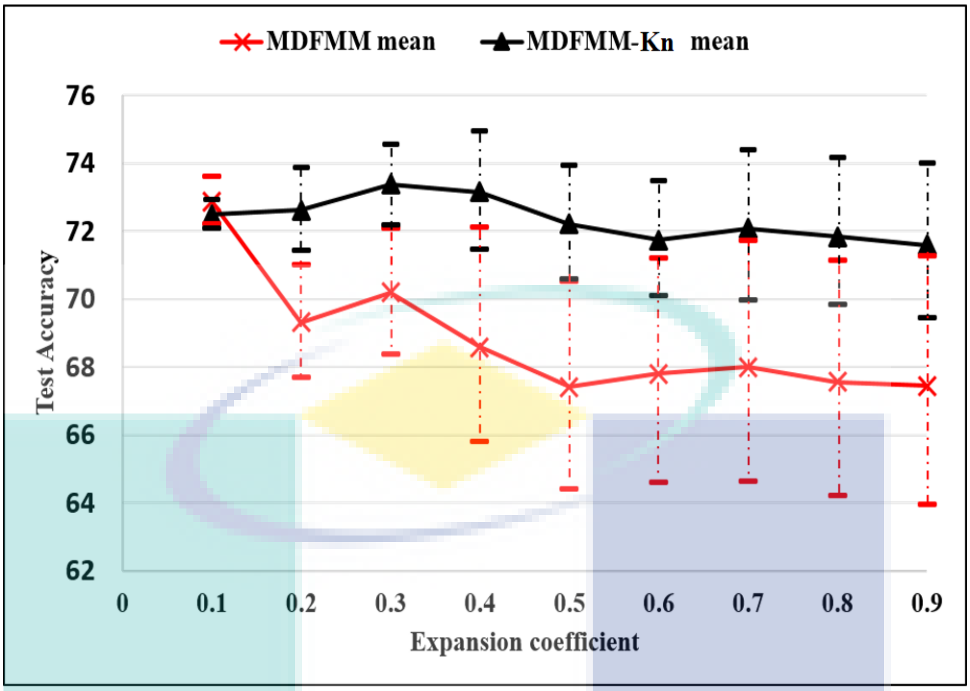


(a) Seed data set

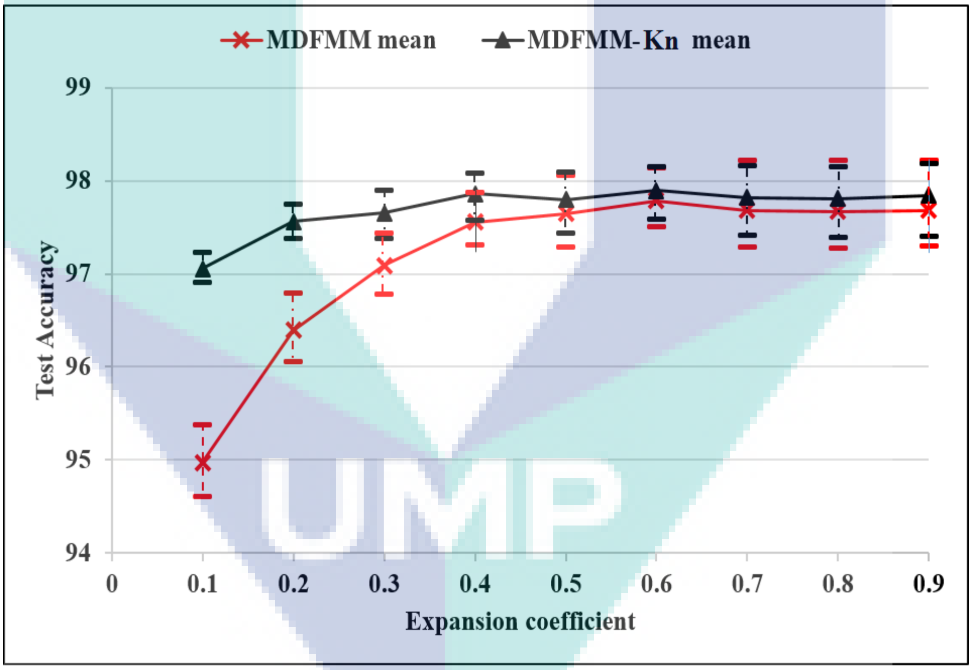


(b) WBC data set

Figure 4.4 The average test accuracy of MDFMM and MDFMM-Kn

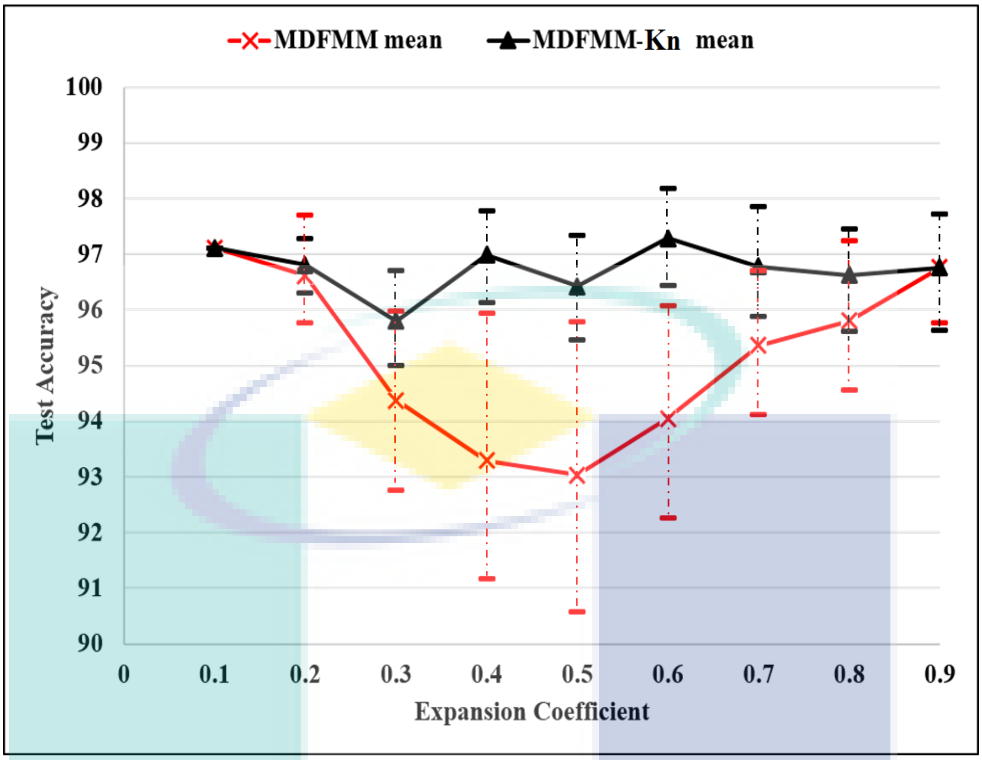


(c) PID data set



(d) Thyroid data set

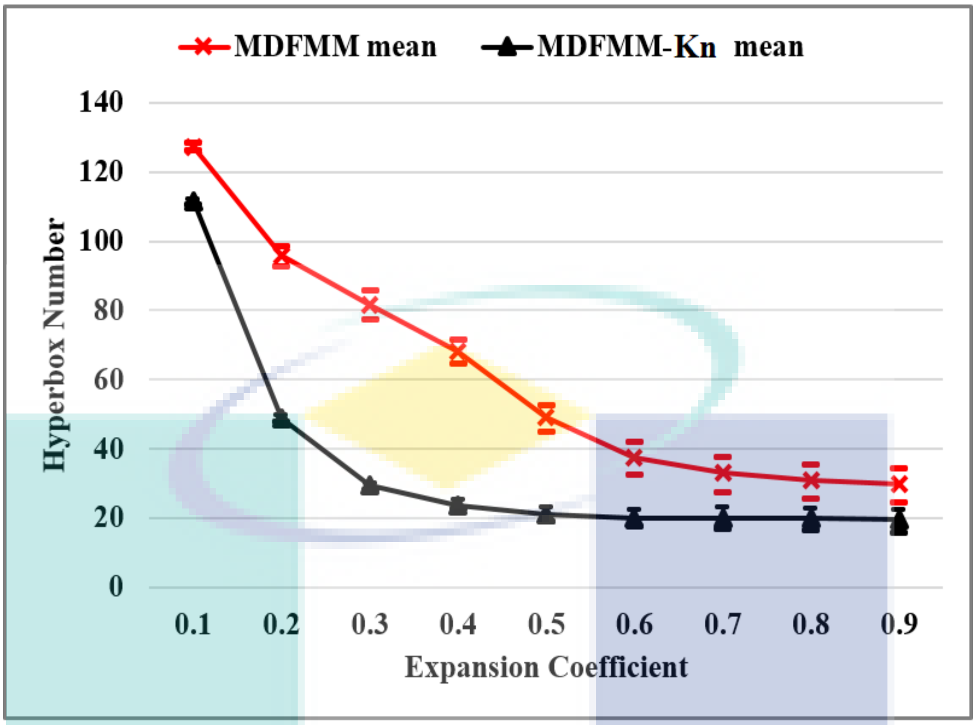
Figure 4.4 Continued



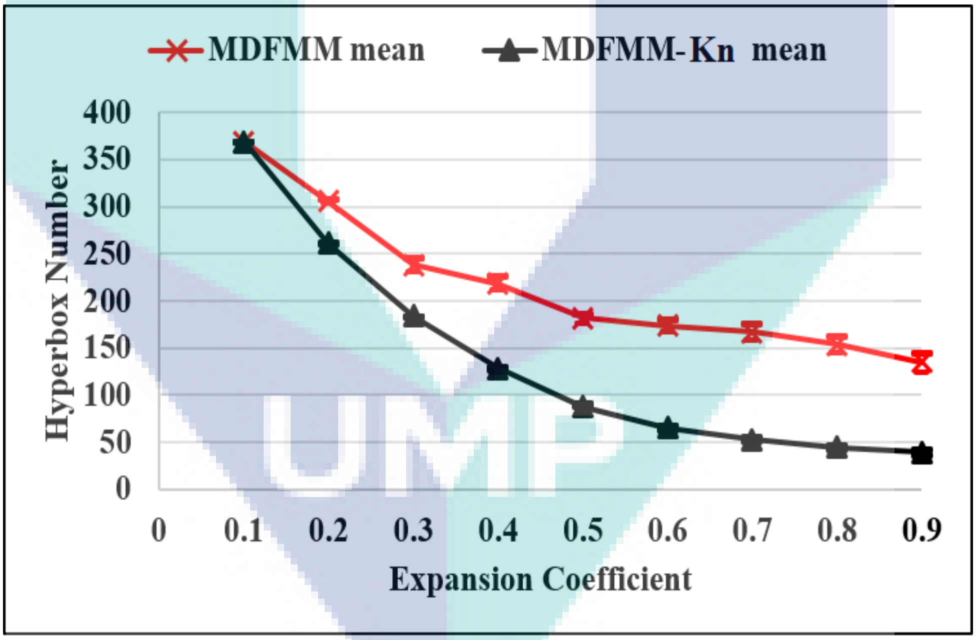
(f) Wine data set

Figure 4.4 Continued

In terms of complexity, Figure 4.5 shows the average number of hyperboxes generated by MDFMM-Kn and MDFMM. All datasets were used for MDFMM and MDFMM-Kn. The result showed that MDFMM generated a more complex network structure than MDFMM-Kn. Many small hyperboxes within the vicinity of the hyperbox were created because MDFMM tends to create numerous hyperboxes to provide a highly accurate description of the network structure and generate substantial precise decision boundaries, as explained in Section 4.2 and shown in Figure 4.2. All these figures show that MDFMM-Kn produced fewer hyperboxes than the MDFMM network. MDFMM-Kn statistically outperformed FMM for all hyperbox size settings, as shown in Figure 4.5. Specifically, MDFMM-Kn can create a parsimonious network structure without negatively affecting the accuracy.

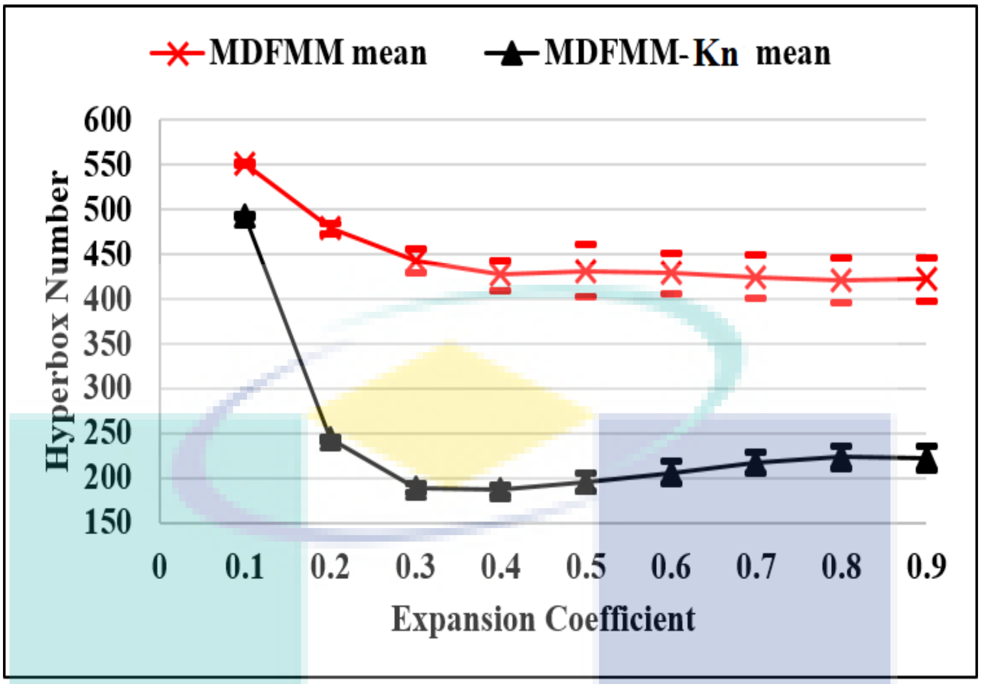


(a) Seed data set

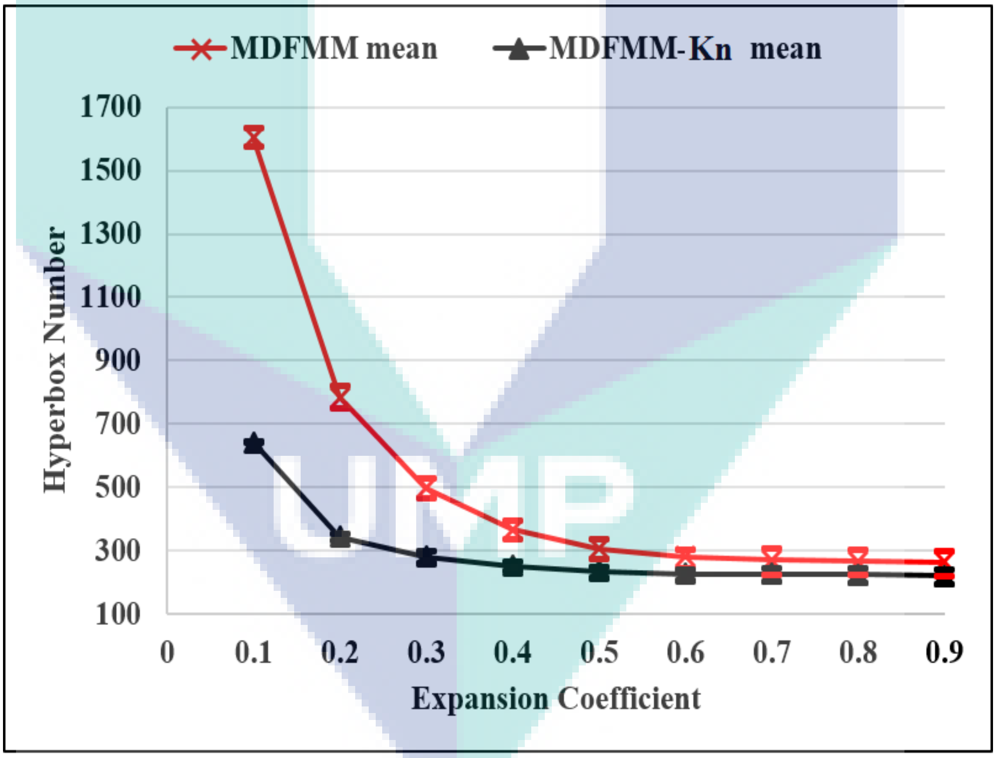


(b) WBC data set

Figure 4.5 The average number of hyperboxes of MDFMM and MDFMM-Kn for different benchmark datasets

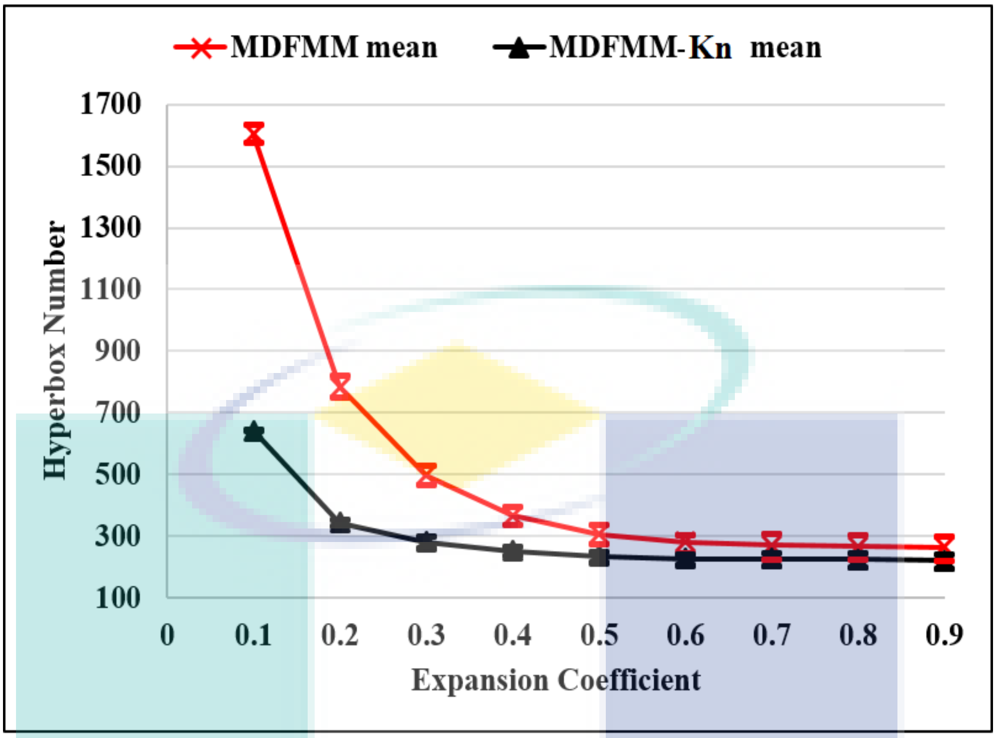


(c) PID data set

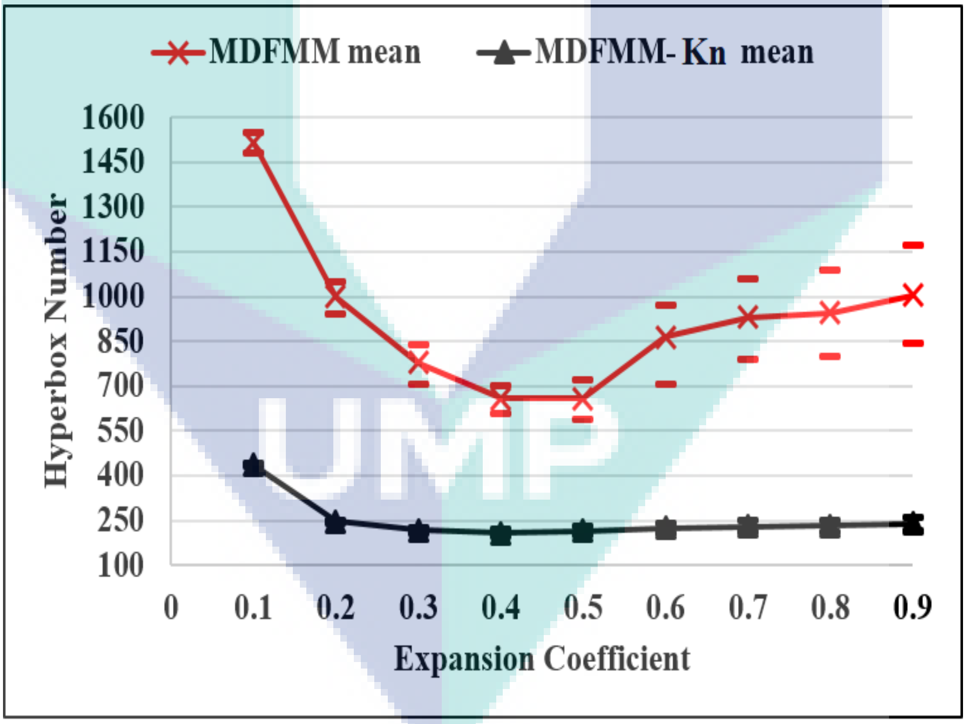


(d) Thyroid data set

Figure 4.5 Continued

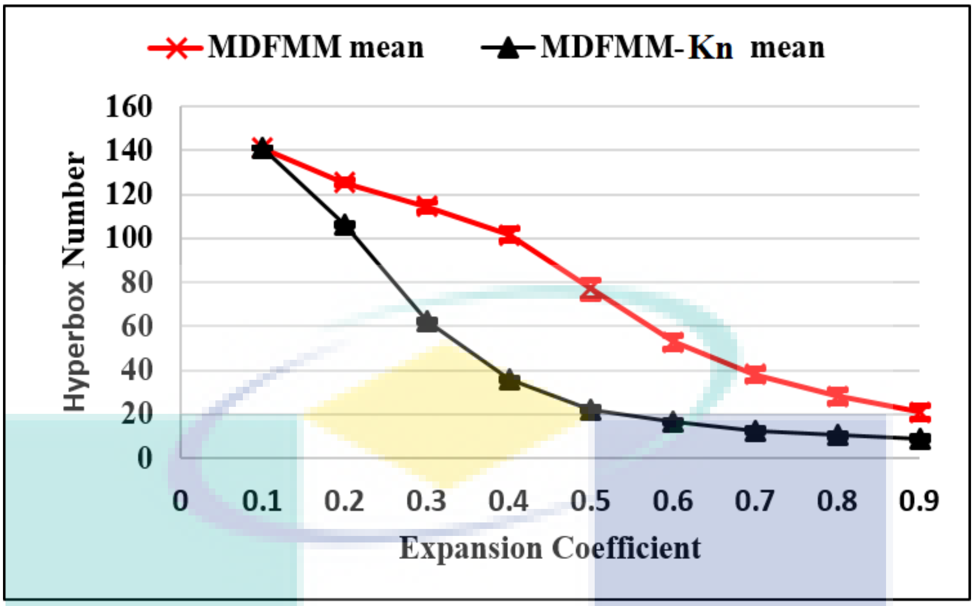


(d) Thyroid data set



(e) Page block data set

Figure 4.5 Continued



(f) Wine data set

Figure 4.5 Continued

4.5.2 Experiment 2

This experiment aims to assess the effect of using different training size samples on the performance of MDFMM-Kn and compare the outcomes with those of MDFMM and FMM networks. Four datasets (Two Spirals, Kernel, Corners and Outlier) were generated for performance validation. Table 4.2 shows the statistical information of the created datasets. In this experiment, the training sizes for all datasets were randomly generated instances and varied from 1000 to 10000 whilst 1000 new instances were used for testing. The hyperbox size was set to a small value (0.05) to maximise the classification performance. Figure 4.6 shows the original image for the created dataset.

Table 4.2 Statistical information of created datasets

Data set	Features	Classes	Training size	Testing size
Two Spiral	2	2	10000	1000
Kernel	2	2	10000	1000
Corners	2	4	10000	1000
Outlier	2	4	10000	1000

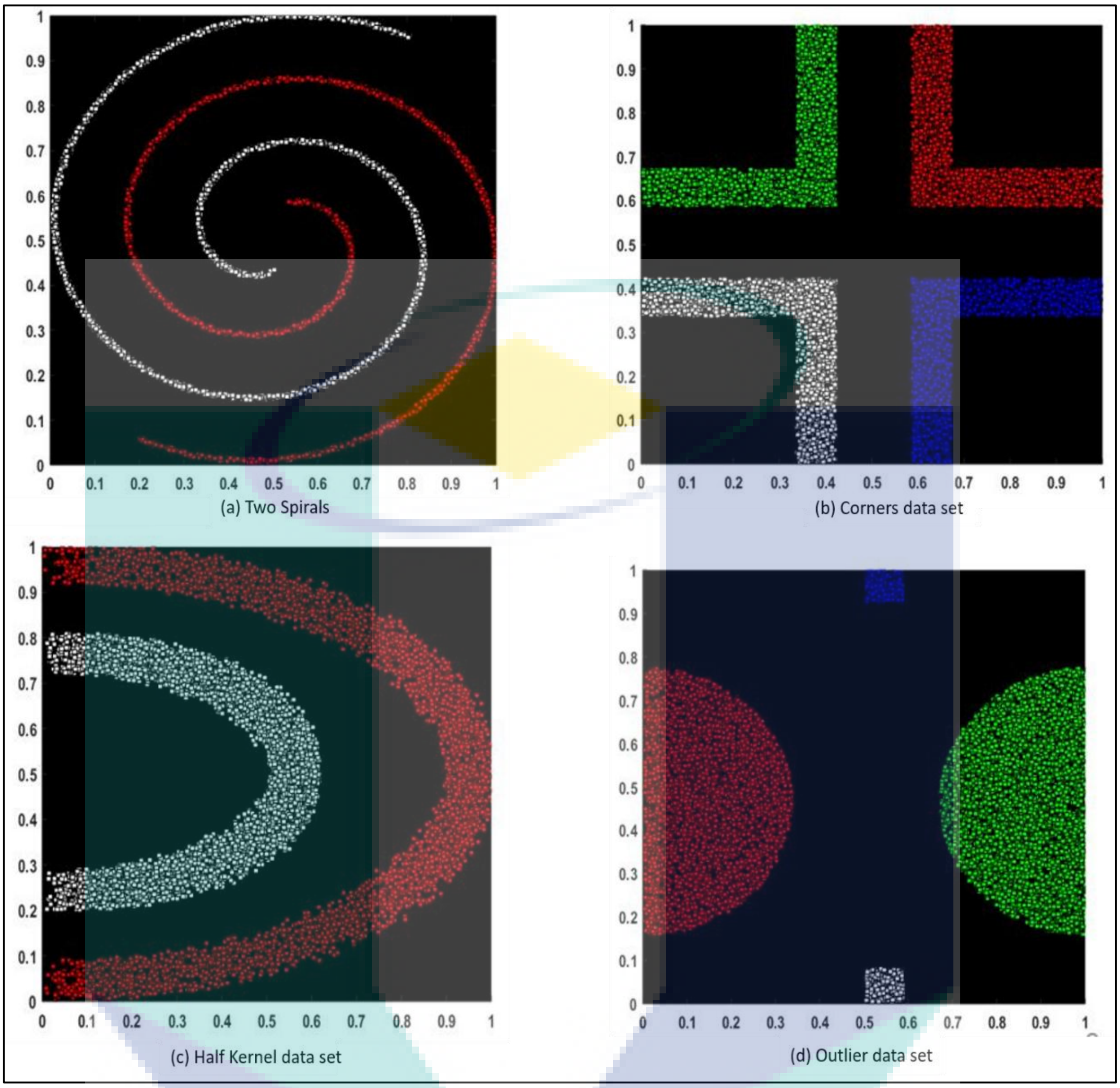


Figure 4.6 The original shape for the created datasets

The results for this experiment are summarised in Table 4.3. The three models (i.e. MDFMM, MDFMM-Kn and FMM) obtained the same accuracy, as shown in the table. In terms of complexity, the MDFMM-Kn outperformed FMM and MDFMM by generating a simple network structure. Based on Table 4.3, the difference in network complexity increases by increasing the training samples, where the MDFMM-Kn generated less network structure as compared with MDFMM and FMM neural network. In other words, integration the MDFMM with Kn expansion rule technique result in reduced network complexity.

These results showed that MDFMM-Kn can maintain its accuracy whilst forming a less complex network structure relative to MDFMM and FMM (Table 4.3). This finding verified the effectiveness of integrating the Kn expansion rule with the MDFMM model in decreasing the MDFMM network structure, increasing/maintaining the classification rate and making the network highly robust towards increasing the hyperbox size.

Table 4.3 Comparison of MDFMM-Kn results with FMM and MDFMM classifiers

Dataset	**Training size**	**Models**					
		FMM		**MDFMM**		**MDFMM-Kn**	
		Accuracy	Complexity	Accuracy	Complexity	Accuracy	Complexity
Two Spirals	1000	100	140	100	163.10	100	**126.70**
	5000	100	320.9	100	300.20	100	**157.10**
	10000	100	515.1	100	411.20	100	**163.50**
Corners	1000	100	277.9	100	244.30	100	**132.30**
	5000	100	823.8	100	707.90	100	**184.90**
	10000	100	1282.7	100	1073.70	100	**198.00**
Half Kernel	1000	100	297.90	100	275.10	100	**153.70**
	5000	100	957.00	100	828.20	100	**237.90**
	10000	100	1493.90	100	1294.10	100	**265.40**
Outlier	1000	100	302.10	100	276.20	100	**148.70**
	5000	100	961.50	100	841.00	100	**216.60**
	10000	100	1505.40	100	1285.80	100	**241.30**

In the second sub-experiment, MDFMM is compared with FMM and MDFMM networks in terms of the creation of small hyperboxes within the vicinity. In this sub-experiment, four 2D artificial datasets (i.e. Two Spirals, Corners, Half Kernel and Outlier) were used. Figures 4.7, 4.8 and 4.9 show the network structure of MDFMM-Kn, MDFMM and FMM, respectively. The training and test sizes for each dataset were 10,000 and 1000, respectively. The hyperbox size was θ=0.05, and this experiment was conducted ten times for each dataset. As shown in Figures 4.7, 4.8 and 4.9, MDFMM-Kn outperformed FMM and MDFMM by generating a simple network structure. It is clear that the winner selection method that used only the sole winning hyperbox leads to the creation of many small hyperboxes within its vicinity; thus increasing the MDFMM and FMM network complexity unnecessarily, as shown in Figures 4.7, 4.8 and 4.9.

These Figures 4.7, 4.8, and 4.9 demonstrate the ability of MDFMM-Kn in preventing the creation of hyperboxes with small sizes. Specifically, the integration of the Kn expansion rule with MDFMM contributed to the decrease of the network complexity.

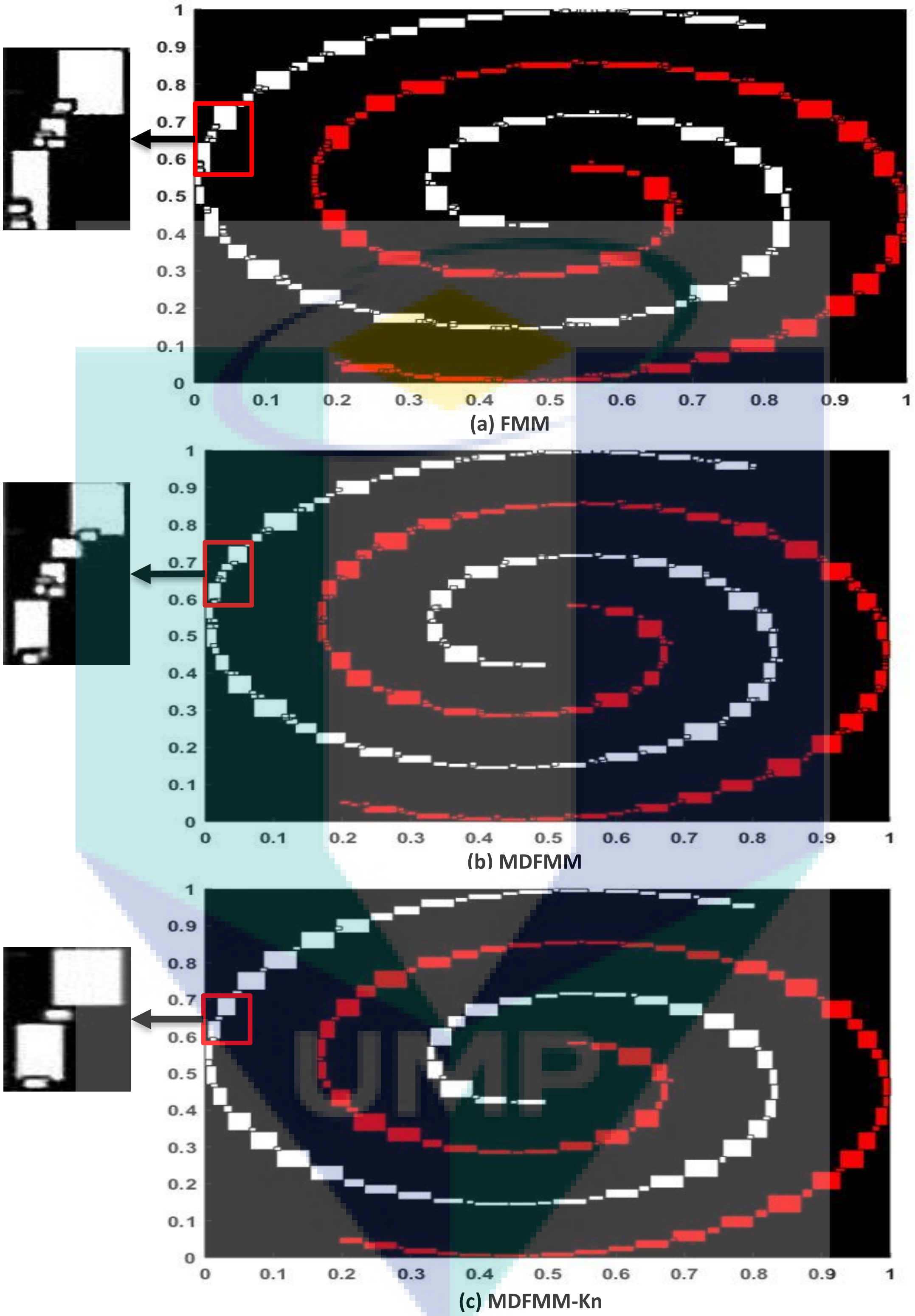


Figure 4.7 The two spirals data set with training and testing size 10000 and 1000, respectively

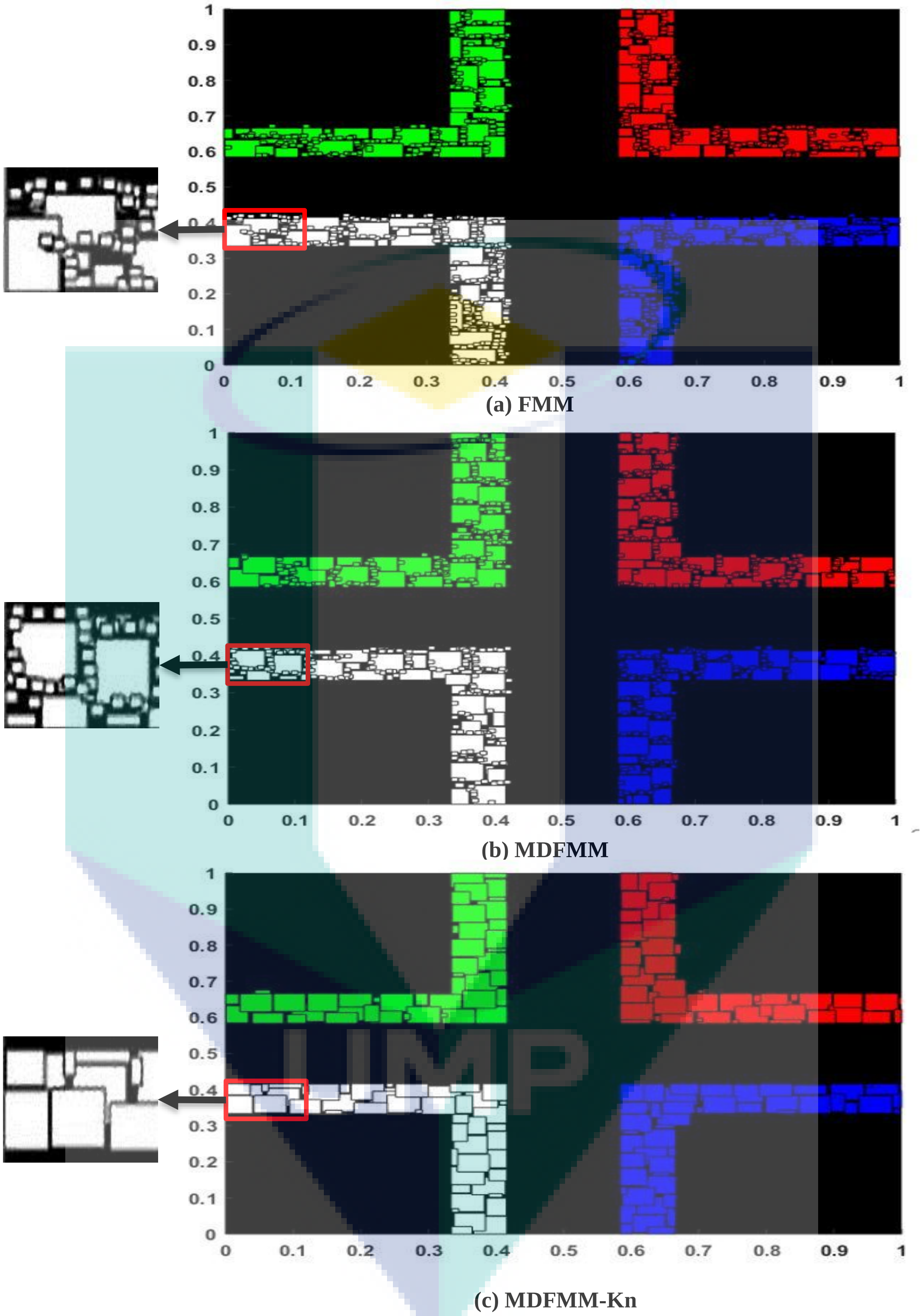


Figure 4.8 The Corners data set with training and testing size 10000 and 1000, respectively

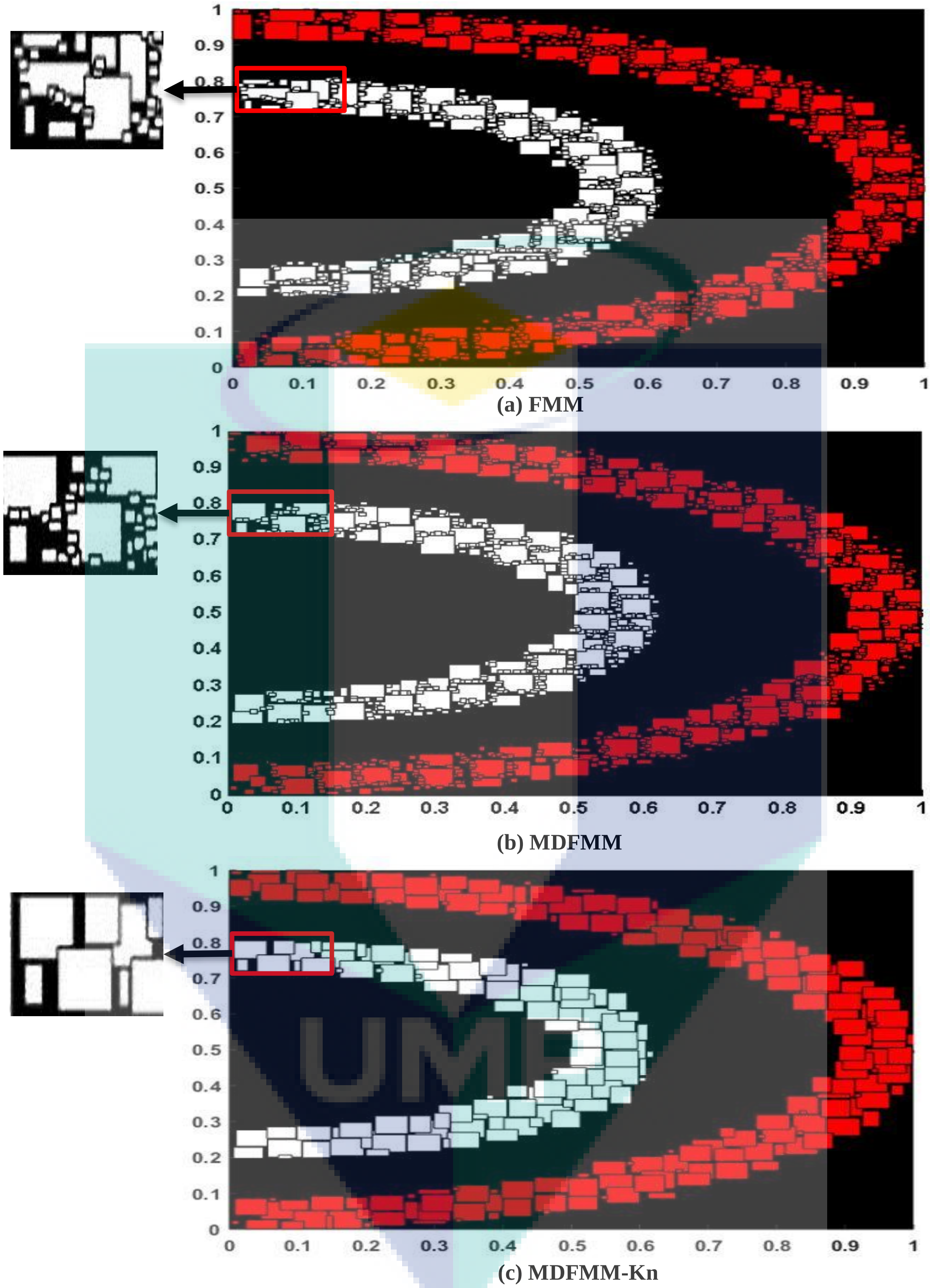


Figure 4.9 The Half Kernel data set with training and testing size 10000 and 1000, respectively

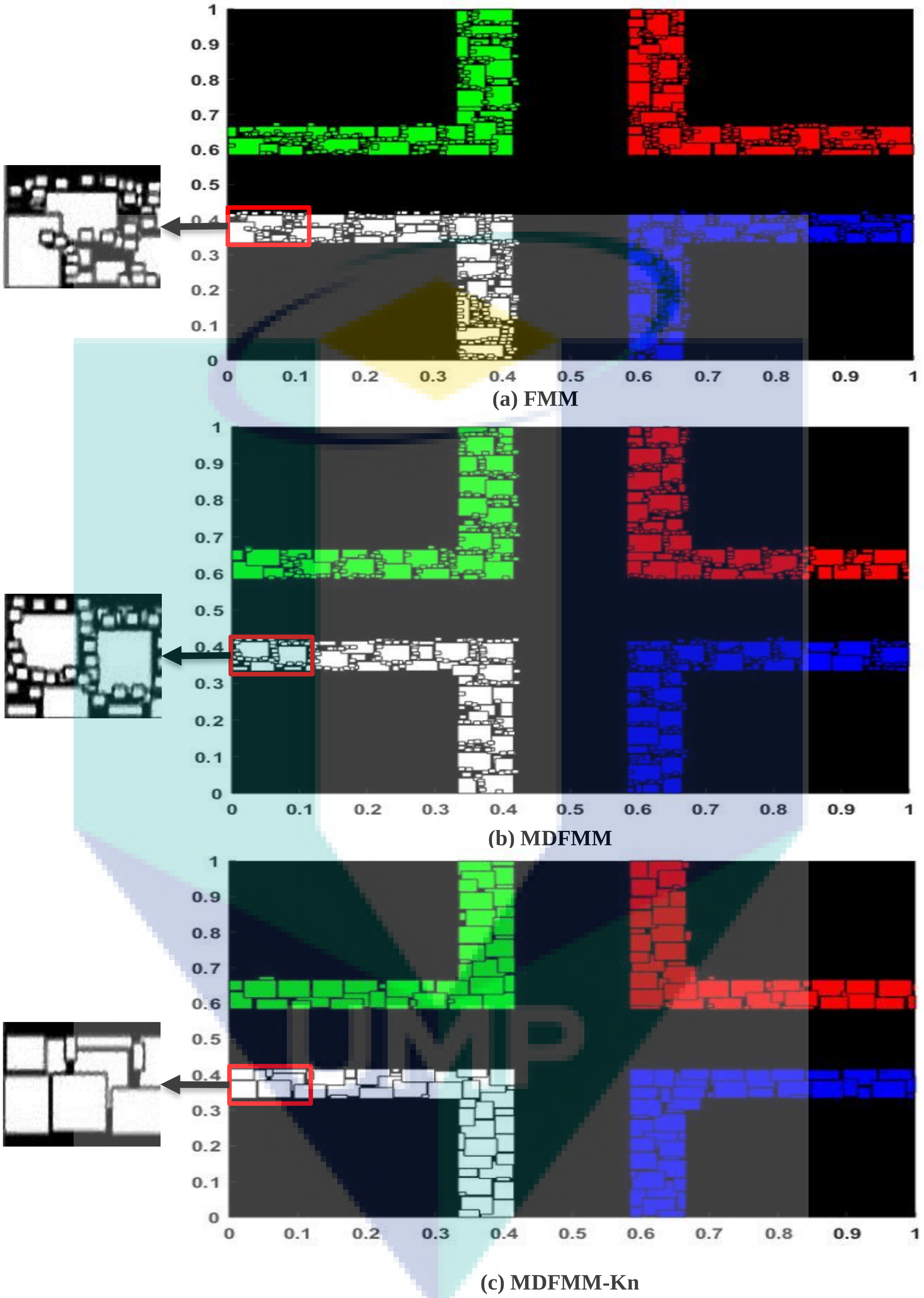


Figure 4.10 The Outlier data set with training and testing size 10000 and 1000, respectively

4.5.3 Experiment 3

This experiment aims to assess the effect of adding noise to the dataset on the performance of MDFMM and MDFMM-Kn networks and compare the results with those of the other common variants of FMM. Nine benchmark datasets (i.e. Iris, Glass, Wine, Ionosphere, Thyroid, Ozone level, Parkinson and Spambase) were used in this experiment. In this experiment, the training and testing samples was corrupted with 5%, 10%, and 15% noise levels. The noise was added to the class labels for each selected benchmark dataset to test the robustness of MDFMM and MDFMM-Kn and other classifiers against noise. A three-fold cross validation was carried out in accordance with the experimental procedures reported by (Davtalab et al. 2014), and the size of the hyperbox (θ) was varied from 0.0 to 0.1 in increments of 0.02. Table 4.4 shows the results of misclassification in different models with noise datasets. As shown in the table, MDFMM-Kn and MDFMM performed better than the other models did because it achieved the minimum misclassification rate for all experiments.

Table 4.4 Furthermore, MDFMM-Kn obtained the lowest misclassification rate for various noise levels relative to the different variants of FMM. However, the MLF neural network showed a low misclassification rate in the Iris dataset at the 5% noise level. By contrast, MDFMM-Kn showed superiority in the same dataset at the 10% and 15% noise levels. As shown in Table 4.4, MDFMM and MDFMM-Kn obtained the same low misclassification rates in the Wine, Ozone level and Parkinson datasets across all noise levels. Moreover, MDFMM-Kn achieved low misclassification rates for the Glass, Thyroid, Spambase, Iris and Ionosphere datasets. In this experiment, MDFMM-Kn showed high tolerance to noise and could thus create a pure decision boundary with few hyperboxes (Section 3.4). In noisy dataset, the noisy samples cause the generation of some hyperboxes with the wrong class label. The size of these wrong hyperboxes relies on the expansion coefficient and, when a large value is selected for hyperbox size (θ), these hyperboxes will enlarge. As a result, a large number of overlapped areas will be generated. This issue leads to increase the misclassification rate in FMM and its variants. As MDFMM-Kn is reduced the misclassification rate in overcoming the irregular dimension shape and overlap leniency by incorporating the proposed modifications. In other words, MDFMM-Kn is able to reduce the overlap between hyperboxes belonging to different classes by preventing the creation of hyperboxes of a large size. In addition, the k-nearest hyperbox selection rule is able to establish accurate decision boundaries by avoiding the creation of too many small hyperboxes in the vicinity of the wining hyperbox (using the k-nearest expansion rule, as explained in Section 4.3 and 4.4). In summary, these results indicate that MDFMM-Kn and MDFMM are not as sensitive to noisy data as other related classifiers are due to the modifications added to both networks. The result is the creation of pure decision boundaries and undistorted hyperboxes.

Table 4.4 Performance comparison in percentage between MDFMM-Kn and MDFMM, and variants of FMM models highlighted in Davtalab et al. (2014) for different benchmark datasets

Dataset	Noise Level	Models							
		GFMM	EFC	FMCN	DCFMN	MLF	EFMM	MDFMM	MDFMM-Kn
Iris	5%	15.56	12.08	13.51	11.34	**6.53**	7.36	8.28	**8.27**
	10%	29.61	27.91	27.51	20.91	15.74	8.13	10.77	**10.69**
	15%	30.23	32.19	29.01	25.95	19.85	10.94	**13.09**	**13.09**
Glass	5%	58.48	50.33	52.68	45.80	39.74	32.76	31.45	**31.32**
	10%	61.20	51.80	53.44	46.10	42.46	33.45	**31.42**	**31.42**
	15%	64.46	57.61	58.53	53.06	44.18	35.56	27.87	**27.8**
Wine	5%	15.61	9.86	11.66	10.23	9.76	11.53	**7.85**	**7.85**
	10%	16.03	12.13	12.82	12.63	11.72	11.72	**8.55**	**8.55**
	15%	23.34	18.94	19.48	18.29	17.72	21.97	**12.90**	**12.90**
Ionosphere	5%	64.10	20.26	16.19	15.83	15.18	17.96	**13.14**	**13.14**
	10%	64.10	27.16	20.10	20.72	19.93	22.16	18.88	**18.47**
	15%	64.10	31.65	28.12	26.68	23. 51	25.60	17.6	**17.43**
Thyroid	5%	33.02	28.23	26.59	18.61	11.94	19.4	10.35	**8.41**
	10%	26.37	44.02	35.46	24.69	13.29	25.20	15.11	**12.57**
	15%	47.48	42.36	41.31	28.70	22.61	31.79	20.29	**16.99**
Ozone Level	5%	27.85	20.29	13.03	13.55	11.90	11.97	**10.17**	**10.17**
	10%	32.11	26.32	18.58	18.39	15.48	16.17	**14.83**	**14.83**
	15%	39.91	29.14	23.84	24.66	19.09	21.02	**17.97**	**17.97**
Parkinson	5%	22.23	18.60	17.99	15.28	15.04	17.85	**13.20**	**13.20**
	10%	23.10	24.47	23.73	19.96	19.72	14.87	**12.08**	**12.08**
	15%	25.93	23.12	22.90	21.64	21.01	16.66	**12.74**	**12.74**
Spambase	5%	51.75	48.61	47.26	39.05	17.09	19.72	15.13	**13.98**
	10%	59.18	57.14	49.71	40.07	21.71	21.53	17.54	**16.49**
	15%	60.22	64.80	52.11	43.07	26.56	22.35	19.04	**18.26**

4.5.4 Overall comparative performance analysis

This section aims to measure the overall complexity (number of generated hyperboxes) improvements percentage of the MDFMM-Kn comparative to other selected models, as described in experimental results of experiment 2. Eq. 4.1 is used to calculate the improvement percentage. It is a basic and well known equation of finding the improvement percentage of the performance testing for a technique

$$PIM_i = \frac{\#Hyperboxs\_E_i - \#Hyperbox\_S}{\#Hyperbox\_E_i} * 100 \quad 4.1$$

Where PIM_i is the percentage improvement of MDFMM-Kn model against the *i*th existing model; $\#Hyperboxs\_E_i$ is the number of generated hyperboxes by the *i*th existing model; *#Hyperboxes_S* is the number of generated hyperboxes by the proposed model MDFMM-Kn.

Figure also shows that the space complexity efficiency of MDFMM-Kn is 63% and 61% better than that MDFMM and FMM models. Furthermore, the average performance of MDFMM-Kn against MDFMM and FMM models indicates that its performance is better than the FMM and MDFMM models in term of complexity at a percentage of 62%.

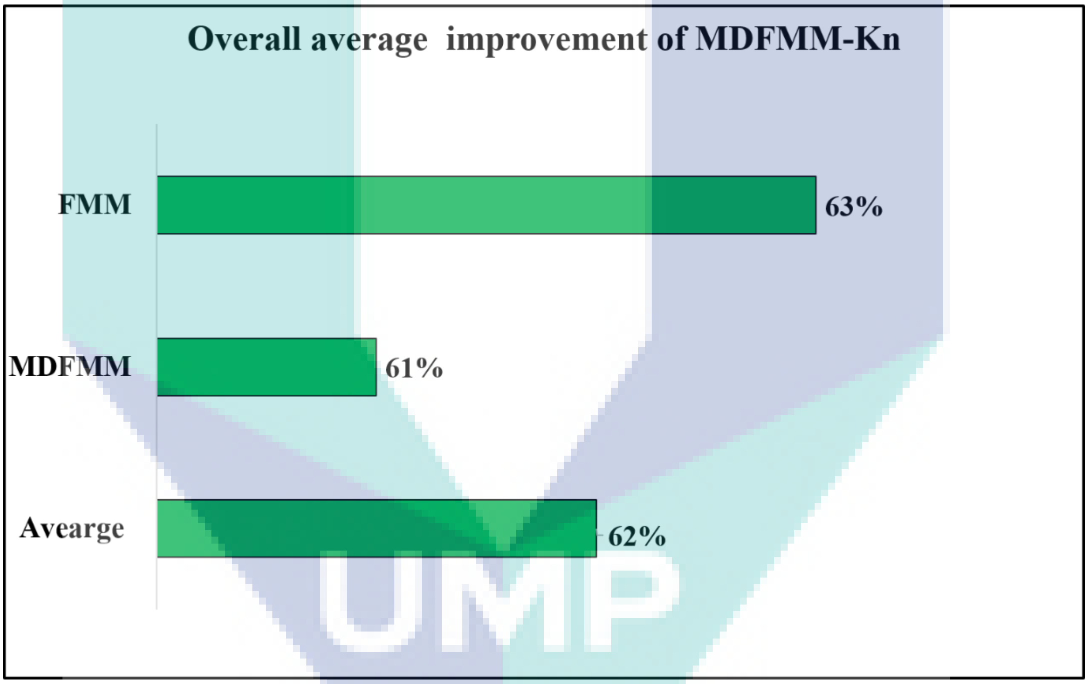


Figure 4.11 performance analysis of MDFMM-Kn with respect to the space complexity

The experimental evaluation result showed that MDFMM-Kn created a balance between complexity and accuracy because it alleviated complexity without diminishing the network performance. Specifically, the performance of MDFMM-Kn was more stable than that of MDFMM, according to Experiments 1, 2 and 3. In addition, MDFMM-Kn decreased the effect of increasing the hyperbox size on the accuracy of the network, thereby indicating that few variations were generated when the maximum sizes of the hyperbox ranged from 0.1 or 0.9. In terms of noisy data, MDFMM-Kn and MDFMM showed an excellent performance relative to other FMM variants.

4.6 Summary

In this chapter, the MDFMM model and K-nearest expansion rule were analysed and explained. MDFMM-Kn was introduced to decrease the number of hyperboxes generated by the MDFMM network. In this approach, Kn was integrated with the MDFMM model (i.e. MDFMM-Kn). The resulting MDFMM-Kn model was evaluated by using different benchmark (i.e. Iris, Wine, Glass, WBC, PID, Thyroid, Page blocks and Ionosphere) and 2D (Two Spirals, Half Kernel, Outlier and Corners) datasets. The experimental results showed that the performance and sensitivity to noise of MDFMM-Kn are better than those of all the variants of FMM. Finally, MDFMM-Kn achieved an excellent performance with a simple network structure

CHAPTER 5

CONCLUSION AND FUTURE WORKS

5.1 Summary of research

This chapter delineates the influence and contributions of this study by summarizing the outcomes presented in the previous chapters. Some recommendations for future works are also provided.

Chapter 1 discusses the extant techniques for pattern classification. The features and benefits of the FMM network, which serves as the backbone of this study, are also presented. The scope, problems, objectives, motivations and methodology of this work are eventually described.

Chapter 2 introduces the cutting-edge innovations in FMM and its variants with and without contraction. A detailed analysis of the FMM network is initially presented before highlighting the limitations in its expansion, contraction, overlap test processes, as well as, in the membership function. Several modifications are also presented to address these limitations.

Chapter 3 introduces the MDFMM model along with the four key contributions of this work (i.e. a new expansion technique, overlap test formula, contraction process, and prediction strategy). The efficacy of the MDFMM model is then evaluated by conducting experimental tests. The model demonstrates significant improvements in its classification accuracy across all tested UCI benchmark datasets. However, given its embedded extraction process and contraction method, the MDFMM network has more complex network structure as compared with the original FMM model, thereby producing more hyperboxes number whilst addressing overlaps.

Chapter 4 addresses the third research objective by introducing the MDFMM-Kn model, which is developed by integrating the k-nearest technique into the MDFMM expansion process. The MDFMM-Kn model is considered as a method for addressing the limitations of the MDFMM network. As indicated by the 'Kn' in its name, the MDFMM-Kn model selects 'the k-nearest' hyperbox during training stage instead of relying on a single winning hyperbox in order to reduce the number of generated hyperboxes. The performance of this model is then evaluated using 11 UCI benchmark datasets and 4 other datasets, including two spirals, corners, outlier and kernel. The evaluation results are compared with those for the MDFMM model and other classifiers presented in the literature. The results of the experiment indicate that the MDFMM-Kn model demonstrates a favorable performance across all datasets and generate a smaller number of hyperboxes with comparable or better accuracy rate comparing with other classifiers.

In sum, this research achieves its objectives by developing, implementing and evaluating MDFMM and MDFMM-Kn models. The objectives of this work are outlined in Table 5.1 along with the chapters where these objectives are addressed.

Table 5.1 The research objective achievment

#	Objectives	Chapter achievement
1	To study and analyse the weakness, limitations and drawbacks of the current FMM model and its variants in handling pattern classification problems.	Chapter 2
2	To propose Modern Fuzzy Min-Max (MDFMM) model by introducing a new expansion technique, new overlap test formula, new contraction process, and new prediction strategy for the test stage.	Chapter 3
3	To further enhance the MDFMM model by integrating k-nearest technique to reduce the network structure complexity and improve the model noise tolerance ability.	Chapter 4
4	To evaluate the usefulness of the proposed models in undertaking pattern classification problems using different benchmark, 2D AI, noise and noise free data sets. Furthermore, quantify their performances using statistical indicators, as well as analyse and compare their effectiveness with different classifiers	Chapter 3 and 4

5.2 Contributions of the Research

This study contributes to pattern classification research by identifying the limitations of the FMM network in both the testing and learning phases as summarized in Table 2.3, Section 2.5, Chapter 2. These limitations are addressed in this work by applying two key modifications to the FMM network. The contributions of this research can be summarized in the following paragraphs.

As its first contribution, this work introduces a new expansion technique that prevents generating overlap cases between hyperboxes from different classes, thereby simplifying the training process in the learning stage. As its second contribution, this work constructs a new overlap test formula that can simplify the overlap test process and determine the potential overlap cases (containment overlaps). This formula uses also to terminate the expansion process in case it led to overlap other hyperboxes from different classes. As its third contribution, this work addresses the data distortion problem and simplifies the extant contraction process by introducing a new contraction process that produces highly accurate hyperbox decision boundaries and subsequently promotes the classification accuracy. As its fourth contribution, this research integrates the distance equation into the membership function to create a new prediction strategy that can solve the randomized decision making problem, especially when the fitness value of the input sample remains the same across multiple hyperboxes from different classes.

All these contributions lead to the formation of the MDFMM network, which performance has also been empirically examined in this work by using 11 benchmark datasets, including Iris, Glass, Heart, Wine, Liver, WBC, Ionosphere, PID, Zoo, Thyroid and Seed. To address the inadequacy of measuring accuracy, statistical analysis was conducted to evaluate the classification performance via bootstrap method and k fold cross validation method. The results of the evaluation are then compared with other models from the literature. The empirical findings indicate that the MDFMM outperforms the other classifiers proposed in the literature in various aspects. Therefore, the aforementioned contributions effectively improve the performance of FMM. The proposed MDFMM model not only shows higher stability in its performance compared with FMM but can also generate a highly accurate hyperbox structure and solve the data distortion problem as discussed in Section 3.3.2. Nevertheless, the structure of the MDFMM network is more complex compared with that of the FMM network because of its behaviour to produce pure decision boundaries. However, complex network structures consider as a critical issue in real world applications. Therefore, reducing the complexity of the MDFMM network warrants further attention as discussed in Chapter 4.

As its fifth contribution, this work integrates the k-nearest algorithm into MDFMM to solve the issues related to the complexity of its network structure. The objective of the resulting MDFMM-Kn model is to use a set of k-nearest hyperboxes in the expansion process instead of relying on a single winning hyperbox. The process conditions Eq. 3.1 and Eq. 3.2 should be satisfied by the selected/winning hyperbox. Specifically, the MDFMM-Kn model prevents the creation of small hyperboxes within the area of the winning hyperbox and reduces the number of hyperboxes that are generated by the MDFMM model in the training stage. The performance of this model was evaluated using 11 benchmark datasets and 4 created datasets. Whilst its effectiveness in reducing the complexity and producing efficient classification accuracy was validated by conducting three experiments, bootstrap method and k fold cross validation, which indicate that this model obtains more stable outcomes, achieves better mean values and produces few hyperboxes compared with MDFMM. Both the MDFMM and MDFMM-Kn models were then evaluated against noisy data to understand how these data influence their performance. The evaluation results were then compared with the results for the other classifiers that have been proposed in the literature. Both models demonstrate a better performance compared with the variants of FMM across all noise levels for different benchmark datasets.

5.3 Future Expansion and Recommendations

This study introduces two models, namely, MDFMM and MDFMM-Kn, as well as several modifications to address the pattern classification issues faced by FMM and its variants. Some directions for future research in this area are also presented as follows.

Both the MDFMM and MDFMM-Kn models have some limitations related to the identification of the maximum hyperbox size (θ) and the expansion process. To address these limitations, a re-expansion method can be implemented after the learning process. This method not only grants some flexibility to hyperboxes without the need to integrate the expansion parameter, θ, but also could generate more accurate decision boundaries for the generated hyperboxes and prevents the formation of any overlaps amongst hyperboxes from various classes. These effects have been proven to improve the performances of both the MDFMM and MDFMM-Kn models.

The MDFMM neural network should be integrated with feature selection algorithms as a pre-processing phase in order to solve issues related to big data. Doing so can also lessen the influence of insignificant features and consequently promote the prediction quality of the membership function and improve the decisions related to feature selection.

Most ANNs, including MDFMM and MDFMM-Kn, suffered from inability to explain its predication (black box). An extraction step can be set up to address this problem by generating compact rules via a genetic algorithm.

The pattern classification conducted in this study adheres to the FMM network and its variants, including MFMM-GA, MFMMN, enhanced FMM network and FMM clustering. Future studies may consider applying the modifications introduced in this study, including the new contraction and expansion methods, Kn and new overlap test formula, into other FMM models.

The proposed MDFMM and MDFMM-Kn models have demonstrated high accuracy and obtained excellent pattern classification results. However, these models show some instability in the learning phase because of the changes in the training sequence that can influence their intricacy and accuracy. Applying the ensemble method can support the predictions derived from multiple classifiers, obtain an integrated output and manipulate the decisions of a multi-classifier framework, thereby enhancing the robustness of both MDFMM and MDFMM-Kn.

REFERENCES

Alan, M. (1950). Turing. Computing machinery and intelligence. Mind, 59, 433-460.

Andrzej Bargiela 1, M. T., Giovanna Castellano 3, Annamaria Fanelli 3. (2004). Adaption of exclusion/inclusion hyperboxes. *CEUR Workshop Proceedings*, 1542, 33-36.

Augasta, M. G., & Kathirvalavakumar, T. (2013). Pruning algorithms of neural networks—a comparative study. *Central European Journal of Computer Science*, 3(3), 105-115.

Azad, C., & Jha, V. K. (2017). Fuzzy min–max neural network and particle swarm optimization based intrusion detection system. Microsystem Technologies, 23(4), 907-918.

Azad, C., Mehta, A. K., & Jha, V. K. (2018). Improved Data Classification using Fuzzy Euclidean Hyperbox Classifier. *Paper presented at the 2018 International Conference on Smart Computing and Electronic Enterprise*.

Bache, K., & Lichman, M. (2013). UCI machine learning repository.

Barde, N. C., & Patole, M. (2016). Classification and Forecasting of Weather using ANN, k-NN and Naïve Bayes Algorithms. *International Journal of Science and Research (IJSR)*, 5(2), 1740-1742.

Bargiela, A., Pedrycz, W., & Tanaka, M. (2004). An inclusion / exclusion fuzzy hyperbox classifier. *KES Journal*, 8, 91-98.

Begum, S. A., & Devi, O. M. (2011). Fuzzy Algorithms for Pattern Recognition in Medical Diagnosis. Assam University *Journal of Science and Technology: Physical Sciences and Technology*, 7(2), 1-12.

Carpenter, G. A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer vision, graphics, and image processing*, 37(1), 54-115.

Carpenter, G. A., Grossberg, S., & Reynolds, J. H. (1991a). ARTMAP: supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *1991 IEEE Conference on Neural Networks for Ocean Engineering*, 4, 341-342.

Carpenter, G. A., Grossberg, S., & Rosen, D. B. (1991b). Fuzzy ART: an adaptive resonance algorithm for rapid, stable classification of analog patterns. *1991 Seattle International Joint Conference on Neural Networks*, 2(6), 411-416.

Davtalab, R., Dezfoulian, M. H., & Mansoorizadeh, M. (2014). Multi-level fuzzy min-max neural network classifier. *IEEE Transactions on Neural Networks and Learning Systems*, 25(3), 470-482.

Donglikar, N. V., & Waghmare, J. M. (2017). An enhanced general fuzzy min-max neural network for classification. *Paper presented at the 2017 International Conference on Intelligent Computing and Control Systems*. 757-764

Duan, Y., Cui, B., & Xu, X. (2007). State space partition for reinforcement learning based on fuzzy min-max neural network. *Paper presented at the 2007 International Symposium on Neural Networks*. 160-169

Fallahnezhad, M., Moradi, M. H., & Zaferanlouei, S. (2011). A Hybrid Higher Order Neural Classifier for handling classification problems. *Expert Systems with Applications*, 38(1), 386-393.

Gabrys, B., & Bargiela, A. (2000). General fuzzy min-max neural network for clustering and classification. *IEEE Transactions on Neural Networks*, 11, 769-783.

Gao, Z., Cecati, C., & Ding, S. X. (2015). A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part I: Fault Diagnosis With Model-Based and Signal-Based Approaches. *IEEE Transactions on Industrial Electronics*, 62(6), 3757-3767.

Genaro, N., Torija, A., Ramos, A., Requena, I., Ruiz, D. P., & Zamorano, M. (2009). Modeling Environmental Noise Using Artificial Neural Networks. *Paper presented at the 2009 Ninth International Conference on Intelligent Systems Design and Applications*. 215-219

Hau, C. C. (2015) Handbook of Pattern Recognition and Computer Vision: World Scientific. 1-1044.

Jain, A. K., Duin, R. P. W., & Jianchang, M. (2000). Statistical pattern recognition: a review. I*EEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), 4-37.

Jang, J. S. R. (1993). ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3), 665-685.

Jawarkar, N., Holambe, R., & Basu, T. (2011). Use of fuzzy min-max neural network for speaker identification. *Paper presented at the 2011 International Conference on Recent Trends in Information Technology*.178-182

Kamruzzaman, S. M., & Jehad Sarkar, A. M. (2011). A new data mining scheme using artificial neural networks. *Sensors*, 11(5), 4622-4647.

Keller, J. M., & Hunt, D. J. (1985). Incorporating fuzzy membership functions into the perceptron algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 7 (6), 693-699.

Kim, H.-J., Ryu, T.-W., Lee, J., & Yang, H.-S. (2006). Face detection using an adaptive skin-color filter and FMM neural networks. *Paper presented at the 2006 9th Pacific Rim international conference on Artificial intelligence*.4099, 1171-1175.

Kim, H. J., Ryu, T. W., Nguyen, T. T., Lim, J. S., & Gupta, S. (2004). A Weighted Fuzzy Min-Max Neural Network for Pattern Classification and Feature Extraction. Paper presented at the *2004 Computational Science and Its Applications*, 3612,1178-1181

Kim Hua, T., Chee Peng, L., Hooi Shen, K., & Platts, K. (2004). Application of the fuzzy ARTMAP neural network to classification of manufacturing technology projects. *WSEAS Transactions on Computers*, 3(3), 636-641.

Kuan, M. M., Lim, C. P., & Harrison, R. F. (2003). On operating strategies of the fuzzy ARTMAP neural network: A comparative study. *International Journal of Computational Intelligence and Applications*, 3(01), 23-43.

Liu, H., Yin, J., Luo, X., & Zhang, S. (2018). Foreword to the special issue on recent advances on pattern recognition and artificial intelligence. *Neural Computing and Applications*, 29(1), 1-2.

Liu, J., Yu, Z., & Ma, D. (2012). An adaptive fuzzy min-max neural network classifier based on principle component analysis and adaptive genetic algorithm. *Mathematical Problems in Engineering*, 2012.

Mermillod, M., Bugaiska, A., & Bonin, P. (2013). The stability-plasticity dilemma: investigating the continuum from catastrophic forgetting to age-limited learning effects. *Frontiers in Psychology*, 4, 1-3.

Mithe, R., Indalkar, S., & Divekar, N. (2014). Optical character recognition. *International Journal of Emerging Technology and Advanced Enginnering*, 4(5), 219-223.

Mohammed, M. F., & Lim, C. P. (2015). An enhanced fuzzy min-max neural network for pattern classification. *IEEE Transactions on Neural Networks and Learning Systems*, 26(3), 417-429.

Mohammed, M. F., & Lim, C. P. (2017a). Improving the Fuzzy Min-Max neural network with a K-nearest hyperbox expansion rule for pattern classification. *Applied Soft Computing*, 52, 135-145.

Mohammed, M. F., & Lim, C. P. (2017b). A new hyperbox selection rule and a pruning strategy for the enhanced fuzzy min–max neural network. Neural networks, 86, 69-79.

Nandedkar, A. V., & Biswas, P. K. (2007a). A fuzzy min-max neural network classifier with compensatory neuron architecture. *IEEE Transactions on Neural Networks*, 18, 42-54.

Nandedkar, A. V., & Biswas, P. K. (2007b). A General Reflex Fuzzy Min-Max Neural *Network. Engineering Letters*, 14, 195-205.

Naz, S., Hayat, K., Imran Razzak, M., Waqas Anwar, M., Madani, S. A., & Khan, S. U. (2014). The optical character recognition of Urdu-like cursive scripts. Pattern recognition, *Pattern Recognition* ,47(3), 1229-1248.

Oludolapo, O. A., Jimoh, A. A., & Kholopane, P. A. (2012). Comparing performance of MLP and RBF neural network models for predicting South Africa's energy consumption. *Journal of Energy in Southern Africa*, 23(3), 40-46.

Pal, S. K., & Mitra, S. (1992). Multilayer perceptron, fuzzy sets, and classification. *IEEE Transactions on Neural Networks*, 3(5), 683-697.

Pawar, D. (2015). Fuzzy min-max neural network with compensatory neuron architecture for invariant object recognition. *Paper presented at the 2015 International Conference on Computer, Communication and Control* (IC4). 1-5

Qu, Z. (2009). Application of Data Mining in Classification Analysis of Safety Accidents Based on Alternate Covering Neural Network. *Paper presented at the 2009 International Conference on Innovation Management*. 47-50.

Quteishat, A., Al-Batah, M., Al-Mofleh, A., & Alnabelsi, S. H. (2013). Cervical cancer diagnostic system using adaptive fuzzy moving k-means algorithm and fuzzy min-max neural network. *Journal of Theoretical and Applied Information Technology*, 57(1), 48-53.

Quteishat, A., & Lim, C. (2007). A modified fuzzy min-max neural network and its application to fault classification. *Soft computing in industrial applications*,39 ,179-188.

Quteishat, A., & Lim, C. P. (2008). A modified fuzzy min-max neural network with rule extraction and its application to fault detection and classification. *Applied Soft Computing*, 8(2), 985-995.

Quteishat, A., Lim, C. P., & Tan, K. S. (2010). A modified fuzzy min–max neural network with a genetic-algorithm-based rule extractor for pattern classification. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 40(3), 641-650.

Raj, M., Swaminarayan, P., Saini, J., & Parmar, D. (2015). Applications of Pattern Recognition Algorithms in Agriculture: A Review. *International Journal of Advanced Networking and Applications*, 6(5), 2495-2502.

Reaz, M. B., Hussain, M., & Mohd-Yasin, F. (2006). Techniques of EMG signal analysis: detection, processing, classification and applications. *Biological procedures online*, 8(1), 11-35.

Saksamudre, S. K., Shrishrimal, P., & Deshmukh, R. (2015). A review on different approaches for speech recognition system. *International Journal of Computer Applications*, 115(22), 23-28.

Seera, M., & Lim, C. P. (2014). Online motor fault detection and diagnosis using a hybrid FMM-CART model. *IEEE Transactions on Neural Networks and Learning Systems*, 25(4), 806-812.

Seera, M., Lim, C. P., Ishak, D., & Singh, H. (2013). Application of the fuzzy min–max neural network to fault detection and diagnosis of induction motors. *Neural Computing and Applications*, 23(1), 191-200.

Shinde, S., & Kulkarni, U. (2016). Extracting classification rules from modified fuzzy min–max neural network for data with mixed attributes. *Applied Soft Computing*, 40, 364-378.

Shrivastava, J. N., & Bindu, M. H. (2014). E-mail Spam Filtering Using Adaptive Genetic Algorithm. *International Journal of Intelligent Systems and Applications*, 6(2), 54-60.

Simpson, P. K. (1992). Fuzzy min-max neural networks. I. Classification. *IEEE Transactions on Neural Networks*, 3(5), 776-786.

Simpson, P. K. (1993). Fuzzy min-max neural networks-part 2: Clustering. IEEE Transactions on Fuzzy systems, 1(1), 32-45.

Smola, A., & Vishwanathan, S. (2008). Introduction to machine learning. Cambridge University, UK, 32, 1-34.

Tkáč, M., & Verner, R. (2016). Artificial neural networks in business: Two decades of research. *Applied Soft Computing*, 38, 788-804.

Umoh, U. A. U., A.A.; James, G.G.; oton, U.U.; Udoudo, J.J.; B.Eng. (2012). Design of Pattern Recognition System for the Diagnosis of Gonorrhea Disease. *International Journal of Scientific & Technology Research*, 1(5), 74-79.

Unar, J., Seng, W. C., & Abbasi, A. (2014). A review of biometric technology along with trends and prospects. *Pattern recognition*, 47(8), 2673-2688.

Upasani, N., & Om, H. (2018). Optimized fuzzy min-max neural network: an efficient approach for supervised outlier detection. *Neural Network World*, 28(4), 285-303.

Viharos, Z. J., & Kis, K. B. (2015). Survey on neuro-fuzzy systems and their applications in technical diagnostics and measurement. *Measurement*, 67, 126-136.

Yilmaz, I., Erik, N. Y., & Kaynar, O. (2010). Different types of learning algorithms of artificial neural network (ANN) models for prediction of gross calorific value (GCV) of coals. *Scientific Research and Essays*, 5, 2242-2249.

Yilmaz, I., & Kaynar, O. (2011). Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils. *Expert Systems with Applications*, 38, 5958-5966.

Zhang, H., Liu, J., Ma, D., & Wang, Z. (2011). Data-core-based fuzzy min–max neural network for pattern classification. *IEEE Transactions on Neural Networks*, 22(12), 2339-2352.

APPENDIX A
LIST OF PUBLICATION

Published Journal Papers:

Sayaydeh, O. N., M.F. Mohammad and C.P. Lim (2018). "Survey of Fuzzy Min–Max Neural Network for Pattern Classification Variants and Applications." IEEE Transactions on Fuzzy Systems 27: 635-645.

Journal Impact Factor: 8.415 (Q1)

Sayaydeh, O. N.and A. Shamaileh (2018). "Fuzzy Min Max Neural Network for pattern classification: An overview of complexity problem." International Journal of Information Technology and Language Studies 2: 110-117.

Published Conference Papers:

Sayaydeh, O. N.and M.F. Mohammad (2019). " Diagnosis of The Parkinson Disease Using Enhanced Fuzzy Min-Max Neural Network and OneR Attribute Evaluation Method 2019 International Conference on Advanced Science and Engineering (ICOASE), IEEE.