Fuzzy Based Approach for Modelling Uncertainty in Classification for a Computer Aided Detection

Rahil Hosseini

Submitted in partial fulfilment of the requirements of Kingston University for the degree of

Doctor of Philosophy

Abstract

A computerized image analysis technology suffers from imperfection, imprecision and vagueness of the input data and its propagation to all individual components of the technology including image enhancement, segmentation and object classification. Furthermore, a computerized medical image analysis system (CMIAS) deals with another source of ambiguity that is inherent in the image-based practice of medicine and intuitive knowledge of experts. Therefore, a CMIAS such as computer aided detection (CAD) technologies implicitly suffer from uncertainty and vagueness both from image analysis techniques and medical diagnosis. Although several technology-oriented studies have been reported for CAD, no attempt has been made to address, model and overcome these types of uncertainty in the design of the CAD. However, uncertainty issues directly affect the accuracy of the system. This study addresses the main sources of the uncertainty in a CAD system. While uncertainty outcomes are latent in the input of a classifier, the aim is to model them in the classification for a CAD application. For this, this research takes advantages of type-2 fuzzy logic (T2FL). Integrating a T2FL model for object classification in CAD architecture allows us to model uncertainty issues. For this, an automatic approach models uncertainty in training dataset using membership function of a type-2 fuzzy set. This approach was applied to the candidate nodule classification problem in a lung CAD application. The ROC (receiver operating characteristic) analysis of the classifier results (with an average accuracy 95% (area under the ROC curve) for nodule classification) reveals that the T2FL is more capable of capturing the uncertainty in the model and achieving better performance results compared to type-1 fuzzy logic counterpart. Furthermore, the research introduces the idea of uncertain rule-based pattern classification in environments which exhibit a lack of expert knowledge and with an imperfect training dataset. An automatic approach for rule extraction is presented which takes advantages of genetic algorithm for learning rule set of an T2FL system from training samples. The proposed approach was applied to the popular Wisconsin breast cancer diagnosis (WBCD) database. Analysis of the performance results reveals that this approach is competitive with the best results of other proposed fuzzy classification methods to date in terms of trade-off between accuracy and interpretability, with an average accuracy of 96.6 % for the breast cancer diagnosis problem.

This study introduces the concept of uncertainty in a CAD application. This is a first attempt toward modelling uncertainty issues in classification component for a CAD. The main contribution is automatically modelling uncertainties using membership functions and a rule set of a type-2 fuzzy logic. The performance evaluation on two different CAD classification problems (1) nodule classification in a lung CAD and (2) the WBCD diagnosis problem using Mammography CAD reveals the superiority of the T2FLS classifier for managing high levels of uncertainty compared to the T1FLS counterpart and providing classification that is more accurate. This approach is significant from two major aspects (1) clinical view: by producing more accurate results for diagnosis problems which can save more human lives, (2) technical view: modelling uncertainties in the design of a classifier using automatically presented approach for IT2FLS membership and rules generation. This is critical for multi-dimensional classification problems with large number of inputs and lack of expert knowledge as is the case for most of medical diagnosis problems.

Contents

Abstract 2

Contents 4

Acknowledgments 8

Publications 11

Glossary of Terms 12

List of Figures 15

List of Tables 19

Chapter 1: Introduction 20

1.1 Computerized medical image analysis 20

1.2 Computer aided detection technology 23

1.3 ROC analysis for evaluation of the classification performance 28

1.4 Aim and objectives 33

1.5 Research motivation 33

1.6 Methodology of the research 34

1.7 Research contribution 35

1.8 Structure of the dissertation 36

Part I 38

Chapter 2: Uncertainty Challenges in Classification for Medical Image Analysis Applications 39

2.1 Introduction 39

2.2 Uncertainty in input data and noisy images 40

2.3 Uncertainty in mathematical models for measuring the complex features 42

2.4 Inter- and intra- observer variability 42

2.5 Uncertainty in words perception from different experts point of view 45

2.6 Uncertainty in non-stationary features 46

2.7 Uncertainty in image processing techniques....47
2.8 Uncertainty in object classification....49
2.9 Probability theory and fuzzy set theory for modeling uncertainty....50
2.10 Summary....51
Chapter 3: Fuzzy Logic for Rule-based Pattern Classification....53
3.1 Introduction....53
3.2 Fuzzy logic....55
3.3 Fuzzifier....57
3.4 Different types of rules in a FL....57
3.5 Fuzzy inference engine....60
3.6 Fuzzy set operations....62
3.7 Example of fuzzy logic inference....64
3.8 Defuzzification....64
3.9 Remaining uncertainty issues in an ordinary fuzzy logic....66
3.10 Summary....68
Chapter 4: Type-2 Fuzzy Logic for Uncertain Rule based Classification....70
4.1 Introduction....70
4.2 An overview of advances of type-2 fuzzy sets in medical imaging applications 72
4.3 Type-2 fuzzy set concepts....74
4.3.1 Footprint of uncertainty....75
4.4 Set theoretic operations on type-2 fuzzy sets....80
4.5 Type-2 fuzzy sets relation and implication....84
4.6 Centroid of a type-2 fuzzy set....86
4.7 Type-2 fuzzy logic....86
4.7.1 Fuzzifier....87
4.7.2 Rules and inference engine....87
4.7.3 Output Producer....90
4.8 Summary....93
Part II....94

Chapter 5: An Approach for Modelling Uncertainty through Gaussian IT2FMFs for Nodule Classification in a Lung CAD Application 95

5.1 Introduction 95

5.2 Learning Gaussian interval type-2 fuzzy membership function 98

5.2.1 Fuzzy partitioning of input sets for generating type-1 membership functions 99

5.2.2 Modeling the footprint of uncertainty 99

5.3 A genetic algorithm for parameter estimation of the FOU 106

5.3.1 GA initialization 107

5.3.2 Cross-validation 107

5.3.3 Fitness evaluation 108

5.3.4 Comparison of the GA IT2FMF learning approach with other GA-based methods 108

5.4 Nodule classification problem for a lung CAD application 111

5.4.1 Overview of the lung CAD system 113

5.5 Experimental results and performance evaluation 115

5.6 Summary 125

Chapter 6: A Genetic Type-2 Fuzzy Approach for Rule Extraction for Uncertain Pattern Classification Problems 127

6.1 Introduction 128

6.2 Interval type-2 fuzzy membership function generation 129

6.3 Rule extraction for interval type-2 fuzzy pattern classification 132

6.3.1 Rule generation for IT2FL 133

6.3.2 Fuzzy reasoning 138

6.4 An interval type-2 fuzzy GA approach for uncertain linguistic rules learning.. 140

6.4.1 The GA overfitting problem 145

6.5 Breast cancer diagnostic problem 146

6.6 Review of the classification methods for the WBCD classification problem.... 148

6.6.1 Rule extraction methods based on deterministic approaches 149

6.6.2 Rule extraction methods based on fuzzy techniques 150

6.7 Experimental results and performance evaluation 152
6.7.1 Estimation of the FOU parameters 153
6.7.2 The GA IT2FL rule learning algorithm 155
6.7.4 Performance comparison with the other fuzzy rule extraction methods 167
6.8 Summary 168
Chapter 7: Conclusion and Future Works 171
7.1 Conclusion 171
7.2 novelty 173
7.3 Contribution 174
7.3.1 Modeling uncertainty associated with the input of classification 174
7.3.2 Extracting rules for uncertain classification problems 175
7.4 Benefits 176
7.5 Future Works 177
Bibliography 179
Appendix A 196

Acknowledgments

First, I would like to thank Prof. Jamshid Dehmeshki, my primary supervisor for his continues commitment, guidance at every stage of my research. I am grateful to him for his constructive criticism, extensive consultation, and long discussions that enabled me to complete this dissertation.

I would like to thank Prof. Tim Ellis, for his support, time and assistance in reviewing my work and providing useful suggestions and comments that helped me immensely in improving the quality of my publications and this dissertation.

I would like to also thank Dr. Sarah Barman for giving me the opportunity to get involved in her project to better understand inter- and intra- uncertainties in CAD applications and her valuable comments and suggestions to improve my manuscripts and my thesis.

I would like to thank Prof. S. D. Qanadli from Lausanne Hospital, Switzerland for providing me lung CT scan datasets and his clinical views and comments on my work.

I would like to thank all reviewers of my papers whose comments helped me to complete my work. I would like to acknowledge Prof. Oscar Castillo from Tijuana Institute of Technology, Mexico and his research group for providing me the framework of type-2 fuzzy logic toolbox in Matlab to implement and test the proposed approach.
I would like to thank Mahdi Mazinani for all his technical helps and emotional supports during some difficult times throughout my study. I am also extremely grateful to his assistance, which help me to stay focused on my study.

I was fortunate enough to meet some very sincere and kind individuals throughout my journey, many of which helped me with my transition in UK. I would like to thank Mr. Mehdi Bashiri for all his invaluable support and guidance.

I would like to express my deepest and emotional appreciation to my family. To my mother, Mina Mahjoobi, who has always been there for me with a ready ear, prayers and unconditional love whenever I need it most and specially, for her great help and support in the last year of my study. My father, Mr. Ali Akbar Hosseini, for his advice, encouragement, dreams, and brightening the research life for me. My brother, Reza Hosseini, who has provided all the support and help that he could offer. My eternal love and gratitude goes to my lovely daughter, Manisa whose presence in my life was my great motivation to complete my PhD, and also my husband, Mahdi Mazinani, without him and his continues support and unconditional love I would never get to this stage of my life. Please accept this thesis in appreciation of all you have done for me.

To my loving daughter Manisa, husband and parents

Publications

Journal Paper:

1. R. Hosseini, S. D. Qanadli, S. Barman, M. Mazinani, T.Ellis, and J. Dehmeshki, An Automatic Approach for Learning and Tuning Gaussian Interval Type-2 Fuzzy Membership Functions Applied to Lung CAD Classification System, *IEEE Transactions on Fuzzy Systems*, vol. 20, No. 2, April (2012), pp. 224-234.

Conference Paper:

1. R. Hosseini, T. Ellis, S. A. Barman, M Mazinani, S.D. Qanadli, and J. Dehmeshki, "A Genetic Fuzzy Approach for Rule Extraction for Rule-Based Classification with Application to Medical Diagnosis", Int. Conference of Knowledge Discovery in Health Care and Medicine, Athens, Greece, September (2011), pp. 2-15.
2. R. Hosseini, J. Dehmeshki, S. A. Barman, M. Mazinani, S.D. Qanadli, "A Fuzzy Logic System for Classification of the Lung Nodule in Digital Images in Computer Aided Detection", In Proc. of Int. Conf. of Digital Society (ICDS), IEEE, September (2010), pp. 255- 259.
3. R. Hosseini, J. Dehmeshki, S. A. Barman, M. Mazinani and S. Qanadli, "A Genetic Type-2 Fuzzy Logic System for Pattern Recognition in Computer Aided Detection Systems", in IEEE Int. Conf. on Fuzzy Systems (FUZZ), Barcelona, Spain, July (2010), pp. 1-7.
4. R. Hosseini, J. Dehmeshki, S. A. Barman, M. Mazinani, S. D. Qanadli, "Modelling Uncertainty in Classification Design of Computer Aided Detection", SPIE Medical Imaging, Computer-Aided Diagnosis, vol. 7624, San Diego, California, USA, February (2010), pp. 76242V.

Glossary of Terms

Acronym	Meaning
AC	Ant Colony
AUC	Area Under the ROC Curve
ASR	Age-Standardized cancer Rates
BC	Bland Chromatin
BIO-RE	Binarized Input-Output Rule Extraction
BN	Bare Nuclei
CAD	Computer Aide Detection
CADx	Computer Aide Diagnosis
CAM	Computer Aided Measurement
CMIAS	Computerized Medical Image Analysis System
CI	Confidence Interval
CT	Computed Tomography
CT	Clump Thickness
CV	Cross-Validation
2D	2 Dimensional
3D	3 Dimensional
DT based	Decision Tree-based
FCM	Fuzzy C-mean Clustering
FGA	Fuzzy Genetic Architecture
FL	Fuzzy Logic
FLS	Fuzzy Logic System
FOU	Footprint of Uncertainty
FP	False Positive
fp rate	False Positive Rate
FN	False Negative
FR	Fail to Reject null hypothesis
GA	Genetic Algorithm
GA IT2FLS	GA for Learning the Membership Function of an IT2FLS

HGA	Hierarchical Genetic Algorithm
HU	Hounsfield Unit
IT2FL	Interval Type-2 Fuzzy Logic
IT2FLS	Interval Type-2 Fuzzy Logic System
IT2FMF	Interval Type-2 Fuzzy Membership Function
IT2FS	Interval Type-2 Fuzzy Set
LIDC	The Lung Image Database Consortium
LOOCV	Leave-One-Out Cross-Validation
MA	Marginal Adhesion
MF	Membership Function
M	Mitoses
MRI	Magnetic Resonance Imaging
NN	Normal Nucleoli
NP	Nodule Possibility
Partial-RE	Partial Rule Extraction
PE	Pulmonary Embolism
PS	Particle Swarm
PVE	Partial Volume Effect
RE	Rule Extraction
ROC	Receiver Operating Characteristic
ROI	Region of Interest
SANIF	Self-Adaptive Neuro-fuzzy Inference System
SEC	Single Epithelial Cell Size
STD	Standard Deviation
SVM	Support Vector Machines
T1FMF	Type-1 Fuzzy Membership Function
T2FL	Type-2 Fuzzy Logic
T2FLS	Type-2 Fuzzy Logic System
T2FNN	Type-2 Fuzzy Neural Network
T1FL	Type-1 Fuzzy Logic
T2FS	Type-2 Fuzzy Set
T2FMF	Type-2 Fuzzy Membership Function
TP	True Positive

tp rate	True Positive rate
TN	True Negative
TKS	Takagi-Sugeno
T-norm	Triangular Norm
UC	Uniformity of Cell Size
UCS	Uniformity of Cell Shape
UK	United Kingdom
WBCD	Wisconsin Breast Cancer Diagnosis

List of Figures

Figure 1.1: A representation of the input data, processes and output of a medical image analysis system..22
Figure1.2: A typical components of a CAD system 24
Figure1.3: Sample of a nodule presented in a CT image slice and manually annotated by a radiologist..27
Figure1.4: Confusion matrix..29
Figure1.5: ROC curve coordinates..30
Figure 1.6: An example of an ROC curve..32
Figure 1.7: Methodology of the research...35
Figure 2.1: Examples of a nodule in lung CT scan; (a) Noisy image (b) Smoothed image 41
Figure 2.2: Different mathematical model results for measuring sphericity 42
Figure 2.3: Illustration of variation between radiologists for nodule identification in the LIDC study (van Ginneken et al., 2010); Ri (i=1, ..., 4) represents one of four radiologists..43
Figure 2.4: illustration of variation between radiologists for nodule identification between three radiologists in (Dehmeshki *et al.*, 2003)..44
Figure 2.5: A trapezoidal membership function for "critically ill" category (Wills *et al.*, 2003) 46
Figure 2.6: Confusion in symptoms duration of the influenza and scarlet fever (Innocent and John, 2004)..47
Figure 2.7: Modeling of three classes for a lung phantom in (Dehmeshki *et al.*, 2008).48
Figure 2.8: The effect of a seed point placed on a nodule attached to blood vessel; (a) Original image (b) A random seed is selected (shown with plus), (c) An optimum seed is selected (shown with a circle) (Dehmeshki *et al.*, 2008) 48
Figure 2.9: A candidate nodule in three adjacent slices (Gomathi and Thangaraj, 2010) 49
Figure 2.10: An example of temperature fuzzy set...51

Figure 3.1: Membership function corresponding to linguistic terms 56
Figure 3.2: Type-1 Fuzzy Logic (FL) 57
Figure 3.3: Type-1 fuzzy inference engine 61
Figure 3.4: (a) Fuzzy sets A and B (b) Union ($A \cup B$) (c) Intersection $A \cap B$ 63
Figure 3.5: Rules and antecedent fuzzy sets 64
Figure 3.6: Inference engine in a type-1 fuzzy logic system 65
Figure 3.7: Footprint of uncertainty (FOU) in a type-2 fuzzy set 67
Figure 3.8: Modeling uncertainty in a rule consequent 68
Figure 4.1: Pictorial view of extension of type-2 fuzzy sets theory and applications according to (John and Coupland, 2006) 71
Figure 4.2: Membership function of a type-2 fuzzy set at $x = x'$ 74
Figure 4.3: Vertical slices view of a type-2 fuzzy set 75
Figure 4.4: Interval type-2 secondary membership function 76
Figure 4.5: Gaussian primary membership function with fixed standard deviation and uncertain mean 78
Figure 4.6: Lower bound and upper bound of a Gaussian primary membership function 79
Figure 4.7: Union and Intersection of type-2 fuzzy sets; (a) two type-2 fuzzy sets, (b) union (c) intersection 82
Figure 4.8: Fuzzy inference engine 85
Figure 4.9: Type-2 fuzzy logic architecture 87
Figure 4.10: Rules in an IT2FL and corresponding antecedent and consequent type-2 fuzzy sets 88
Figure 4.11: Inference engine of a type-2 fuzzy logic system 89
Figure 4.12: A firing interval of output of a rule of an IT2FLS for a given input $x = x1,\ x2$ 90
Figure 4.13: Fired consequent sets for three consequent of three rules 93
Figure 5.1: Typical structure of the chromosome for a feature 100
Figure 5.2: Type-2 fuzzy Gaussian membership function of a typical feature; (a) A linguistic term, (b) Three linguistic terms 101
Figure 5.3: An example of a Gaussian membership function with uncertain mean (a) Original membership function with m, s=(8, 2); (b) Lower bound membership

function with m = 6.2; (c) Upper bound membership function with m = 9.8; (d) The shaded area shows the FOU 103
Figure 5.4: An example of a Gaussian membership function with uncertain STD; (a) Lower bound membership function with s = 1; (b) Upper bound membership function with s = 4; (c) The shaded area shows the FOU 105
Figure 5.5: Flowchart of the GA learning MFs method for an IT2FLS 106
Figure 5.6: The structure of the chromosome 107
Figure 5.7: Worldwide age-standardized cancer rate for both male and female per 100,000, (Ferlay *et al.*, 2011) 111
Figure 5.8: The distribution of the incidence and mortality in the world compare to the United Kingdom, (Ferlay *et al.*, 2011) 112
Figure 5.9: Sample of a nodule presented in a CT image slice and manually annotated by a radiologist 114
Figure 5.10: Major components of a lung CAD system: (a) Original image and (b) Pre-processing of the image to smooth the lung area; (c) Lung segmentation, applied to the smoothed image; (d) Object detection applied to the lung region 115
Figure 5.11: Implemented IT2FLS for classification of candidate nodules in a lung CAD (a) Interval type-2 fuzzy sphericity membership function (b) Interval type-2 fuzzy membership functions for 7 features of candidate lung nodules (c) IT2 Fuzzy inference system for candidate nodule classification (d) Rule set of the IT2FLS119
Figure 5.12: The ROC curve of the T1FLS compared to the proposed IT2FLSs 123
Figure 5.13: Examples of the T1FLS and the IT2FLS nodule classification: detected nodules are circled in red (a) True Positive (TP) detected by the T1FLS and the IT2FLS (nodules), (b) False Positive (FP) detected by the T1FLS and the IT2FLS, (c) FP detected by the T1FLS but not the IT2FLS (d) TP detected by the IT2FLS but not by the T1FLS 124
Figure 6.1: Type-2 fuzzy partitioning of the pattern space 130
Figure 6.2: Genetic algorithm for learning IT2FL rules 141
Figure 6.3: Chromosome structure with numerical coding 142
Figure 6.4: An example of crossover operation on two individuals at crossover point, *p* 144
Figure 6.5: An example of mutation operation on an individual at random mutation point * 145

Figure 6.6: Incidence and mortality age-specific breast cancer rate for women per 100,000 in the UK in 2008 (Ferlay *et al.*, 2011) 147

Figure 6.7: Two view projections of the breast through mammographic screening from (Sampat, Markey and Bovik, 2005) 148

Figure 6.8: Membership function of IT2FLS_1, $k_m = 0.1, k_v = 0.8$, for nine features of the WBCD 158

Figure 6.9: ROC curve of the IT2FLS_1 for the WBCD dataset 159

Figure 6.10: Membership function of the IT2FL_2, $k_m = 0.1, k_v = 0.9$, for nine features of the WBCD dataset 160

Figure 6.11: ROC curve of the IT2FLS_2 for the WBCD dataset 161

Figure 6.12: Membership function of IT2FLS_3, $k_m = 0.2, k_v = 0.7$, for nine features of the WBCD dataset 163

Figure 6.13: ROC curve of the IT2FLS_3 for the WBCD dataset 163

Figure 6.14: Membership function of IT2FLS_4, $k_m = 0.3, k_v = 0.7$, for nine features of the WBCD 165

Figure 6.15: ROC curve of the IT2FLS_4 for the WBCD dataset 165

Figure A.1: IT2FLS software architecture 197

Figure A.2: Components of automatic IT2FLS generation software and relation between them 197

List of Tables

Table 1.1 Different thresholds for drawing an ROC curve............32
Table 5.1 Comparison of the chromosome structure in the GA IT2FMF learning and tuning methods............110
Table 5.2 Lung nodule candidate features............117
Table 5.3 List of nodule classification rules............117
Table 5.4 Analysis of the results of the T1FLS and the IT2FLSs............120
Table 5.5 Two sample left-tailed t-test analysis of the results of the T1FLS and proposed IT2FLSs for lung nodule dataset............121
Table 6.1 Estimation of the FOU parameters............154
Table 6.2 Relation between different mutation and crossover rates and classification performance............156
Table 6.3 Relation between different values of the classification weights and classification performance............156
Table 6.4 Relation between different values of the learning weights and the classification performance............156
Table 6.5 Extracted rules of IT2FLS_1 for the WBCD dataset............159
Table 6.6 Extracted rules for IT2FLS_2 for the WBCD dataset............161
Table 6.7 Extracted rules of IT2FLS_3 for the WBCD dataset............162
Table 6.8 extracted rules of IT2FLS_4 for the WBCD dataset............164
Table 6.9 Summary of the performance results of the IT2FLSs............166
Table 6.10 Comparison of the performance of the fuzzy rule extraction methods for the WBCD dataset............167
Table 6.11 Rules generated for classifying benign objects............169
Table 6.12 Rules generated for classifying malignant objects............170

Chapter 1: Introduction

This chapter presents an introduction to medical image analysis technologies such as computer aided detection (CAD) and the uncertainty issues associated with CAD applications. It provides an overview of the major components of a CAD application, particularly, the role and importance of a classification in this technology, and current challenges. It then follows by a brief explanation of common performance evaluation method for classification. Lastly, it presents the research motivation, contribution to the field and the structure of thesis.

1.1 Computerized medical image analysis

Advances in computerized medical image analysis system (CMIAS) technologies provide helpful tools for diagnosis and treatment process. Multidimensional digital image processing techniques present hidden characteristics of the images which are sometimes difficult to see with a naked eye. This is the reason that the application of CMIAS have emerged in recent years. Nowadays, medical image analysis technologies such as computer-aided detection (CAD), diagnosis (CADx) and measurement (CAM) are playing a vital role as a second reader (a person who can read and analyse images) in diagnostic radiology. Furthermore, intelligent systems embedded in a CAD for extracting knowledge from imprecise and noisy information of the processed images can extremely improve image understanding and the analysis process.

Various techniques are employed in a computerized medical imaging system such as image processing, computer vision, image understanding and analysis. Image processing usually includes procedures in which both input and output are images (Gonzalez, Woods and Eddins, 2004). However, in computer vision, the ultimate goal is to emulate the human vision using learning and inference techniques based on visual input which is a branch of artificial intelligence. The image analysis (or understanding) field is between image processing and computer vision and is the focus of this study. Referring to the Gonzalez book (Gonzalez, Woods and Eddins, 2004), there is no clear boundary between the continuum of the field of image processing toward computer vision and image analysis. Gonzalez defines a paradigm which considers three levels of process: low-level, mid-level and high-level. In low-level processes both input and output are images. This process includes initial operations on the image for improving the quality of the image, like noise reduction and contrast enhancement. Mid-level processes involve tasks such as segmentation (partitioning an image into regions or objects) and classification (recognition) of individual objects. Mid-level process inputs are images but their outputs are attributes extracted from those images (e.g. edges, contours, and identity of individual objects). Lastly high-level process involves "making sense" of an ensemble of recognized objects, as in image analysis, and, at the far end of the continuum, performing the cognitive function normally associated with human vision (Gonzalez, Woods and Eddins, 2004). The processes of medical image processing and analysis are illustrated in Figure 1.1. According to Gonzalez definition (Gonzalez, Woods and Eddins, 2004), classification involves recognition of the patterns (classes) of objects in the image, analysis and diagnosis, normally based on the observations (selected features for image representation). For example, features selected from a segmented object in the lung area are classified and interpreted for diagnosis of the lung cancer (nodule). In another reference (Sonka, Hlavac and Boyle, 1999), recognition is considered as the end task in image processing and it is often used in image understanding techniques. The focus of this study is on the classification for a medical image analysis application such as CAD as a high-level process after segmentation and detection of regions of interest (ROI).

One of the common characteristics of a CMIAS is its imperfection. The obtained results suffer from imprecision and vagueness in all level of the processes such as image enhancement, segmentation and classification technologies, uncertainties of mathematical models for measuring complex features as well as imprecision in input data and noisy images. In addition, medical image analysis applications deal with other sources of uncertainties which are inherent in image-based practice of medicine and intuitive knowledge of experts such as inter- and intra observer variability for making decisions, ambiguity in the perception of the clinical vocabulary from different experts point of view, and the dynamic process of disease. Therefore, computerized medical image analysis such as CAD technologies suffer inherently from uncertainty and vagueness both from digital image analysis techniques and medical diagnosis.

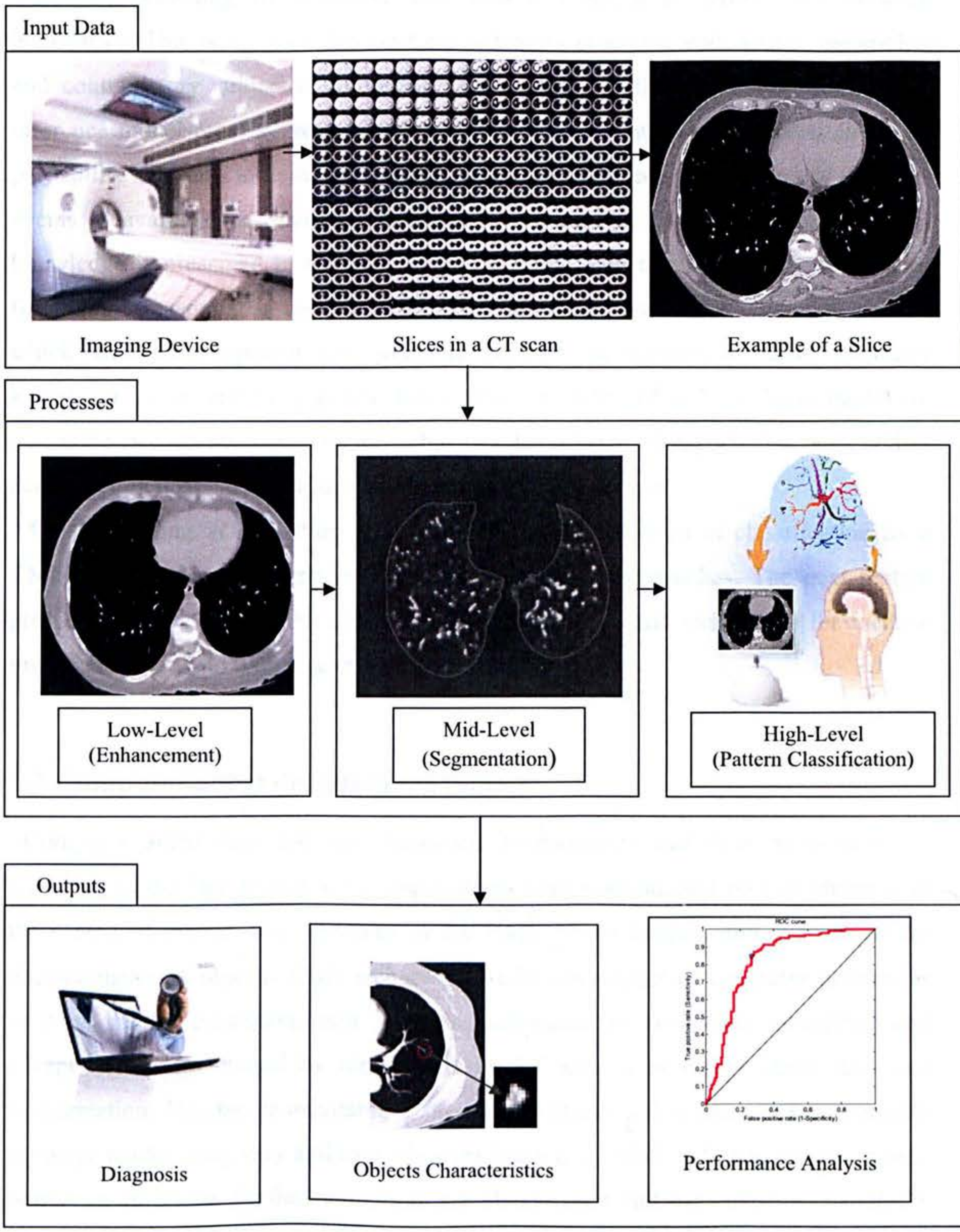


Figure 1.1: A representation of the input data, processes and output of a medical image analysis system

Probability and fuzzy methods have been widely applied for managing the problem of uncertainty in various applications. Choosing a suitable method depends on the nature of the problem and the available data and knowledge. Fuzzy logic has the capability for reasoning according to imprecise and uncertain linguistic terms with un-sharp boundaries. This is the case for most classification problems with vague, incomplete and contradictory subjective knowledge of experts. On the other hand, a Bayesian inference model provides reasoning based on precise knowledge of experts about the probability of events in terms of numbers. When the knowledge about the probability of events is available, a Bayesian inference model can be applied. Otherwise, if the knowledge is presented in terms of linguistic terms with overlapped boundaries then fuzzy logic reasoning is beneficial. A fuzzy inference model is a reasoning method which performs computing with words to manage uncertainties in vague, imprecise knowledge of experts in linguistic terms. The capability of a fuzzy logic model for managing uncertainty issues in the subjective knowledge of experts has been widely considered in medical diagnosis problems in the recent decade.

This study aims at modelling uncertainty issues in the design of classification for a CMIAS such as a CAD application using fuzzy logic approaches. The next section provides an overview of the role and importance of a CAD technology for medical image analysis along with its common components.

1.2 Computer aided detection technology

Computer aided detection and diagnostic technologies and their application are emerging in the last decade. CAD applications have concentrated on identification of the region of interest (e.g., nodules in the lung) in the images, and analysis of the characteristics of objects. CAD technologies take advantages of computer systems as well as image processing and analysis techniques to overcome perceptual and interpretive errors caused by radiologists in the process of image observation and interpretation. This fact is inevitable in the context of medical practice since experience of image readers may vary and each observer (such as a radiologist) has a unique visual search on an image. Furthermore, there are always inter and intra observer variability for interpretation of an image. The main advantages of medical image analysis using a computer system such as a CAD application are:

1. Consistency; i.e., they produce the same result in the same circumstances
2. Speed up very complex analyses
3. Avoid human errors, i.e., distraction and fatigue
4. Improve precision; i.e., using mathematical model
5. Perform parallel processes

Initially, CAD applications have been developed to help radiologists to identify objects in the ROI in an image. Advanced applications of the CAD are expected not only to identify (detect) the candidate objects but also to classify (diagnose) their malignancy risk. However, the majority of the CAD applications only provide the detection process. On the other hand, advances in medical imaging modalities such as X-ray computed tomography (CT) and magnetic resonance imaging (MRI) provide high quality images (slices) per scan. A CT scan slice presents a two-dimensional (2D) X-ray image orthogonal to a section of the body. Modern scanners represent volumetric three-dimensional (3D) images to facilitate the process of image analysis and interpretation through computer. Many technology-oriented studies have been reported for identification and interpretation of the objects in the images and some of them have been commercially developed for screening mammography (Baker *et al.*, 2003) (Image Checker M1000, version 2.5 and R2 Technology, Sunnyvale, CA), for lung screening (Kakeda *et al.*, 2004) (SecondLook, version 4.0, CADx Medical Systems) and for colon screening (Yoshida and Dachman, 2004).

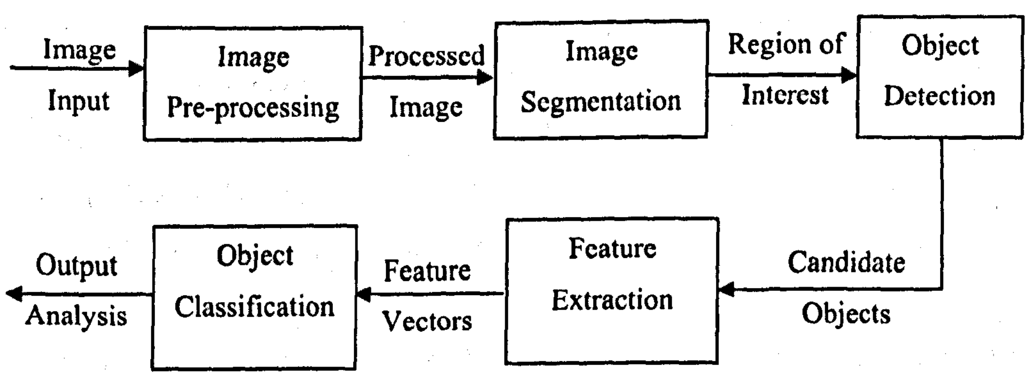


Figure1.2: A typical components of a CAD system

A CAD system can improve image quality through pre-processing, segmentation and detection of the region of interest and candidate objects, recognize and classify the candidate objects and analyse and interpret the system diagnosis. Various techniques are combined to carry out all the above-mentioned processes such as thresholding, clustering, classification and segmentation methods. Artificial intelligence techniques with learning capability for a CAD application can help when dealing with the processes which are efficiently performed by human but are difficult to do with a computer. While CAD systems are application-oriented, a typical CAD application consists of the following major components, as shown in Figure 1.2:

1. *Image pre-processing:* this component enhances the image by minimizing noise and eliminating unwanted structures. The input to this process is acquired images from an imaging device, which contain noise (e.g., caused by motion blur). The noise may affect the region of interest in an image and consequently accuracy of the analysis. Various techniques have been proposed for noise removal in an image such as mathematical morphology (Peters, 1995) and the wavelet transform (Jansen, 2001). The output of this step is a processed image.
2. *Image segmentation*: a segmentation algorithm is applied to extract important objects or regions in an image. The segmentation process usually begins with a processed image and completes when the region of interest for a specific problem is extracted and isolated from the rest of the image. For example, when using CT images of the thoracic area, in one application the region of interest can be nodules while in another application it might be pulmonary embolism (PE). There are various techniques for image segmentation such as edge detection, thresholding, and region growing (Gonzalez, Woods and Eddins, 2004). The proposed approaches can be classified into two main categories: (1) intensity based, and (2) model based approaches. The first group uses the intensity of the pixels such as thresholding methods (Sezgin and Sankur, 2004) and clustering methods (Pham, Xu and Prince, 2000). The second group uses a model to extract the structure of the region of interest such as level set (Sethian, 2003).
3. *Object detection*: this stage detects all possible objects that have the potential to be an object of interest (e.g., nodules in the lung) but further processing is required to remove false positive objects. This step is expected to correctly detect all the interesting objects, but it may also detect unwanted objects. For example, in a lung

CAD application, object detection techniques are expected to extract all nodules inside the lung area but may also detect unwanted objects such as vessels and bronchi. There are various techniques for detection of the candidate objects in the region of interest such as intensity thresholding.

4. *Feature extraction*: based on the knowledge available either from experts or from developers, a set of features is extracted from candidate objects to identify and differentiate objects of interest from other structures. These features are extracted for each of the potential candidates found in the object detection stage. A feature describes scalar properties of an object. An object is typically represented using a group of descriptors called a feature vector (Sonka, Hlavac and Boyle, 1999). Numerical feature vectors are often inputs to a pattern classification application. In machine learning, training or testing datasets usually include a number of numerical vectors of the features. For example, contrast and sphericity are two features of a nodule candidates in a lung CT scan image and the feature vector $X = (contract, sphericity)$ can be used for classification of the objects to nodule and non-nodule classes. The features are categorized in three major classes for region representation (Dhawan, 2011): 1) *statistical Pixel-Level features* which provide quantitative information about the pixels in segmented regions. This type of feature can be the contrast, edge gradient of the boundary, mean, variance, and histogram of the gray level of the pixels within the segmented region; 2) *shape features* provide information about the characteristic of the shape of the boundary of the region. This type of feature can be circularity, compactness or moments. Morphological methods can also be used for description of the shape (Dougherty and Lotufo, 2003); 3) *texture features* provide the local texture information of the segmented region. Local texture information can be computed using wavelets, second-order histogram statistics, co-occurrence matrices or spatio-frequency analysis (Dhawan, 2011); and 4) *relational features* which provide relational and hierarchical information about the structure of regions associated with an object or a group of objects. More details about feature extraction techniques are explained in (Dhawan, 2011).

5. *Object classification*: classification of all objects into classes, according to the feature characteristics extracted from candidate objects (in Step 3). A class describes common characteristics (features) which distinguish an object from other objects in the pattern space (an image). Classification is a process which makes a decision about

the class of a pattern (object). Artificial intelligence (AI) methods are common for analysis of the intelligent behaviour of the systems, classification of the objects and representation of the knowledge for image understanding and analysis (Sonka, Hlavac and Boyle, 1999). In a CAD technology, an intelligent classifier is analogous to a radiologists' diagnostic model. Furthermore, it is needed to manage the subjective knowledge of experts, and learn from new observations. Various techniques have been proposed for classification of the objects in images, such as statistical classification, neural networks (Wu, Jiang and Peng, 2009), rule-based classifier (Dhawan, 2011), and fuzzy classifier and combination of these methods such as Neuro-fuzzy approach, and fuzzy rule-based. The classifiers method can be either supervised or un-supervised. Unsupervised methods classify objects based on similarity in the feature space. K-means clustering and FCM (fuzzy c-mean clustering) clustering are examples of un-supervised classifications. Supervised methods classify objects using a training set of labelled feature vectors into the actual class. Probabilistic methods have been employed to incorporate prior knowledge to improve the accuracy of the classifier. Nearest neighbour, maximum likelihood and Bayes classifiers are widely applied for object classification in medical image analysis applications as supervised classification approaches (Dhawan, 2011).

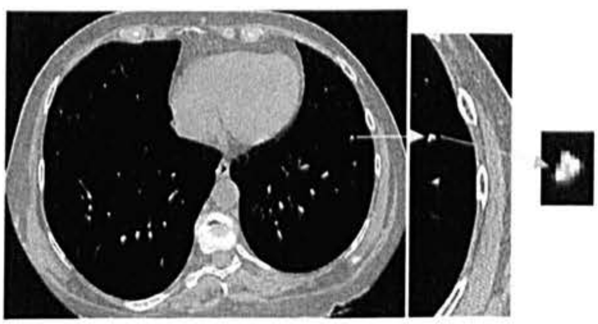


Figure1.3: Sample of a nodule presented in a CT image slice and manually annotated by a radiologist

An example of a CAD system is a lung CAD application for automatic detection of lung pulmonary nodules presented in the CT scan images (see Figure 1.3). A typical lung CAD consists of abovementioned components. After applying image pre-processing and segmenting the lung region, the next step is to define candidate objects which could potentially be a nodule. It proceeds by extracting features such as sphericity, contrast, and volume from candidate objects. A classifier can distinguish between a nodule and non-nodule object according to their characteristics. CAD applications usually provide analysis and interpretation of the output diagnosis. ROC (Receiver Operating Characteristic) is a commonly applied technique for this purpose. The next section explains the ROC analysis method and its application for measuring classification performance in CAD applications.

1.3 ROC analysis for evaluation of the classification performance

The efficiency of a medical image analysis application is usually analyzed using ROC analysis. An ROC curve demonstrates the trade-off between system benefits and its cost as the observer changes the decision threshold. The area under the ROC curve is a measure of the performance of the system as a single measure. The application of ROC curve analysis in machine learning is explained in (Spackman, 1989) to evaluate and compare the efficiency of algorithms. The result of an ROC curve analysis is usually used to evaluate performance of classifiers and to select an appropriate classifier for diagnostic problem in medical image analysis applications.

A classification system predicts the class label of an object. Each object has an actual label *(L_a)* and a predicted label *(L_p)*. If the object is positive (abnormal) and is classified as positive, it is called a true positive (*TP*), but if it is classified as negative (normal) it is called a false negative (*FN*). If an object is negative and is classified as negative, it is called true negative *(TN)*, but if classified as positive, it is called false positive *(FP)* and *N* and *P* are the number of negative and positive samples, respectively. As shown in Figure 1.4, a two-dimensional confusion matrix can be constructed using the classification results of a set of objects as follows (Fawcett, 2006):

Actual label

		P	N
Predicted label	P	True Positive	False Positive
	N	False Negative	True Negative

Figure1.4: Confusion matrix

In a confusion matrix (Figure 1.4), the values along the major diagonal represent the correct diagnosis and the number along the other diagonal represents the classifier error or confusion. According to the numbers in the confusion matrix, following metrics can be measured (Fawcett, 2006):

The true positive rate, which is also, called hit rate or recall is calculated as follows:

$$tp\ rate \approx \frac{Positives\ correctly\ classified}{Total\ positives} = \frac{TP}{P} \tag{1.1}$$

The false positive rate, also called false alarm or error of classifier, is calculated as follows:

$$fp\ rate \approx \frac{Negatives\ incorrectly\ classified}{Total\ negatives} = \frac{FP}{N} \tag{1.2}$$

The sensitivity of ROC curve is equal to tp rate and the specificity is defined as:

$$specificity = \frac{True\ negatives}{False\ positives + True\ negatives} = \frac{TN}{N}$$
$$= 1 - fp \tag{1.3}$$

Positive predictive value or precision is defined as follows:

$$\frac{True\ positive}{False\ positives + True\ positives} = \frac{TP}{TP + FP} \tag{1.4}$$

The accuracy of the classifier is defined as (Fawcett, 2006):

$$accuracy = \frac{True\ positive + True\ nagatives}{Positives + Negatives} = \frac{TP + TN}{P + N} \tag{1.5}$$

The ROC curve (Figure 1.5) depicts the trade-off between the classifier's benefits (*tp rates*) and its cost (*fp rates*) (Fawcett, 2006). Each point of this curve shows the $(fp\ rate, tp\ rate)$ for a classifier decision based on the selected threshold. There are several important points on an ROC space; the lower left point (0, 0) represents the situation where no correct decision is made by the classifier, although this classifier makes no false positive decision but also no true positive. The opposite point is (1, 1) which makes only positive decisions but with high errors. The ideal classifier is (0, 1) which has the maximum accuracy while its opposite point (1, 0) has the minimum accuracy.

The region near the top left represents the classification with higher tp and lower fp rates. In an ROC graph, the area on the line $y = x$ represents a classifier with 50% accuracy which is a statistical random classifier with equal probability of true positive and false positive, see Figure 1.5 (a). The area above $y = x$ has better classification accuracy than the area under it. In Figure 1.5, the accuracy of the classifier with ROC curve (c) is better than the classifier with ROC curve (b).

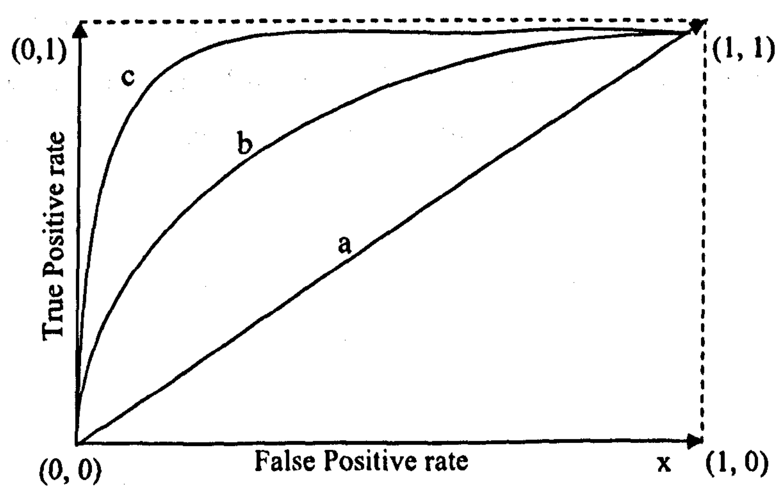


Figure1.5: ROC curve coordinates

ROC curve analysis has been frequently applied for performance evaluation in medical diagnostic application with or without using a CAD application. However, it has been frequently applied to machine learning applications for evaluating the cost of classification. This study applies an ROC curve analysis for performance evaluation of the classification applications. In machine learning applications, an ROC curve is usually plotted after applying a classifier to a set of feature object vectors. The classifier results are usually a class label with a degree of certainty of assigning the object to a specific class. Some classifiers such as fuzzy rule-based systems provide a membership degree to which an object belongs to a class with un-sharp boundary. In a Neural network or Naive Bayes classifiers, a probability is assigned to the classifier decision which represents the degree to which an object is a member of a class (Fawcett, 2006). For each classification result, an ROC thresholding table can be constructed which includes the object number, actual class label and the degree of certainty of classifying an object as an actual class label. By sorting the table based on the degree of certainty in descending order as shown in Table 1.1, different thresholds can be applied. The threshold interval can vary from $-\infty$ to $+\infty$. If the result of classifier is equal or above the threshold, the classifier considers the object as positive otherwise it is negative. Considering different degrees of certainty as thresholds provides different points with different true positive and false positive rates on the ROC space. For illustration, consider a classifier with 10 objects; 5 positive and 5 negative objects.

Table 1.1 demonstrates different selection of the threshold (t) values and the corresponding true positive and false positive rates. Each threshold represents one point on the ROC curve as shown in Figure 1.6. An ROC curve is a step plot. The smoothness of the ROC curve increases by adding more samples to the graph.

Table 1.1 Different thresholds for drawing an ROC curve

Object#	Actual Class	Degree of Certainty	Predicted Class									
			*t**=0.9	0.85	0.8	0.7	0.65	0.55	0.5	0.45	0.41	0.4
1	P	0.9	P	P	P	P	P	P	P	P	P	P
2	P	0.85	N	P	P	P	P	P	P	P	P	P
3	N	0.8	N	N	P	P	P	P	P	P	P	P
4	N	0.7	N	N	N	P	P	P	P	P	P	P
5	P	0.65	N	N	N	N	P	P	P	P	P	P
6	N	0.55	N	N	N	N	N	P	P	P	P	P
7	P	0.5	N	N	N	N	N	N	P	P	P	P
8	N	0.45	N	N	N	N	N	N	N	P	P	P
9	P	0.41	N	N	N	N	N	N	N	N	P	P
10	N	0.4	N	N	N	N	N	N	N	N	N	P
		tp rate	0.2	0.4	0.4	0.4	0.6	0.6	0.8	0.8	1	1
		fp rate	0	0	0.2	0.4	0.4	0.6	0.6	0.8	0.8	1

* *t*: represents a threshold value

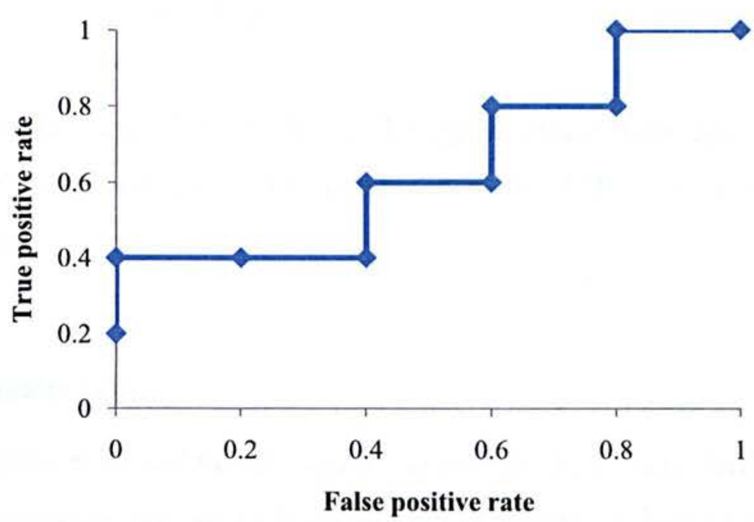


Figure 1.6: An example of an ROC curve

Different points on an ROC curve are obtained by varying the threshold value and measuring true positive and false positive rates. The area under the ROC curve has been frequently employed as a performance measure for measuring and evaluating the cost of classification in machine learning approaches (Bradley, 1997). One of the simplest ways to measure the area under the ROC curve is to estimate trapezoidal integration, the details of this method is explained in (Bradley, 1997).

1.4 Aim and objectives

This study aims at presenting an approach for modelling uncertainties in classification for a CAD application. For this purpose, following objectives are considered:

- Model uncertainties associated with the input of classification using type-2 fuzzy membership functions. The accumulative effect on uncertainty sources in the input datasets for a classification is represented and modelled using type-2 fuzzy sets.
- Model uncertainties in the process of rule extraction for uncertain rule-based pattern classification problems which exhibit a lack of expert knowledge and with an imperfect and imprecise training dataset.
- Evaluate the performance of the proposed approach by applying it to two different medical classification problems: (1) nodule classification in a lung CAD application, and (2) the anonymous Wisconsin breast cancer diagnosis (WBCD) classification problem.

The aim is to show the superiority of the T2FLS (type-2 fuzzy logic system) classifier for managing high levels of uncertainty compared to the T1FLS (type-1 fuzzy logic system) counterpart.

1.5 Research motivation

Despite the progress achieved in developing various medical image analysis systems such as CAD applications, one major issue of current systems is lack of discussion on the impact of the uncertainty in such systems. However, uncertainty issues are major hidden barriers to better performance of a pattern classification. Therefore, it is important to model uncertainties in the design of a classifier. The research aims at presenting a type-2 fuzzy approach to model uncertainty issues in the design of a classification for generally medical image analysis applications and particularly CAD applications with the following properties:

1. *Interpretability:* The classifier needs to be comprehensive to communicate with the medical experts such as radiologists in the process of object detection and diagnosis. A reasonable classification for medical image analysis is expected to maintain the trade-off between accuracy and interpretability.

2. *Accuracy:* This factor is one of the most important aspects of a classifier for medical diagnosis. In such a system, the accuracy of the classification is as important as its interpretability. While an accurate classifier can save lives, a poor classifier with high false positive rates can threaten human lives.
3. *Learning:* a classifier for medical image analysis is expected to learn from new observations in the application domain. Each year, many new cases of abnormality are diagnosed around the world. The classifier is needed to be dynamic to learn from new samples.
4. *Generalization:* design a classifier with the capability to be applied more widely to pattern classification problems in uncertain environments; i.e., with imprecise, vague, and imperfect training datasets, which exhibit a lack of expert knowledge as it is the case for many classification problems.

1.6 Methodology of the research

The methodology of this research is depicted in Figure 1.7.As shown in this Figure, it includes data collection for preparing the training dataset of candidate nodule features which is the input of classification. This part is related to clinical aspects which was performed in parallel to technical part. It also provides a rule set for differentiating nodules from non-nodules identified by radiologist. Details of the features and lung nodule dataset are provided in Chapter 5, Section 5.5. The technical part includes literature review of current technologies and their challenges. Then it introduces the methodology suggested in the study to address uncertainty issues. Then it follows by presenting an approach to tackle uncertainties in membership function and rules of an Interval type-2 fuzzy logic (IT2FL). The performance of the proposed approach is investigated on two CAD classification problems: (candidate nodule classification in a lung CAD and WBCD problem)

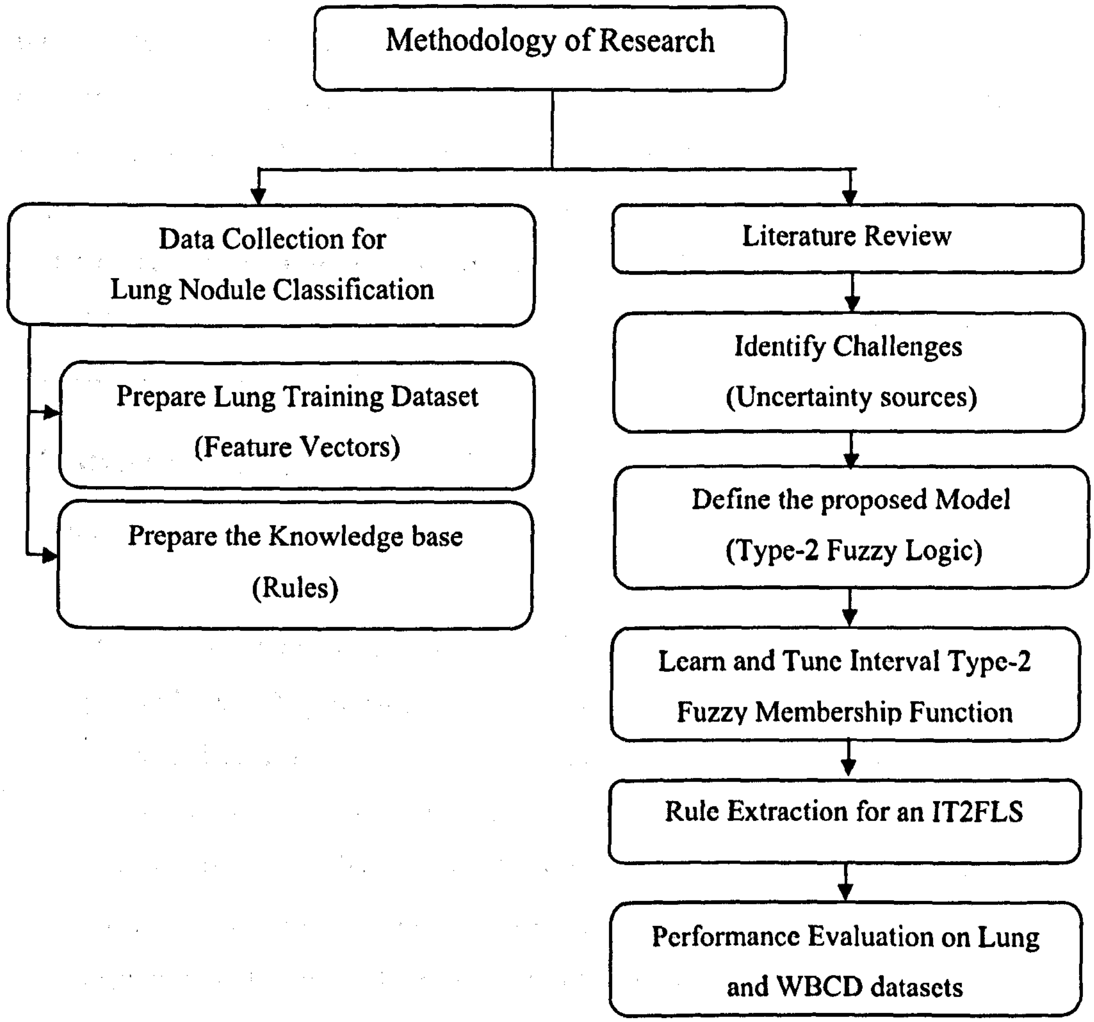


Figure 1.7: Methodology of the research

1.7 Research contribution

This research, for the first time, addresses the uncertainty sources for object classification in CAD applications. It also attempts to tackle the uncertainty problems (explained in Section 1.1) in a classification component of a CAD system. For this, type-2 fuzzy logic capabilities are employed for managing high levels of uncertainties such as inter- and intra observer variability, word perception, and uncertainty in imprecise and inadequate input datasets. Type-2 fuzzy logic, an extension of ordinary fuzzy logic was introduced for managing uncertainty issues which are not managed using ordinary fuzzy sets. Integrating a type-2 fuzzy logic model for object classification in a CAD architecture allows us to represents and model accumulative effect of these uncertainties. For this, an automatic approach models uncertainty in a training dataset using membership function of type-2 fuzzy sets.

On the other hand, in most medical applications, extracting knowledge (rules) from various experts and fusing their knowledge is difficult, sometimes impossible or expensive to achieve. For this, the research introduces the idea of uncertain rule-based pattern classification to extract rules automatically from imperfect and imprecise training dataset to classify the patterns. This approach models uncertainties in classification in the process of learning rules for a type-2 fuzzy logic.

1.8 Structure of the dissertation

The thesis consists of two main parts which are explained as follows:

- **Part I:** this includes a discussion of uncertainty issues in a general CMIAS and particular CAD applications in Chapter 2. A fuzzy logic (FL) rule-based model is suggested as a solution to address the uncertainty issues. It is followed by an overview of fuzzy logic approaches and advantages for tackling the problem of uncertainty in rule-based classification. Particularly, type-1 FL, type-2 FL, and interval type-2 FL (an extension of general type-2 FL) theories and concepts are explained in Chapters 3 and 4. This study uses an interval type-2 FL (IT2FL) as a practically applied type-2 FLS to overcome the uncertainty issues associated with rule-based classifiers.

- **Part II:** The second part presents the approach proposed in this research for modelling uncertainty issues. This includes:
  1. An automatic approach for modelling uncertainty associated with the input dataset through type-2 fuzzy membership functions. This method is applied and integrated to the classification component in a lung CAD application. Chapter 5 provides results of this application.
  2. The idea of uncertain rule-based pattern classification for classification problems with a lack of expert knowledge and imprecise and imperfect datasets. This presents an approach for modelling uncertainties in the process of extracting rules for a type-2 fuzzy logic classifier from imprecise and imperfect

training datasets. The proposed methods for modelling uncertainty in membership function and rules are applicable for the classification problems with lack of expert knowledge and with an incomplete and imprecise training dataset. In order to verify the generalization property of the proposed approach, it has been applied to the popular Wisconsin breast cancer diagnosis (WBCD) database. The features in this database are computed using image processing techniques and are visually assessed by an expert. Furthermore, the noisy and variation in this dataset in addition to uncertainty in image processing techniques makes it a suitable choice for validation of the proposed approach. This approach and the results of applying that for the WBCD classification problem are presented in Chapter 6.

Lastly, the dissertation is concluded in Chapter 7 with a discussion of the research achievements and future investigations.

Part I

Chapter 2: Uncertainty Challenges in Classification for Medical Image Analysis Applications

This chapter aims at addressing different sources of uncertainty associated to CMIAS such as CAD technology. Uncertainty issues are addressed and explained using examples of actual cases in CAD applications or medical diagnosis. Furthermore, it briefly provides an overview of the probability and fuzzy set theory capabilities for managing uncertainty issues.

2.1 Introduction

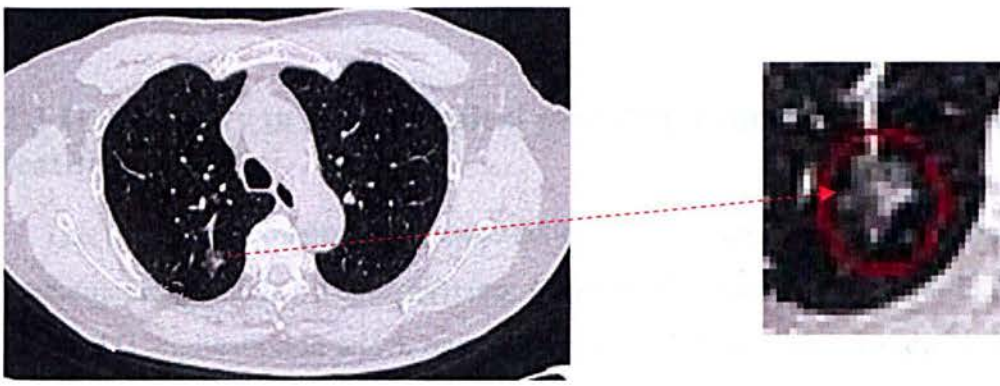
Computerized image analysis suffers from imperfection, imprecision and vagueness in the input data and its propagation to individual components of the technology including image enhancement, segmentation and pattern recognition. Furthermore, a computerized medical image analysis system (CMIAS) as a 'human-centred' problem, deals with other sources of uncertainties that are inherent in image-based practice of medicine, including uncertainties in: (1) the 'medical science' and incomplete knowledge of human, (2) subjective knowledge of experts for diagnosis and clinical applications, and (3) one expert or between experts' for diagnosis about the same clinical case in the same image. This is inevitable in clinical diagnosis because of an experts' experiences and previous knowledge, their observations in laboratory examinations and subjective information, and the patients' history. Moreover, the expert final decision is usually associated with a confidence level. Therefore, a CMIAS such as computer aided detection (CAD) system suffers inherently from uncertainties and imprecision both from digital image analysis techniques and medical diagnosis. The main sources of uncertainty in a CAD technology can be summarized as follows:

1. Imprecision in input data and noisy images (*input data*)
2. Uncertainty in all processes of image enhancement, segmentation, edge detection and converting a 3D image to a 2D image.
3. Uncertainty in mathematical models for measuring the *complex features* of images (such as degree of circularity)
4. Uncertainty in the final decision of an expert about a pattern (*intra-uncertainty* (Mendel, 2007b), called intra-observer variability in the clinical context)
5. Uncertainty between experts for making a decision about a pattern (*inter-uncertainty* (Mendel, 2007b), called inter-observer variability in the clinical context)
6. Uncertainty in the meaning of different linguistic terms used by experts (*word perception*)
7. Dynamic aspect of disease progress and its influence on the diagnosis process (*non-stationary features*)

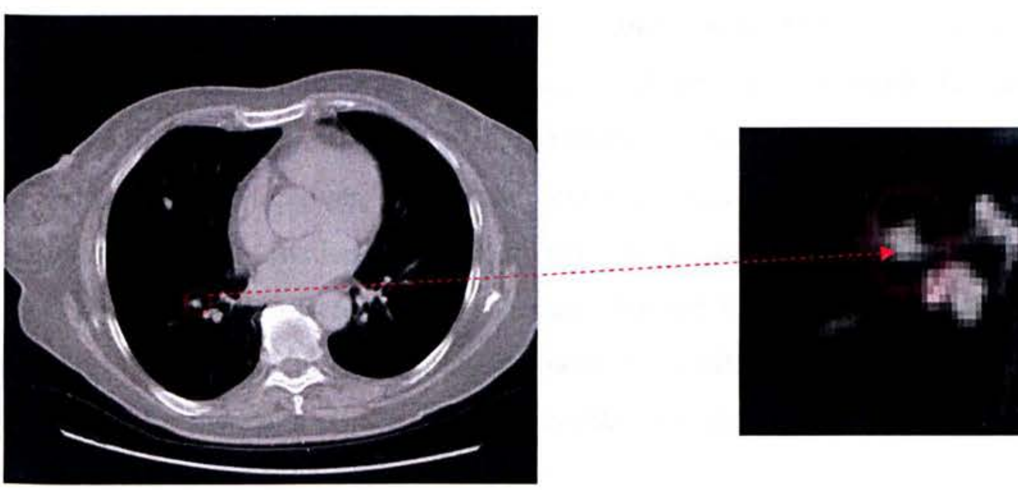
While there are several technology-oriented studies reported in developing CAD applications such as mammography CAD (Leon *et al.*, 2009), colon CAD (Halligan *et al.*, 2006; Hein *et al.*, 2010) and lung CAD (Lee *et al.*, 2001; Dehmeshki *et al.*, 2003; Dehmeshki *et al.*, 2007; Ye *et al.*, 2009), no attempt has been made to address, model and integrate these types of uncertainty in the design of the components of a CAD system, even though uncertainty issues directly affect the performance of the system and the accuracy of the results. The rest of this chapter explains the uncertainty sources associated with a CAD system with examples of real cases.

2.2 Uncertainty in input data and noisy images

One of the characteristics of information derived from an imaging device is its imperfection because of the noise in the surrounding environments. Therefore, the acquired image can be inconsistent, incomplete and blurred. Furthermore, the partial volume effect (PVE) is a common problem that arises from image acquisition. The PVE problem is related to resolution limitation of digital imaging devices, which may lead to inaccurate classification and measurement results. Image enhancement techniques such as smoothing are frequently applied to reduce the effect of the PVE and noise in the image (as shown in Figure 2.1).



(a)



(b)

Figure 2.1: Examples of a nodule in lung CT scan; (a) Noisy image (b) Smoothed image

On the other hand, the features of a detected object in an image (e.g., sphericity, volume) are either measured manually by experts (e.g., radiologist) or automatically by image processing and analysis application such as CAD. In both cases, measuring the characteristics of objects with noisy and blurred borders and texture may lead to inaccurate or imprecise results. These measures are usually used in a training or testing dataset for learning a classifier. Interpretation and analysis of the medical images using these imprecise measurements affect the accuracy and precision of the classification results.

2.3 Uncertainty in mathematical models for measuring the complex features

This section explains the uncertainty in measuring complex features using mathematical models by exemplifying. Consider the nodule classification problem in a lung CAD application in which one of the features of the nodule candidate is the degree of sphericity. In the study reported in (Dehmeshki *et al.*, 2007; Ye *et al.*, 2009), three radiologists annotated nodules in the computed-tomography (CT) scans of the lung for training purpose. Two mathematical models including shape index (shape information) and intensity feature (local intensity distribution) were defined as indicators of the sphericity of the nodule and the non-nodule objects in a training dataset. There is variation between the results obtained by theses two mathematical models (as illustrated in Figure 2.2). Moreover, each mathematical model produces its own error. By considering more mathematical methods, more variations are observed in the measurement results. Refer to the sphericity feature for nodule detection in the lung, applying two different measurement methods result in two different degrees of sphericity for the same object. This fact emphasizes the need for managing uncertainty issues in mathematical models.

Object #	Shape Index	Intensity Distribution
1	0.7	0.78
2	0.6	0.63

Figure 2.2: Different mathematical model results for measuring sphericity

2.4 Inter- and intra- observer variability

In a CMIAS, knowledge is usually extracted from a group of experts (physicians or radiologists). While there are variations between an expert's decisions (inter- observer variability or within an expert decision (intra- observer variability), an expert system is expected to manage these sources of uncertainty. A panel of experts is usually considered to form a consensus about the class of an object in an image during blind or un-blind sessions. It should be noted that the process of providing a group of qualified experts and fusion between them is very expensive and time-consuming. A classifier for a CAD system is expected to be aware of and to manage these sources of uncertainty. The rest of this section discusses inter- and intra- observer variability examples in a CMIAS.

The variation between radiologists in the identification of a nodule has been reported in the lung image dataset consortium (LIDC) study in (van Ginneken *et al.*, 2010). In this study, four radiologists identified nodules in two blind and un-blind sessions. A total of 174 nodules were identified at least by one of the four radiologists, 146 by at least two, 121 by at least three and only 90 nodules were identified by all four radiologists. The LIDC study indicates about 50% variation between the radiologists' annotations. Figure 2.3 illustrates the agreed results between four radiologists in the LIDC for nodule detection.

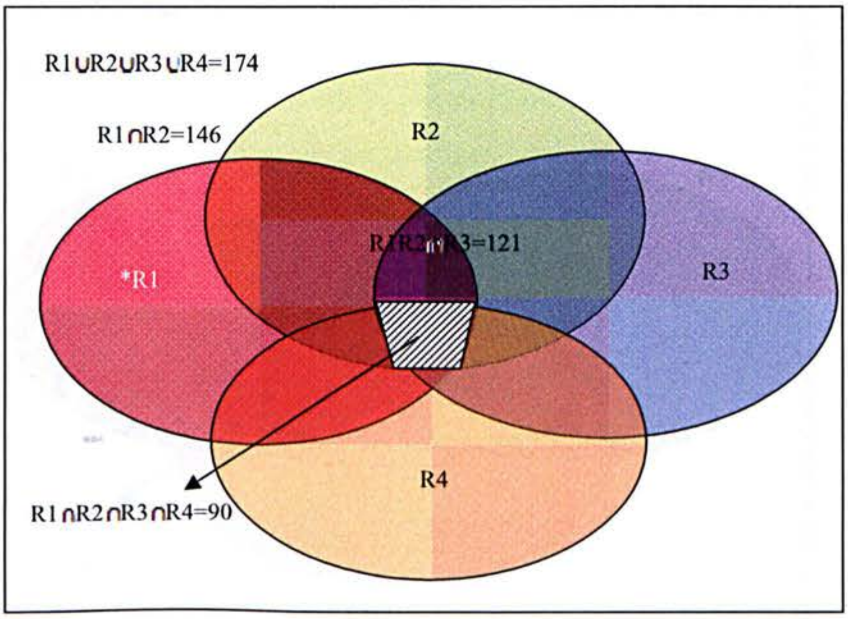


Figure 2.3: Illustration of variation between radiologists for nodule identification in the LIDC study (van Ginneken et al., 2010); R_i (i=1, ..., 4) represents one of four radiologists

In another study reported in (Dehmeshki *et al.*, 2003), radiologists' performance were compared with or without using computer assisted detection (CAD) software for pulmonary nodule detection in CT scan images. The CAD system applied image processing and classification techniques. Three radiologists agreed that 34 of the 42 scans were abnormal (in only one scan) and the remaining 8 scans were considered as normal. In total, 207 nodules were identified at least by one of three radiologists, and only 49 nodules were identified by all three radiologists. This study reported a very high variability for nodule detection by the three radiologists as shown in Figure 2.4; 76% of the nodules were undetected by at least one radiologist. The size of nodules missed by the radiologists varied as follows: 0-5 mm (61 nodules), 5-10 mm (84 nodules), 10-20 mm (11nodules) and >20 mm (2 nodules). In comparison, the accuracy of the CAD results were promising and consistent for detection of the nodules; 100% of pulmonary nodules were identified by at least one radiologist. Furthermore, some potential nodules were detected by the image classification techniques which were not identified by the radiologists. This study clears the role of the CAD systems for analysis of the medical images for tackling the inter-uncertainty issue.

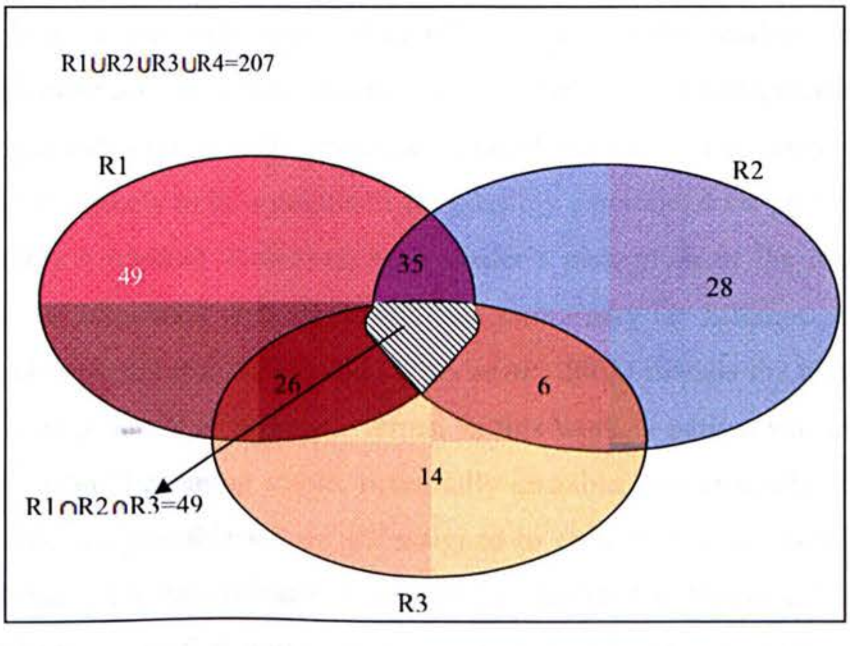


Figure 2.4: illustration of variation between radiologists for nodule identification between three radiologists in (Dehmeshki *et al.*, 2003)

In a study reported for assessment of the performance of ten radiologists without computed topographic colonography CAD and with the CAD after two months, the CT colonography images of 60 out of 107 patients with 142 polyps were investigated (Halligan *et al.*, 2006). Using CAD, 41 patients with polyps were identified and the radiologists' sensitivity per patient increased to 70% (an average of 12 more polyps) and the interpretation times decreased significantly. This example shows that colonography CAD can significantly decreases intra- observer variability.

2.5 Uncertainty in words perception from different experts point of view

Expert knowledge is usually expressed using linguistic terms. The extracted knowledge is based on a fusion between experts. Different experts with different levels of expertise may have a different diagnosis for a common clinical case (John and Lake, 2001) (e.g., classify an object in an image). Moreover, the experts' diagnosis is associated with a confidence score for different linguistic terms. However, the perception of linguistic words associated to confidence scores may vary from various expert's point of view. This is known as a *word perception* problem.

For illustration, in the study reported in (Roos *et al.*, 2010), readers interpreted CT images to identify all pulmonary nodules with diameter equal and greater than 3 mm. The readers classify objects with confidence rating from 1 to 5 (1 for very unlikely to be a nodule; 2 for unlikely to be a nodule; 3 for possibly a nodule; 4 for probably a nodule; 5 for definitely a nodule). However, each reader's perception of the meaning of the linguistic terms associated with the confidence rating may be different. Another study reported in (John and Lake, 2001) and (Wills *et al.*, 2003) models the clinical intuition of nursing staff in terms of linguistic terms. In this work, a patient was categorized in five groups: stable, becoming stable, potentially unstable, and critically ill. Lower and upper probable and possible values are assigned to each of five categories. The lower and upper bounds for the critically ill category is depicted in Figure 2.5. These values are elicited from a qualified nurse while it may change from one nurse to another. By preparing a survey of lower and upper bound from the perspective of different nurses, the mean and variance for each of the five categories are estimated. The nurses were asked to define the possible starting and ending values for each category such as "critically ill". As shown in Figure 2.5, the defined starting interval from the point of view of different nurses is [7.44 8.72] and the ending interval was defined as [9.72 10.44]. For example for at value equal to 9.26 instead of a function value, we have an interval. This is showing an example of uncertainty in the meaning of the words.

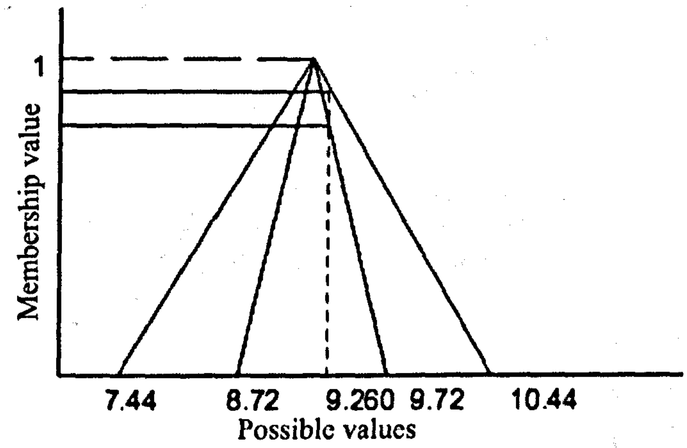


Figure 2.5: A trapezoidal membership function for "critically ill" category (Wills *et al.*, 2003)

2.6 Uncertainty in non-stationary features

A non-stationary feature has statistical attribute (mean and variance parameters) that changes by varying time or space (Zeng and Liu, 2007), and (Priestley, 1988). The development of a disease is a dynamic process over time and has a non-stationary feature; it begins at an early stage and continues to an end stage. A patient passes the symptom information to the expert using imprecise linguistic terms. A clinician makes the preliminary diagnosis based on the vague knowledge. There is also the possibility of confusion in an expert diagnosis because of similarities of the different symptoms of the diseases in the early stages (John and Innocent, 2005). In a model proposed for modelling uncertainty in clinical diagnosis, (John and Innocent, 2005), the stage of the disease is estimated using duration constraints of the disease and related symptoms. Figure 2.6 (Innocent and John, 2004) indicates the relation between support (compatibility degree to which a diagnosis follows a related evidence) and the symptoms in the first few days of appearance the disease of influenza and scarlet fever. These symptoms represent what are expected for influenza disease, i.e., fever, headache, vertigo, back pain, muscle pains, collapse, cough, running eyes and running nose and sore throat. The slope of influenza disease is greater than scarlet fever, which indicates faster growth of the influenza symptoms compared to the other disease symptoms, but in the first two days, it is actually difficult to distinguish between these two diseases according to their symptoms.

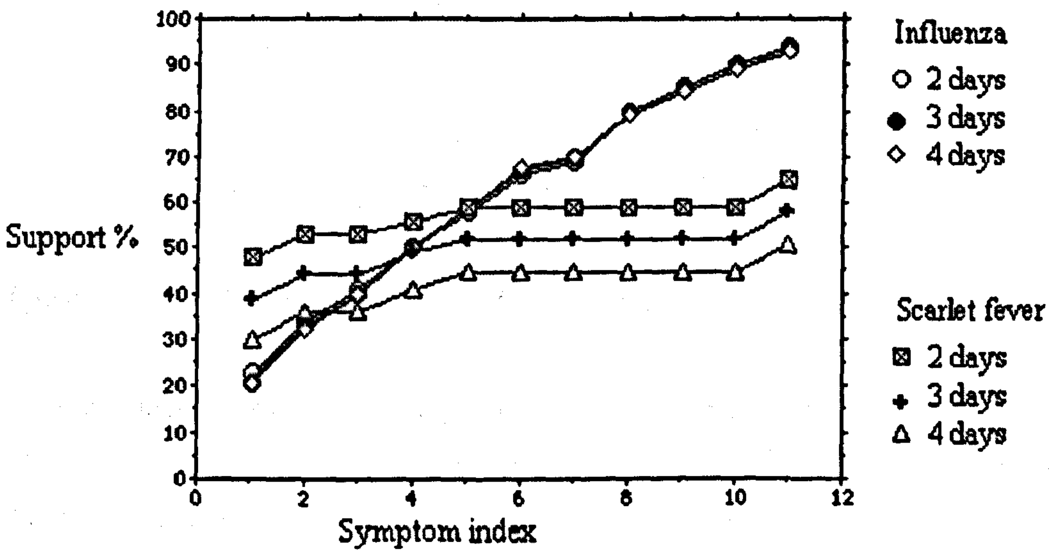


Figure 2.6: Confusion in symptoms duration of the influenza and scarlet fever (Innocent and John, 2004)

2.7 Uncertainty in image processing techniques

One of the common characteristics of information derived from an imaging device is its imperfection. The acquired image can be inconsistent, and incomplete. According to the resolution of an imaging technology, a voxel (3D pixel) in an image may belong to more than one class while object classification techniques may assign it to only one class. Furthermore, the noise in an image affects the accuracy of the image processing results. In the study reported in (Dehmeshki *et al.*, 2008), for segmentation of the pulmonary nodules in the CT images, according to the intensity, each pixel is assigned to one of the three classes: foreground, background, or partial volume or a combination of partial volume and the two other classes whilst the distribution of each class is considered as Gaussian (see Figure 2.7). The overlapped areas in Figure 2.7 are somewhere the uncertainty in edge detection takes place.

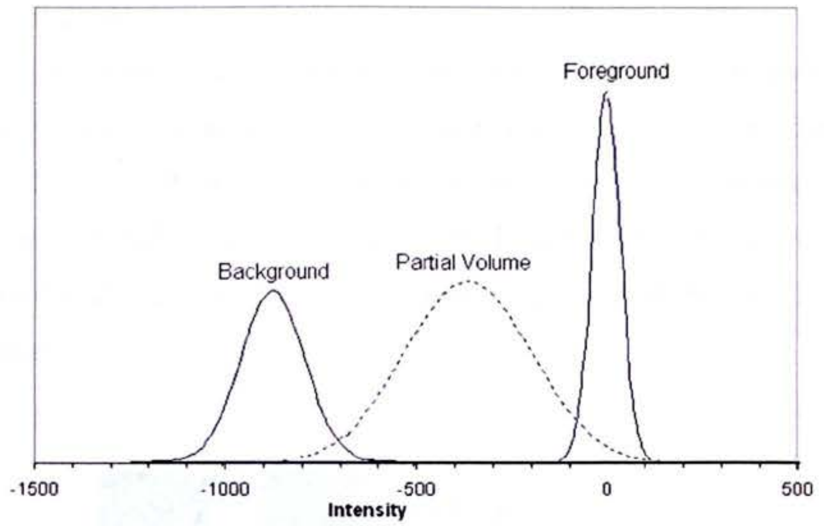


Figure 2.7: Modeling of three classes for a lung phantom in (Dehmeshki *et al.*, 2008)

For segmentation of the pulmonary nodules in a CT image of lung in (Dehmeshki *et al.*, 2008), a region growing in combination with a fuzzy technique is employed. The seed point is randomly selected in the nodule region. For a nodule attached to the blood vessel, Figure 2.8 illustrates how the selected seed point affects on the region growing result. A randomly selected seed point has leakage into the blood vessel (see Figure 2.8 (b). However another seed ponit selected by automatic region growing (red circle in Figure 2.8 (c)) detects the nodule but there is uncertainties in the border of object where it is attached to blood vessel. In most CAD applications, radiologists usually select seed points. This fact makes the segmentation process more complicated because of human errors (such as eyestrain).

(a) (b) (c)

Figure 2.8: The effect of a seed point placed on a nodule attached to blood vessel; (a) Original image (b) A random seed is selected (shown with plus), (c) An optimum seed is selected (shown with a circle) (Dehmeshki *et al.*, 2008)

In a CMIAS, a 3D object is constructed using 2D slices (Figure 2.9) and the relation between different parts of objects in several slices. The accuracy of the object detection result is affected by slice thickness; i.e. some part of the important information in small objects, such as a nodule with less than 2mm diameter or a blood vessel may be ignored. Moreover, this complicated task is usually associated with the PVE problem and missing slices.

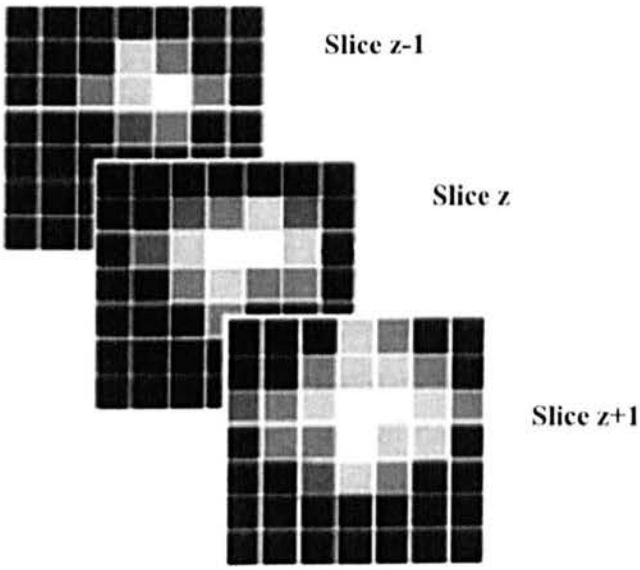


Figure 2.9: A candidate nodule in three adjacent slices (Gomathi and Thangaraj, 2010)

Consequently, a major hidden barrier to better performance of a CAD application results from imprecision and vagueness associated with image analysis (see Figure 1.1); such as uncertainty in the gray level, texture, contours, edge detection, converting a 3D image to a 2D image, and the relationship between two segments of an image.

2.8 Uncertainty in object classification

The objective of a classification system for a CMIAS such as a CAD system is to assign the candidate objects in an image to their corresponding classes. One of the common issues in the design of a classifier for a CAD is the lack of consideration of the effect of uncertainty issues (discussed in Sections 2.1 to 2.7). A reliable classification is expected to address and model vagueness and imprecision in the following ways:

1. ***Input of classifier:*** this means features extracted from processed objects in an image, and mathematical modelling of them, and inter- and intra- uncertainties in the training datasets
2. ***Design of the classifier:*** select an appropriate classification method to cope with uncertainties for assigning an object to a class
3. ***Output decision:*** define the degree of certainty of the classifier output. Ignoring uncertainty issues results in a poor performance and inaccurate results.

The main aim of this research is to manage the accumulative effect of uncertainty issues in the design of a classification for a CAD application.

2.9 Probability theory and fuzzy set theory for modeling uncertainty

Probability represents the likelihood of the occurrence of an event. The Bayesian inference model provides reasoning based on probability of events. For example, the symptom of high fever in a patient maybe cause of flu, cold, malaria, Scarlet fever or other diseases. A probability can be assigned to each of these possibilities given the fever symptom. Using the Bayes' theorem (Box and Tiao, 1992), the conditional probability of a random patient has flu, $P(A)$, knowing the probability that a patient has high fever $P(B)$ is calculated using the following equation:

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} \tag{2.1}$$

where $P(B)$ and $P(A \text{ and } B)$ are known, and $P(A \text{ and } B) = P(B \mid A)\,P(A)$. If the knowledge about P(A and B) and P(B) are certainly either available or straight forward to calculate, the probability theory can lead to a certain results, otherwise, fuzzy set theory can help. An example of such knowledge is "a patient suffering from flu shows 60% of all cases a high fever". However, there are uncertainties in measuring probability of a patient with 38.5°C (as a case of high fever) and the probability may vary from different experts' point of view. Fuzzy set theory takes advantage of membership functions to define a degree to which an object belongs to a linguistic fuzzy set with un-sharp boundaries. Fuzzy logic provides an inference model for reasoning with linguistic words which mimics human reasoning in vague environments. For the above example, the high fever can be modeled using fuzzy sets of "Low", "Medium" and "High" fever. Then 38.5°C can belong to one or more than one fuzzy set with different degree of memberships (Figure 2.10).

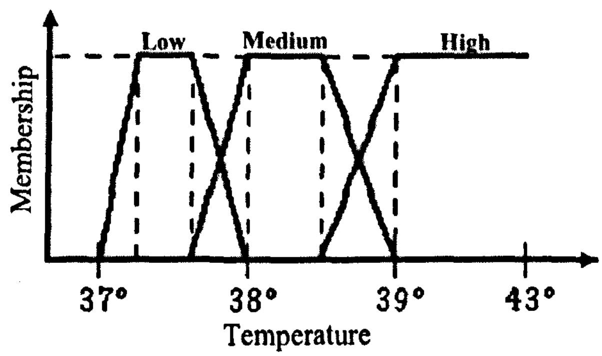


Figure 2.10: An example of temperature fuzzy set

One can conclude that a combination of fuzzy and probabilistic approaches can bridge the gap between fuzzy and probabilistic methods for modelling uncertainties. The fuzzy-probability hybrid approach takes advantages of both fuzzy and probability techniques for modelling uncertainty (Ross, Booker and Parkinson, 2002). This research takes advantages of fuzzy logic for modelling uncertainty sources in a CAD application. It also utilizes probability techniques whenever they are required such as modelling parameters of a membership function using a Gaussian distribution.

2.10 Summary

This chapter presented the uncertainty issues associated with a CMIAS such as CAD application. This is the first time that the concept of uncertainty in a CAD application as a medical image analysis tool is addressed and discussed in details. Different sources of uncertainty were explained using examples of either real medical image analysis or CAD applications. Furthermore, it discussed probability and fuzzy set capabilities for handling the problem of uncertainty in the systems. However, the focus of this study is on fuzzy based approaches for modelling uncertainty but taking advantage of probability methods whenever they are required, such as modelling the membership function of fuzzy sets using Gaussian distribution properties. The next two chapters explain how fuzzy logic and its extension type-2 fuzzy logic are capable of capturing and modelling uncertainties discussed in this chapter for classification.

Chapter 3: Fuzzy Logic for Rule-based Pattern Classification

This chapter presents the importance and role of fuzzy logic for handling the problem of uncertainty in rule-based classification. The theory and concepts of ordinary fuzzy logic as a rule-based system and its components are reviewed. The fuzzy inference model is illustrated for handling the problem of uncertainty for the nodule classification in a lung CAD application. It also discusses uncertainty issues which are not modelled using ordinary fuzzy sets and introduces type-2 fuzzy set for managing the remaining uncertainty issues.

3.1 Introduction

There are various approaches for pattern classification. The main taxonomy can be defined as (Kandel and Friedman, 1999): distance function and clustering, statistical and syntactic, neural networks (Bishop, 1995; Meyer-Baese, 2004), and fuzzy rule-based approaches. Furthermore, there are combinations of these approaches such as fuzzy-neural networks (Meyer-Baese, 2004; Abe, 2001; Pal and Mitra, 1999), fuzzy clustering (Kerre and Nachtegael, 2000). Among them, the fuzzy rule-based approach provides the most comprehensive rule set for user interaction. The capability of handling uncertainty in classification in addition to high interpretability of the fuzzy rules has been the focus of this study for the problem of object classification in a medical imaging application such as CAD.

Zadeh introduced the theory of fuzzy sets in 1965 (Zadeh, 1965). In a fuzzy logic, an event can belong to more than one sample space with un-sharp boundaries. Main useful features of the fuzzy set for tackling uncertainty issues are as follows (Chi, Yan and Phạm, 1996):

1. Fuzzy logic models uncertainty sources of imprecise and vague data
2. Models uncertainty in classification problems with un-sharp boundaries between classes
3. Fuzzy reasoning mimics an expert reasoning model based on subjective knowledge and vague linguistic terms
4. Provides capabilities of handling uncertainty in linguistic term in fuzzy sets
5. In the circumstances of the vague and inadequate data where it is difficult to define randomness of the data, fuzzy sets are applicable
6. Models mathematically ill-defined applications

The superior ability of fuzzy set theory for handling uncertainty sources is briefly quoted by Zadeh (Zadeh, 2000) and cited in (Bezdek and Pal, 1992):

"The conventional approaches to the analysis of large-scale systems were ineffective in dealing with systems that are complex and mathematically ill-defined, as are most of the real-world systems in which human perception and intuitive judgments play important roles".

Bezdek and Pal (Bezdek and Pal, 1992) collected seminal published papers on the theory and application of pattern recognition based on fuzzy sets in their book. According to their work, the most popular applied fuzzy methods for pattern recognition are: the use of the fuzzy membership function, fuzzy clustering, fuzzy entropy, fuzzy decision tree, fuzzy measure and fuzzy integral, and fuzzy rule-based systems. Among them, fuzzy rule-based classifiers are more analogous to the model of human reasoning using imprecise and vague information (Ishibuchi and Nojima, 2008). Furthermore, a fuzzy rule-based classifier is a practical classifier with high interpretability (Ishibuchi and Nojima, 2008; Arakawa, Kerre and Nachtegael, 2000). The linguistic terms in a fuzzy rule-based classifier provide the most comprehensive rule-based model for environments with the need for user interaction such as medical applications. This is one of the reason that the focus of this study is on fuzzy rule-based classification. In a fuzzy rule-based classifier, rules show expert's knowledge and are expressed using linguistic terms. The general form of If-Then rules is: *If <antecedent> Then*

<Consequent>, e.g. *"If the object's shape in the lung is round and the object is bright then it is likely to be a nodule"*.

Fuzzy logic as a rule-based classifier is approximately equal to computing with words in which reasoning is performed using vague, and imprecise words instead of numbers (Zadeh, 1994). The nature of computing with words has made fuzzy logic significantly different in reasoning in comparison to other methods such as predicate logic, Bayes theory, probability theory and neural network theory. There is some argument about the tradeoffs between interpretability and accuracy in fuzzy rule-based systems (Casillas, 2003; Ishibuchi and Nojima, 2007). Hybrid learning techniques are frequently incorporated into fuzzy rule-based system for improving accuracy. Examples of such a systems are Neuro-fuzzy (Meyer-Baese, 2004; Abe, 2001; Pal and Mitra, 1999) and genetic fuzzy (Herrera, 2008; Casillas *et al.*, 2005) systems which take advantage of neural network learning techniques and the global optimization properties of evolutionary computing, respectively. This study takes advantage of learning methods to improve the trade-off between accuracy and interpretability in the classification through fuzzy logic.

3.2 Fuzzy logic

Fuzzy logic has been introduced to manage uncertainty in numerical information (Zadeh, 1965). Fuzzy representation of the data structures using linguistic terms is a very interpretable and a comprehensive way to model intuitive and vague knowledge of the experts. Fuzzy set A in universe of discourse X can be defined as a set of ordered pairs of element x in X and the grade of membership of $x, \mu_A(x)$, to fuzzy set A (Zadeh, 1965), i.e.,

$$A = \{(x, \mu_A(x)) | x \in X\} \tag{3.1}$$

where the two dimensional membership function $\mu_A(x)$ is a crisp value between 0 and 1 for all $x \in X$. Linguistic terms are modelled using fuzzy sets. One of the parameters in the design of a fuzzy logic is the number of fuzzy sets associated to a linguistic term. For illustration purpose, the linguistic variable "sphericity" can be represented with three terms ("weak sphere", "moderate sphere", and "strong sphere"), see Figure 3.1.

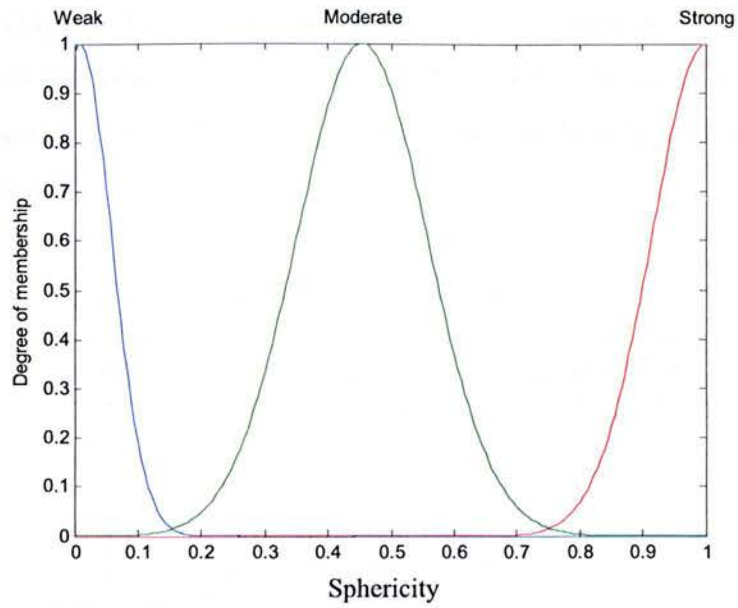


Figure 3.1: Membership function corresponding to linguistic terms

Fuzzy logic as a soft computing method mimics cognitive reasoning of the human mind based on linguistic terms for performing tasks in a natural environments. Zadeh (Zadeh, 1994) defined fuzzy logic as a logical system that helps formalizing approximate reasoning. He also introduced it as a branch of fuzzy set theory which includes concepts of linguistic variables, fuzzy if-then rules, fuzzy quantifier and reasoning models. Zadeh clarified misleading arguments about a fuzzy logic. He stated that FL as a precise logic of imprecision and approximate reasoning which is not fuzzy in nature.

The main capabilities of fuzzy logic are described by Zadeh (Zadeh, 2008) as

" ... Fuzzy logic may be viewed as an attempt at formalization/mechanization of two remarkable human capabilities. First, the capability to converse, reason and make rational decisions in an environment of imprecision, uncertainty, incompleteness of information, conflicting information, partiality of truth and partiality of possibility – in short, in an environment of imperfect information. And second, the capability to perform a wide variety of physical and mental tasks without any measurements and any computations."

Fuzzy logic (FL) defines the general architecture for designing a classifier with linguistic terms (Figure 3.2). There are four major components in a fuzzy logic architecture (Cazarez-Castro, Aguilar and Castillo, 2010): 1) fuzzifier, 2) inference engine, 3) rules, and 4) defuzzifier. The rest of this section briefly describes the role of each component in a FL.

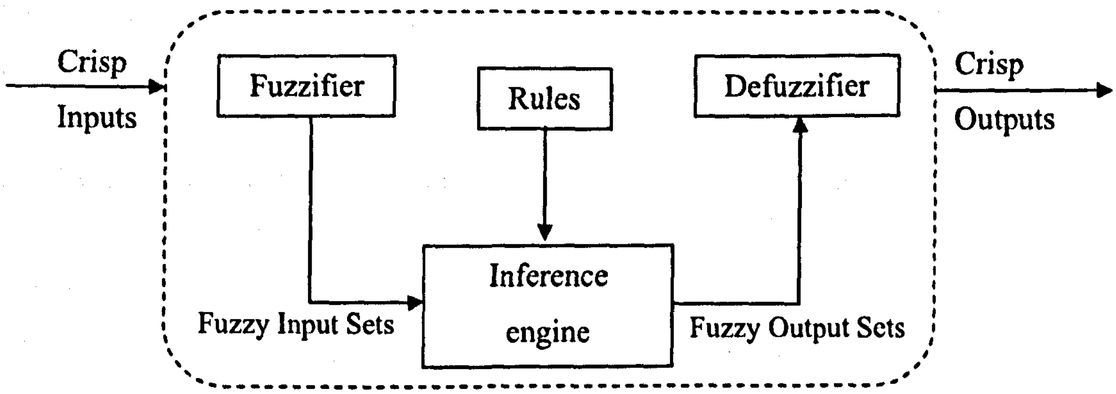


Figure 3.2: Type-1 Fuzzy Logic (FL)

3.3 Fuzzifier

In the fuzzification process, a membership function is assigned to each input value x_i in universe of discourse X (all possible values for that variable). For an n-dimensional input $X = (x_1, x_2, \dots, x_n)$, this component assigns a membership function $\mu_{A_{ij}}(x_i)$ to each input value x_i.

3.4 Different types of rules in a FL

In a fuzzy logic with n inputs $x = (x_1, x_2, \dots, x_n)$ and one output $y \in Y$, a general form of rule is,

$$Rule\ R_i : If\ x_1\ is\ A_{i1}\ and\ \dots and\ x_n\ is\ A_{in}\ then\ y\ is\ B_i$$

where $i = 1, \dots, r$ is the number of rules, $A_{i1}, \dots, A_{in}$ are antecedent fuzzy sets associated to linguistic terms, y is an output variable and B_i is a consequent fuzzy set.

Various types of rules have been proposed in the literature. Mendel (Mendel, 2001) has proposed six categories of rules as follows:

Incomplete rules: rules for which antecedent is a subset of the whole input sets. e.g.,

$$Rule\ R_i : If\ x_1\ is\ A_{i1}\ and\ \dots and\ x_n\ is\ A_{im}\ then\ y\ is\ B_i$$

Compare to the general form of the rules, a subset of all antecedent fuzzy sets are involved in this type of rule.

Mixed Rules: the rules with a mixture of "and" and "or "connectives.

$Rule\ R_i$:

$$If\ (x_1\ is\ A_{i1}\ and\ \dots and\ x_n\ is\ A_{im})\ or\ (x_{m+1}\ is\ A_{i\,m+1}\ and\ \dots and\ x_n\ is\ A_{in})\ then\ y\ is\ B_i$$

We can split mixed rules into several "and" rules.

Fuzzy statement: some rules without any antecedent are called a fuzzy statement e.g., "y is high". This type of rule is an extreme type of incomplete rule.

Comparative rule: some rule are in the following form, e.g. The brighter the x_1 the greater the y. This form of rules can be converted to the standard for of If-then rules, e.g.,

$$Rule\ R_i : If\ x_1\ is\ bright\ then\ y\ is\ big$$

where bright and big are fuzzy sets.

Unless rules: rules sometimes are expressed using the "unless" phrase, e.g.,

$$y\ is\ abnormal\ unless\ x_1\ is\ A_{i1}\ and\ \dots and\ x_n\ is\ A_{im}$$

This type of rule can be converted to standard form using logical operations, e.g.,

$$If\ not\ (x_1\ is\ A_{i1}\ and\ \dots and\ x_n\ is\ A_{im})\ then\ y\ is\ abnormal$$

Quantifier rule: Some rules include "some" or "all" phrases. This form of rules is called a quantifier rule and can be reformed to a standard form of rule. For rules with "some" quantifier, union is applied to the antecedent or consequent and for rules with the "all" quantifier intersection is applied to the antecedent or consequent which "all" applies.

Rules can be categorized based on their consequent type. There are three common types of rules based on their consequent in a fuzzy rule-based system, as cited in (Mendel, 2001):

For a given $x = (x_1, x_2, \dots, x_n)$ in an n-dimensional pattern space, the *ith* fuzzy rule (R_i) can be one of the following types:

1. Mamdani rule type (Mamdani, 1977):

$$Rule\ R_i : If\ x_1\ is\ A_{i1}\ and\ \dots and\ x_n\ is\ A_{in}\ then\ y = B_i$$

where $A_{i1}, \dots, A_{in}$ are antecedent fuzzy sets and B_i is consequent fuzzy set associated to linguistic terms, y is an output variable, and B_i is a consequent fuzzy set. An example of this type is "If x_1 is small and x_2 is small then y is large"

2. Takagi and Sugeno rule type with a linear function in its consequent (Takagi and Sugeno, 1985):

$$Rule\ R_i : If\ x_1\ is\ A_{i1}\ and\ \dots and\ x_n\ is\ A_{in}\ then\ y = f_i(x)$$

where $A_{i1}, \dots, A_{in}$ are antecedent fuzzy sets associated to linguistic terms, y is an output variable, and $f_i(x) = b_{i0} + b_{i1}x_1 + \dots + b_{in}\,x_n$ *and* b_{ij} is a real number $(j = 1, 2, \dots, n)$. It is clear that because of the linear function in consequent part of rule, this type of rule has less interpretability than previous type. An example of a fuzzy rule of this type is "If x_1 is small and x_2 is small then $y=0.2+0.5x_1-0.3x_2$".

There is also a simplified version of Takagi-Surgeon fuzzy rules as follows:

$$Rule\ R_q : If\ x_1\ is\ A_{i1}\ and\ \dots and\ x_n\ is\ A_{in}\ then\ y = b_i$$

where b_i is a consequent real number. One can consider this simplified type as another version of Mamdani's rule type (Mamdani, 1977). Because of the difficulty in interpretation of the consequent real number, this type of rule is not very interpretable. An example of this type is If x_1 is small and x_2 is small then y=0.25

3. *Ishibuchi rule type* with a class label, C_i, in the consequent part (Ishibuchi, Nakashima and Murata, 1996) :

$$Rule\ R_i\!: If\ x_1\ is\ A_{i1}\ and\ \dots and\ x_n\ is\ A_{in}\ then\ Class\ C_i$$

where $A_{i1}, \dots, A_{in}$ are antecedent fuzzy sets associated to linguistic terms. There is also another extension of this type which assigns each rule a weight, W_i, or a soundness degree which shows the degree of certainty of the rule (Chi, Yan and Phạm, 1996):

$$Rule\ R_i : If\ x_1\ is\ A_{i1}\ and\ \dots and\ x_n\ is\ A_{in}\ then\ \ Class\ C_i\ with\ W_i$$

An example of this rule type is: If x_1 is small and x_2 is small then Class 2 with 0.75.

4. *Fuzzy rules with a certainty degree for all classes in the consequent* (Cordón, del Jesus and Herrera, 1999). This type of rule shows a membership of pattern to different classes in its consequent part:

$$Rule\ R_i : If\ x_1\ is\ A_{i1}\ and\ \dots and\ x_n\ is\ A_{in}\ then\ \ r_{i1}, \dots, r_{iM}$$

where M is the number of classes and r_{ij} shows the membership degree of the given pattern $\boldsymbol{x}$ to class C_j.

3.5 Fuzzy inference engine

An inference engine uses fuzzy logic to combine rules. Each rule in a fuzzy logic is a relation between input sets and output fuzzy set. This relation maps input fuzzy sets $(A_1 \times \dots \times A_n)$ in the antecedent to output fuzzy set B in the consequent as follows:

$$R_i : A_{i1} \times \ldots \times A_{in} \rightarrow B_i = F_i \rightarrow B_i \qquad (3.2)$$

The membership function of R_i for input $x = (x_1, x_2, \ldots, x_n)$ is described as $\mu_{R_i}(x, y)$ as follows:

$$\begin{aligned} \mu_{R_i}(x, y) = \mu_{F_i \rightarrow B_i}(x, y) &= \mu_{A_{i1} \times \ldots \times A_{in} \rightarrow B_i}(x, y) \\ &= \mu_{A_{i1} \times \ldots \times A_{in}}(x) * \mu_{B_i}(y) \\ &= \mu_{A_{i1}}(x_1) * \ldots * \mu_{A_{in}}(x_n) * \mu_{B_i}(y) \\ &= \left[T_{j=1}^{p} \mu_{A_{j1}}(x_j)\right] * \mu_{B_i}(y) \end{aligned} \qquad (3.3)$$

where multiple antecedents are connected using t-norm (T); * denotes t-norm operator.

For a given input vector $I_j = (x_{j1}, x_{j2}, \ldots, x_{jn})$, the compositional rule of inference determines the fuzzy set $C = I_j(x) \circ R_i(x, y)$, where $\circ$ denotes the compositional operator, Figure 3.3 illustrates the fuzzy inference process.

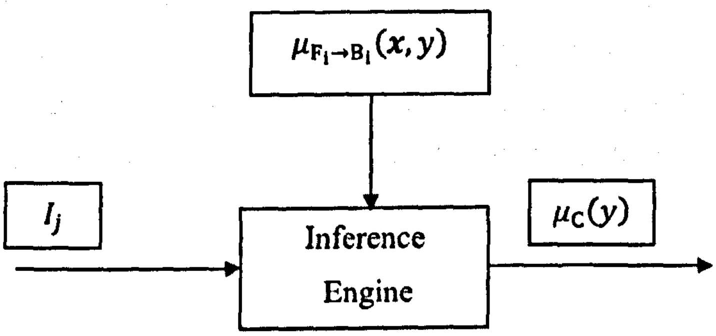


Figure 3.3: Type-1 fuzzy inference engine

The Mamdani inference model is the most commonly used method for fuzzy logic systems (Mendel, 2001). This inference model applies min as the rule implication operator and max-min as the rule composition operator as,

$$\begin{aligned} \mu_C(y) = \mu_{I_x \circ R_i}(y) &= max_{j=1..n}\left[\mu_{I_{x_j}}(x_j) * \mu_{F_i \rightarrow B_i}(x, y)\right] \\ &= max_{j=1..n}\left[\min(\mu_{I_{x_j}}(x_j), \mu_{F_i \rightarrow B_i}(x, y))\right] \end{aligned} \qquad (3.4)$$

The output of the inference engine is a fuzzy set C as,

$$C = I_x \circ [\mathrm{R}_1, \mathrm{R}_2, \ldots, \mathrm{R}_r]$$
$$= \bigcup_{i=1}^{r} I_x \circ R_i \tag{3.5}$$

where r is number of rules.

3.6 Fuzzy set operations

Set operation on ordinary sets can be extended to fuzzy sets. The set operations on fuzzy sets are computed using t-norm (triangular norm) or t-conorm (triangular co-norm) operators. The t-norm is a binary operator ($*$) on [0 1] which satisfies the following properties (Hájek, 1998):

1) *Commutative*, $a * b = b * a$
2) *Associative*, $(a * b) * c = a * (b * c)$
3) *Monotonic*, $(a * b) \leq (c * d)$ if $a \leq c$ and $b \leq d$
4) *Boundary conditions*, $(0 * a) = 0$, and $(a * 1) = a$

The t-conorm operator must satisfy the same conditions as t-norm except unit element, i.e., $(a * 1) = a$. The most common form of t-norm operation is product t-norm or minimum t-norm. The common t-conorm operations are summation and maximum.
Intersection, union and complement of two fuzzy set A and B are defined as follows (Zadeh, 1965):

Intersection:

$$(A \cap B)(x) = min\,[A(x), B(x)] \tag{3.6}$$

Union:

$$(A \cup B)(x) = max[A(x), B(x)] \tag{3.7}$$

Complement:

$$A'(x) = 1 - A(x), \text{ and } B'(x) = 1 - B(x) \tag{3.8}$$

Figure 3.4 demonstrates examples of intersection, union and operation on two fuzzy sets A and B.

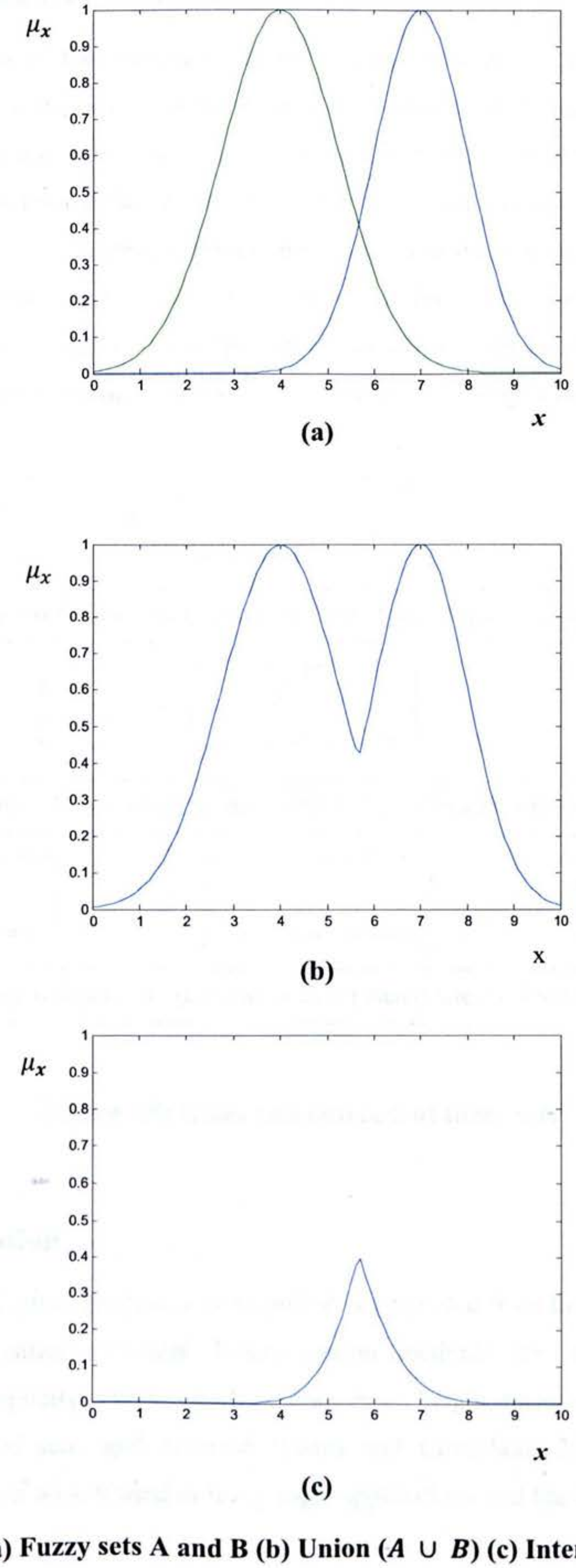


Figure 3.4: (a) Fuzzy sets A and B (b) Union ($A \cup B$) (c) Intersection $A \cap B$

3.7 Example of fuzzy logic inference

For illustration of the inference engine process, consider a fuzzy logic for lung nodule detection with two input fuzzy sets of sphericity and contrast. As shown in Figure 3.5, there are three Mamdani fuzzy rules in the system.Each row shows a linguistic rule and related fuzzy sets invoved in that. Each column is related to one of the input features. For example, column one is related to fuzzy sets of sphericity feature of candidate objects which is modelled using three linguistic terms (Weak, Moderate and Strong). Last column is the output fuzzy set whcich can be modeled using three linguistic fuzzy sets ("Probably not nodule", "Likely nodule", "very likely nodule").

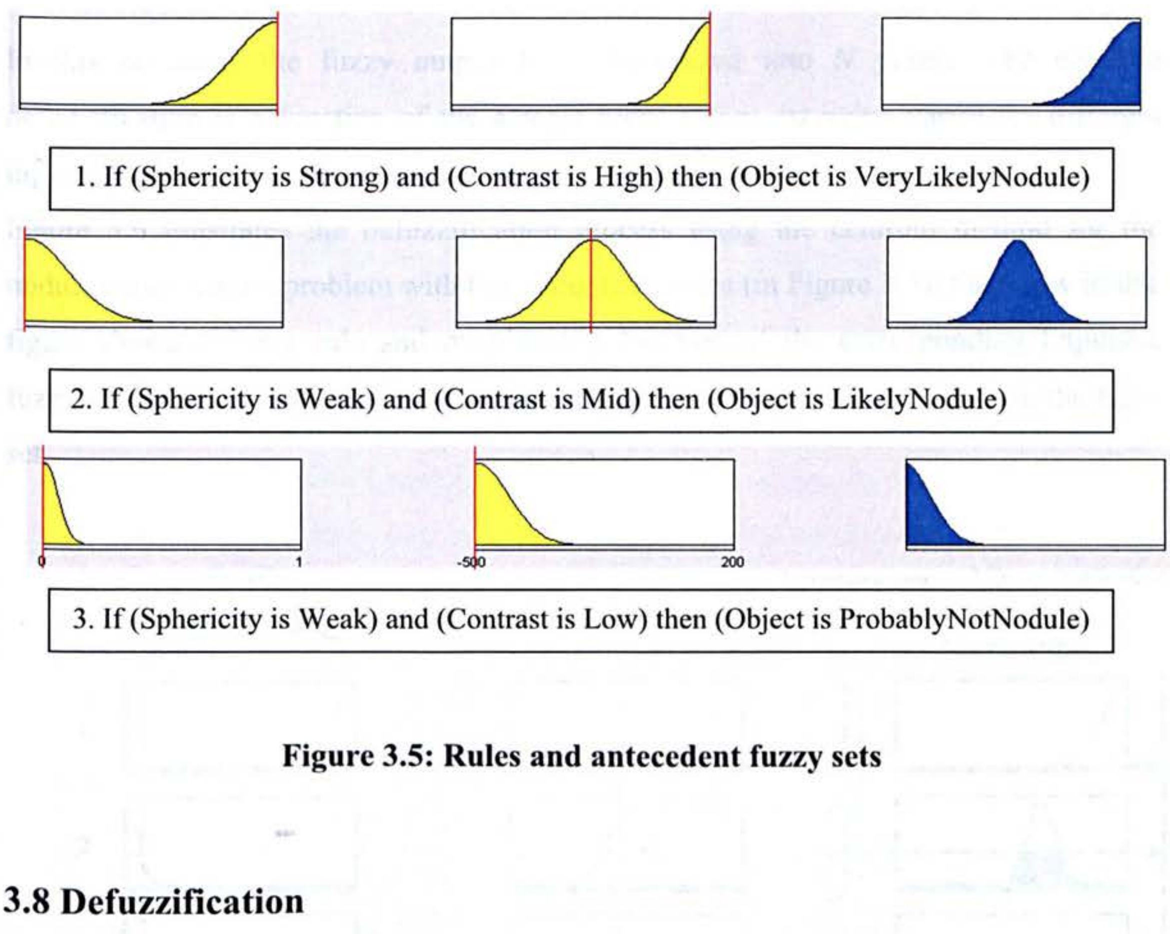


Figure 3.5: Rules and antecedent fuzzy sets

3.8 Defuzzification

In the defuzzification process, a crisp output is generated from the fuzzy set obtained after rule composition. Various defuzzification methods are proposed based on computational simplicity criterion such as: maximum, height, mean of maximum, centre of sum, centre of sets, and centroid (Pham and Castellani, 2002). The centroid defuzzifier has been widely used in fuzzy logic applications and has been utilized in this study.

In the centroid defuzzifier (Mendel, 2001), all output fuzzy sets resulting in fuzzy implication are combined using union as follows,

$$C = \bigcup_{i=1}^{r} C_i \tag{3.9}$$

where $\mu_{C_i}(y)$ is the membership function of the ith rule. The centroid of the union of the output fuzzy sets is as,

$$\mu_{\text{Centroid}}(\boldsymbol{x}) = \frac{\sum_{i=1}^{N} y_i \, \mu_{C_i}(y_i)}{\sum_{i=1}^{N} \mu_{C_i}(y_i)} \tag{3.10}$$

In this equation, the fuzzy output C is discretized into N points. The centroid defuzzification is a function of the sample input $\boldsymbol{x}$, i.e., its value varies for different input vectors.

Figure 3.6 illustrates the defuzzification process using the centroid method for the nodule classification problem with three linguistic rules (in Figure 3.5). Each row in this figure shows a fuzzy rule and membership function of the corresponding linguistic fuzzy set in the antecedent part of the rule and each column shows a feature in the input set.

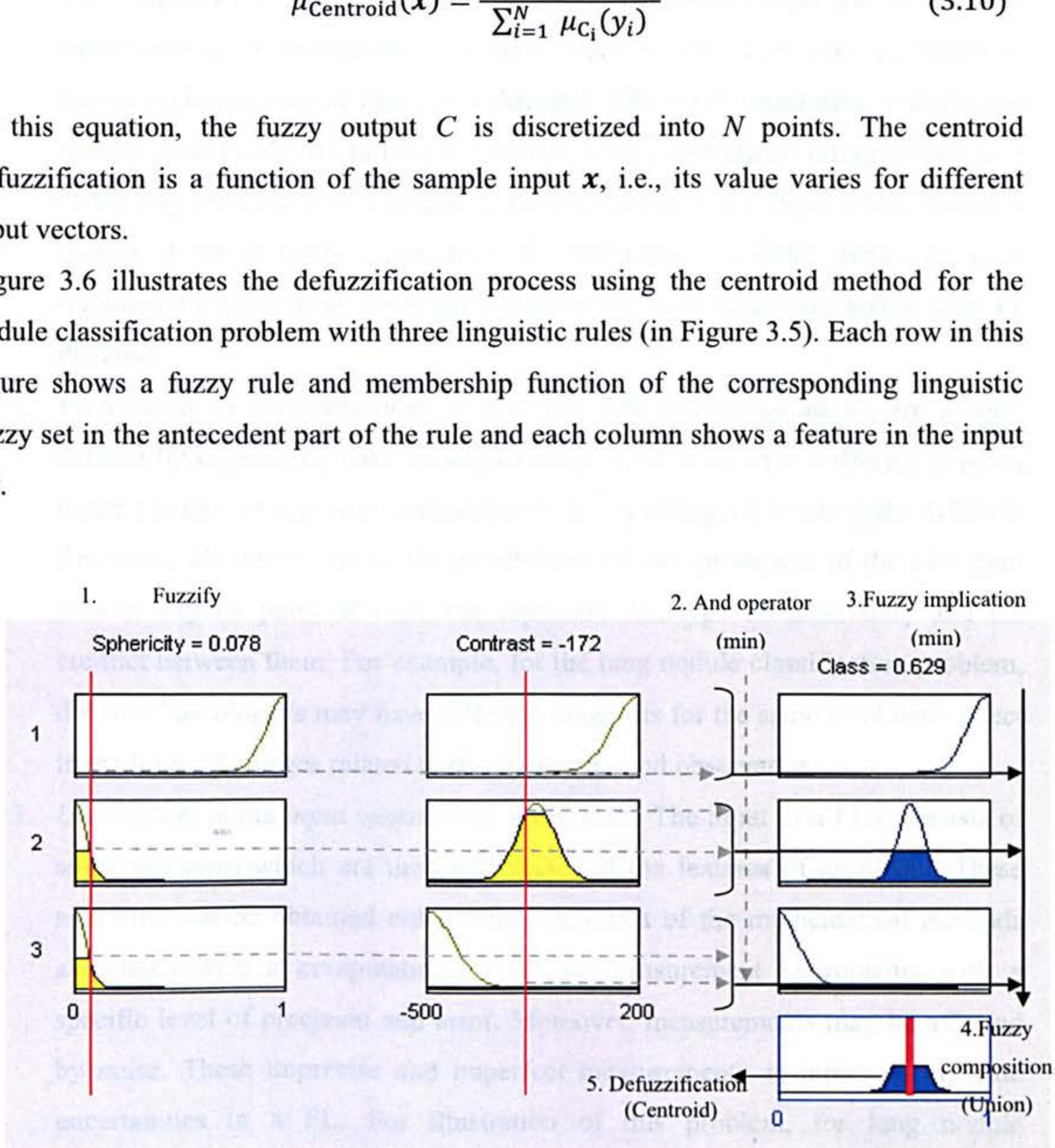


Figure 3.6: Inference engine in a type-1 fuzzy logic system

3.9 Remaining uncertainty issues in an ordinary fuzzy logic

Despite the advantages of ordinary fuzzy sets for modelling uncertainty in comparison to the conventional deterministic methods, some sources of uncertainty still remain which require more consideration. This section explains these uncertainty issues and introduces the method presented to manage them.

The main sources of uncertainty in a FLS are described by Mendel (Mendel, 2001) as follows:

1. *Uncertainty in the meaning of the words which are linguistic terms in the rules:* The concept of words means different things to different people may incorporate some sources of uncertainty in a fuzzy logic system. This issue is related to imprecise boundaries of fuzzy sets (Mendel, 2001). For illustration, radiologists usually classify objects in lung CT images with a confidence rating from 1 to 5 (1 for very unlikely to be a nodule; 2 for unlikely to be a nodule; 3 for possibly a nodule; 4 for probably a nodule; 5 for definitely a nodule). However, each radiologist's perception from the meaning of these linguistic terms may be different.
2. *Uncertainty in the consequent of the rule*: The rules consequents are usually defined by experts, by data mining techniques, or from data. Different experts, either a person or a system) with different levels of expertise may make different decisions. The histogram of the possibilities of the consequent of the rule from various experts point of view can represent this type of uncertainty and the conflict between them. For example, for the lung nodule classification problem, different radiologists may have different diagnosis for the same candidate object in the lung CT images related to their expertise and observations.
3. *Uncertainty in the input measures of the system:* The input of a FLS consists of some measures which are the input vector of the features of an object. These measures can be obtained either from the result of the mathematical methods associated with a computational error or measurement instruments with a specific level of precision and error. Moreover, measurements may be affected by noise. These imprecise and imperfect measurements as inputs incorporate uncertainties in a FL. For illustration of this problem, for lung nodule classification, sphericity is a complex feature of a candidate object which can be measured using different mathematical models such as shape index and dot

enhancement (Ye *et al.*, 2009). However, each metric produces a different number between 0 and 1.

Furthermore, a FL may have some parameters which are required to estimate and tune. One of the common methods for doing this uses a training set and applies learning and tuning techniques. The training set may include measured input-output pairs or linguistic terms expressed by experts. Moreover, learning techniques for tuning and estimating a FL and its parameters suffer from imprecise and imperfect measurements associated with the abovementioned sources of uncertainty.

A type-2 fuzzy set (T2FS) has the potential to overcome uncertainty issues in ordinary type-1 fuzzy sets (Mendel, 2001). Type-1 fuzzy set theory aims at capturing uncertainty about words by assigning a membership function to words. Once the membership function is defined all the uncertainties about the word disappear. On the other hand, T2FS main contribution is to model uncertainties in the type-1 membership function by blurring the boundaries of the membership function. This blurred area creates a bounded region which is called the footprint of uncertainty (FOU). However, the membership function of a type-2 fuzzy set is precise but the FOU provides a degree of freedom for managing the uncertainty, see Figure 3.7.

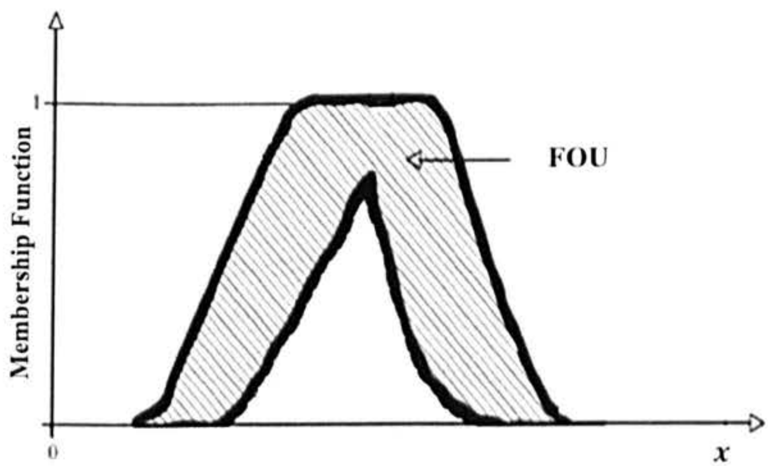


Figure 3.7: Footprint of uncertainty (FOU) in a type-2 fuzzy set

Uncertainty in the consequent of the rules can be managed by modelling the histogram of the possibilities of the consequent of the rule from various experts point of view through the FOU of a type-2 fuzzy set. Figure 3.8 illustrates how the FOU is constructed where there are variations between experts' decisions or measurement results.

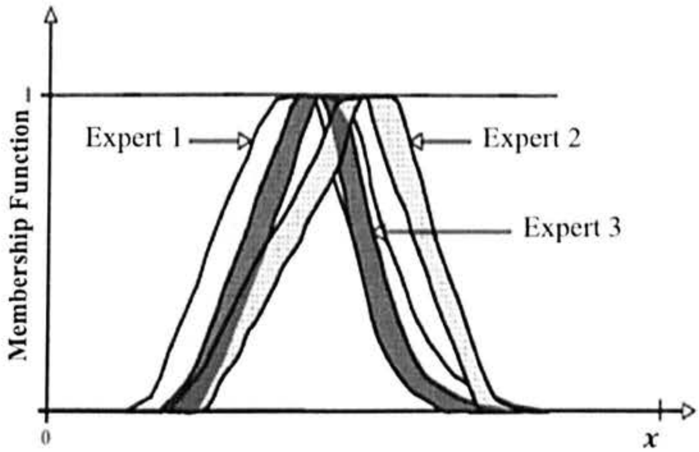


Figure 3.8: Modeling uncertainty in a rule consequent

Uncertainty in the measurement techniques which are latent in the training set can be managed using fuzzy sets. Both type-1 and type-2 fuzzy sets are able to capture uncertainties in the measurements of the applied method. Although, applying different methods (mathematical techniques) with different degrees of precision for the same feature may result in different measures, the histogram of the possibilities of the measurement results can be modelled using type-2 fuzzy membership functions.

3.10 Summary

This chapter presented the importance of fuzzy logic for handling the problem of uncertainty in rule-based classification. The architecture of a FL and its components were explained using theory and concepts of ordinary fuzzy sets. The fuzzy reasoning model was illustrated for handling the problem of uncertainty for the nodule classification of a lung CAD application. Lastly, remaining uncertainty issues in a FL were discussed and illustrated for the nodule classification problem using a FL. Type-2 fuzzy set potential for managing the remaining uncertainty issues in type-1 FL were explained. The next chapter provides an overview of the type-2 fuzzy set concepts, theory and type-2 fuzzy logic components, and the reasoning model. It also represents how type-2 FLS handles the problem of uncertainty for nodule classification in a CAD application.

Chapter 4: Type-2 Fuzzy Logic for Uncertain Rule based Classification

This chapter presents the potential of type-2 fuzzy sets for handling uncertainty issues in rule-based classification. It provides an overview of the theory and concepts of type-2 fuzzy sets and interval type-2 fuzzy set (its practical extension) which has been the focus of this study. The architecture of type-2 fuzzy logic as a rule-based classifier and its main components are reviewed. Type-2 fuzzy set operations and inference model are illustrated for handling the problem of uncertainty for the nodule classification problem in a lung CAD application.

4.1 Introduction

Type-2 fuzzy sets (T2FSs) were introduced for uncertain environments where it is even difficult to define the membership grade in the interval [0, 1]. T2FSs (an extension of the fuzzy set) are known as fuzzy-fuzzy sets in which membership function is not a real number in [0, 1] and is itself a fuzzy set (Mendel, 2001). T2FSs were first introduced by Zadeh in 1975 (Zadeh, 1975) and extended by Karnik and Mendel in 1998 (Karnik and Mendel, 1998). However, T2FSs and ordinary fuzzy sets were emerged at the same time, theory and concepts of it has been extended along with its application in the last decade. The reason for the late application of T2FS is that without extension and application of ordinary type-1 sets, their limitations and challenges and the need for T2FSs would not be clear. Type-2 fuzzy sets have the capabilities to tackle uncertainty issues that are not addressed by ordinary type-1 fuzzy sets (explained in Chapter 3, Section 3.9). Type-2 fuzzy sets are beneficial for classification when dealing with systems with the following properties (Zeng and Liu, 2007; Mendel and John, 2002; Mendel, 2001):

1. *Mathematically ill-defined features:* when the values of the features in the classification vary using different mathematical models or over time
2. *Non-stationary features:* statistical properties of the features in system vary by time
3. *Uncertainty between experts:* knowledge extracted from linguistic terms of a group of experts who may have different perceptions about the meaning of the words
4. *Uncertainty in linguistic terms:* expert knowledge expressed using linguistic terms which are not even easily measurable by the same expert

Mendel introduced uncertain rule-based fuzzy logic systems in 2001(Mendel, 2001). Historical development of T2FL and its applications are demonstrated in (John and Coupland, 2006) and depicted in Figure 4.1. Type-2 fuzzy sets have been applied for handling the problem of uncertainty in medical applications since 1998. The next section provides an overview of the application of type-2 fuzzy sets for managing the uncertainties in medical applications as well as image analysis technologies.

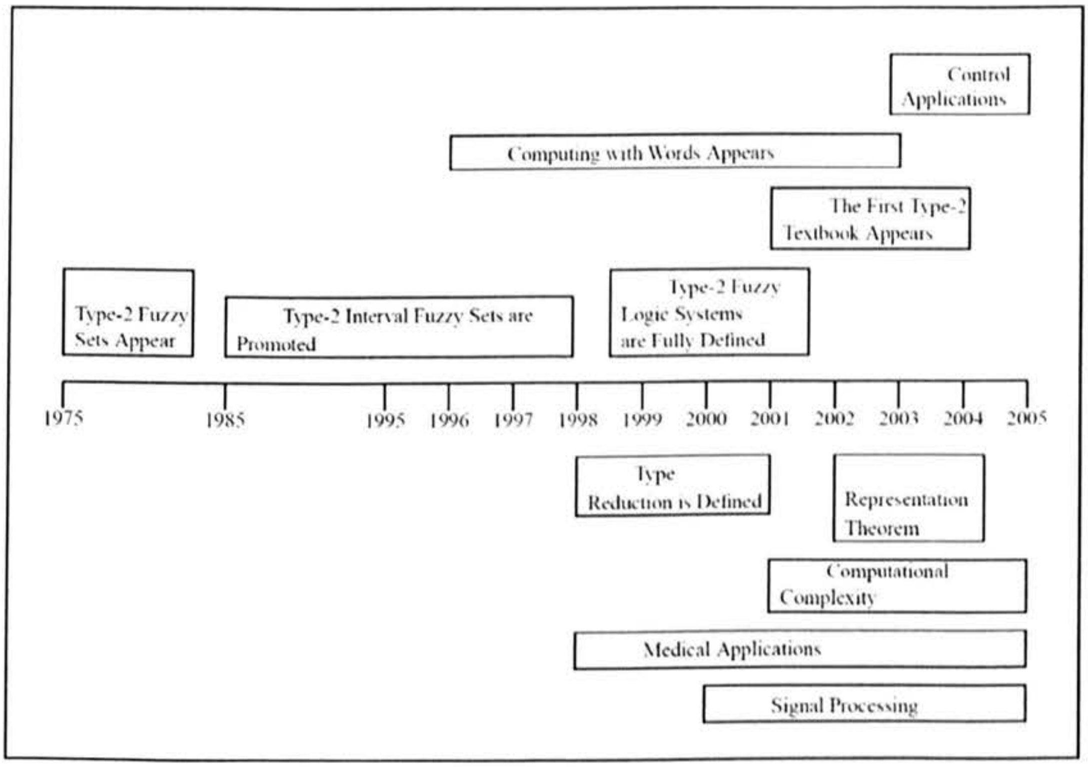


Figure 4.1: Pictorial view of extension of type-2 fuzzy sets theory and applications according to (John and Coupland, 2006)

4.2 An overview of advances of type-2 fuzzy sets in medical imaging applications

The capability of type-2 fuzzy sets for modelling uncertainty has been concentrated in medical diagnosis as well as control and signal processing applications in recent years. The T2FSs application in medical diagnosis has been taking place at the same time of the extension of its theory and it is an open research area (John and Coupland, 2006). The applications of T2FSs in medical imaging applications can be summarized in three main areas: (1) medical diagnosis, (2) image enhancement and segmentation, and (3) pattern recognition, which are briefly reviewed as follows:

1) *Advances in medical diagnosis:* this is one of the main areas applied the general type-2 fuzzy logic where precision in diagnosis is more important than computational complexity. This includes application of type-2 fuzzy sets for modelling intuitive and subjective knowledge of expert and managing inter- and intra- uncertainties in imaging applications. Innocent et. al. applied T2FSs to describe the medical perception of lung scan images in order to predict the pulmonary emboli (Innocent *et al.*, 2001). John et al. (John, Innocent and Barnes, 2000) designed a type-2 fuzzy logic for pre-processing of tibia radiographic images in a Neuro-fuzzy clustering method. Ozen and Graibaldi designed a smart adaptive fuzzy expert system for assessing the health of newborn baby using a type-2 fuzzy system for acid-base balance in blood (Ozen and Garibaldi, 2003). Furthermore, they modelled the inter- and intra- expert variability found in the decision making process using a type-2 fuzzy expert system (Garibaldi and Ozen, 2007). John and Lake (John and Lake, 2001) used type-2 fuzzy sets to model the clinical intuition of nursing staff in terms of linguistic terms. John and Innocent (Innocent and John, 2004) proposed computer aided fuzzy medical diagnosis for modelling uncertainty in medical diagnosis and defining the relation between symptom strength and disease using type-2 fuzzy rules (John and Innocent, 2005; Innocent and John, 2004) . Lascio and Nappi applied type-2 fuzzy sets for modelling differential of the medical diagnosis through linguistic terms related to compatibility of disease and symptoms (Di Lascio, Gisolfi and Nappi, 2005). Recently, Lee et. al. proposed an ontology model based on interval type-2 fuzzy sets for personal diabetic-diet recommendation(Lee, Wang and Hagras, 2010). They constructed an intelligent diet-recommendation agent based on type-2

fuzzy sets, which is able to recommend a menu plan mechanism after diet validation by experts.

2) *Advances in image processing technologies:* Castillo and Melin used T2FSs for modelling uncertainty in edge detection (Castillo and Melin, 2008). Ensafi and Tizhooshe introduced a type-2 fuzzy image contrast enhancement approach (Ensafi and Tizhoosh, 2005). They showed that a type-2 fuzzy logic enhancement method performs better than type-1 fuzzy logic counterpart. Tizhoosh presented a thresholding approach based on type-2 fuzzy sets and a measure of ultra-fuzziness to quantify vagueness (Tizhoosh, 2005). He also proposed a global and a spatial type-2 fuzzy segmentation method (Tizhoosh, 2008).

3) *Advances in pattern recognition:* Zeng and Liu presented the state-of-the-art of type-2 pattern recognition and its success in solving the problem of fuzziness (Zeng and Liu, 2007). According to their work, T2FS framework accounts for randomness in the primary membership function of the feature. It also tackles the fuzziness of the primary membership function using the secondary membership function. Furthermore, the robustness of T2FSs for handling uncertainties in feature space was compared to statistical pattern recognition such as Bayesian methods (Zeng and Liu, 2007). Their results reveal that T2FLS outperforms Bayesian methods and type-1 fuzzy logic for a Chinese character recognition problem. Mitchell proposed a similarity measure for measuring compatibility between two T2FSs and applied it to the problem of automatic detection of welded structures in radiographic images (Mitchell, 2005). Choi and Rhee proposed three methods for automatically defining an internal type-2 membership function from patterns (Choi and Chung-Hoon Rhee, 2009). Their methods are based on heuristic, histograms, and interval type-2 fuzzy C-mean methods. Hidalgo et al. applied an interval type-2 fuzzy inference engine as an aggregation method in a modular neural network with application to biometric patterns. They stated that type-2 fuzzy integration shows better results than type-1 integration method (Hidalgo, Castillo and Melin, 2008).

4.3 Type-2 fuzzy set concepts

This section presents a brief overview of theory and concepts of type-2 fuzzy set. The membership function of a type-1 fuzzy set is a crisp number, whereas in a type-2 fuzzy set (T2FS), the MF is a subset of a fuzzy set and is itself fuzzy. The membership function of a type-2 fuzzy set $\tilde{A}$ (see Figure 4.2), is defined by the function $\mu_{\tilde{A}}(x,u)$, where $0 \leq \mu_{\tilde{A}}(x,u) \leq 1$ and is called the secondary membership function of a type-2 fuzzy set, where $x \in X$ (universe of discourse), and $u \in J_x \subseteq [0,1]$ (Mendel, 2001). Fuzzy set $\tilde{A}$ is defined as (Mendel, 2001):

$$\tilde{A} = \{((x,u),\ \mu_{\tilde{A}}(x,u)) \mid \forall x \in X,\ \forall u \in J_x \subseteq [0,1]\} \tag{4.1}$$

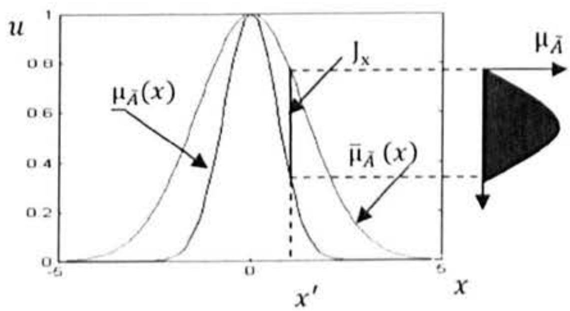


Figure 4.2: Membership function of a type-2 fuzzy set at $x = x'$

A type-2 fuzzy set $\tilde{A}$ can also be defined using the union over all possible paired values of x and u as (Mendel, 2001),

$$\tilde{A} = \int_{x \in X} \int_{u \in J_x \subseteq [0,1]} \mu_{\tilde{A}}(x,u)/(x,u) \qquad J_x \subseteq [0,1] \tag{4.2}$$

where $\iint$ is the union.

At each point $x = x'$, a 2D plane with axes u and $\mu_{\tilde{A}}(x',u)$ is called the vertical axes of $\mu_{\tilde{A}}(x,u)$. A secondary membership function is a *vertical slice* of $\mu_{\tilde{A}}(x,u)$ at $x = x'$ and $\forall\, u \in J_{x'}$ as (Mendel, 2001),

$$\mu_{\tilde{A}}(x = x',u) = \mu_{\tilde{A}}(x') = \int_{u \in J_{\dot{x}}} f_{x'}(u)/u \qquad u \in J_{x'} \subseteq [0,1] \tag{4.3}$$

where $0 \leq f_{x'}(u) \leq 1$.

The domain of a secondary membership function is called the primary membership function of x and $J_{x'}$ in (4.1) is a primary membership function at $x = x'$, (see Figure 4.2). For simplicity $\mu_{\tilde{A}}(x,u)$ can be written as $\mu_{\tilde{A}}(x)$ (Mendel, 2001). The membership function of an n-dimensional variable $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ is as follows (Mendel, 2001):

$$\mu_{\tilde{A}}(\boldsymbol{x}) = \mu_{\tilde{A}_{x_1}}(x_1) \sqcap \mu_{\tilde{A}_{x_2}}(x_2) \sqcap \ldots \sqcap \mu_{\tilde{A}_{x_n}}(x_n) \tag{4.4}$$

Example 1: Figure 4.3 illustrates the three-dimensional membership function of a type-2 fuzzy set. In this example, $X = \{1,2,3,4,5\}$, $U = \{0,0.25,0.5.0.75,1\}$, *and* $J_1 = \{0,0.25,0.75\}$. For instance, at $(x,u) = (1,0)$, $\mu_{\tilde{A}}(x,u) = 0.5$, the primary membership functions are u=*0*, *0.25*, and 0.75 and the secondary membership function at $x = 1$ is 0.5/0+ 0.1/0.25+0.25/0.75.

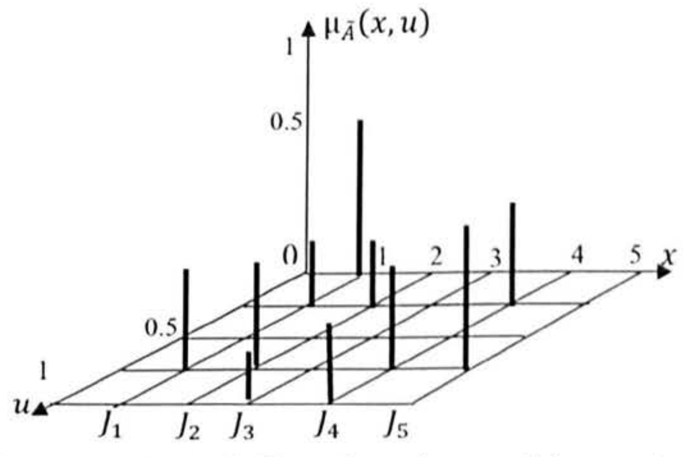


Figure 4.3: Vertical slices view of a type-2 fuzzy set

4.3.1 Footprint of uncertainty

Uncertainty in the primary membership function of a type-2 fuzzy set, $\tilde{A}$ includes a bounded and blurred region which is called the ***Footprint of Uncertainty (FOU)***. It can be written as the union of all primary memberships as follows (Mendel, 2001):

$$FOU\,(\tilde{A}) = \bigcup_{x \in X} J_x \tag{4.5}$$

An ***upper and a lower membership function*** bounds the FOU of a type-2 fuzzy set (see Figure 4.2). The *upper membership function* is a type-1 membership function:

$$\bar{\mu}_{\tilde{A}}(x) \equiv \overline{FOU(\tilde{A})} \qquad \forall x \in X \qquad (4.6)$$

and the *lower membership function* is also a type-1 membership function:

$$\underline{\mu}_{\tilde{A}}(x) \equiv \underline{FOU(\tilde{A})} \qquad \forall x \in X \qquad (4.7)$$

A type-2 membership function is called using the type of its secondary membership function which can be any type such as Gaussian, trapezoidal, triangular or interval. For example, if the type of secondary membership function is Gaussian, it is called Gaussian type-2 membership function.

When the secondary membership functions are interval sets, we have ***interval type-2 membership function*** (Figure 4.4). It means the uncertainty is uniform along its primary membership function and is equal to unity (Mendel, 2001):

$$f_x(u) = 1 \qquad \forall u \in J_x \subseteq [0,1] \qquad (4.8)$$

An interval type-2 membership function can be represented by left and right points of its domain interval $[l, r]$ or by its centre and spread $[c - s, c + s]$ where $c = (l + r)/2$ and $s = (r - l)/2$ (Mendel, 2001). Interval type-2 fuzzy sets are practical type-2 fuzzy sets and have been frequently used because of their low computational complexity and high simplicity. In an interval type-2 fuzzy set $J_x = \left[\underline{\mu}_{\tilde{A}}(x), \bar{\mu}_{\tilde{A}}(x)\right]$ for $\forall x \in X$.

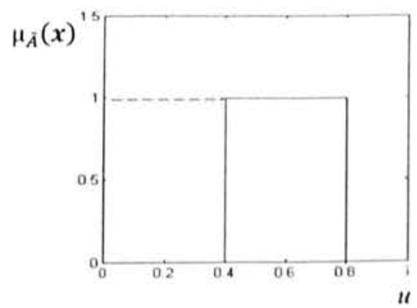


Figure 4.4: Interval type-2 secondary membership function

A type-2 fuzzy set can be considered as a collection of type-2 fuzzy sets which are embedded in the FOU. For a continuous universe of discourse X *and* U, an *embedded* type-2 set $\tilde{A}_e$ is defined as follows (Mendel, 2001):

$$\tilde{A}_e = \int_{x_i \in X} [f_{x_i}(u_i)]/x_i \qquad u_i \in J_{x_i} \subseteq U = [0,1] \tag{4.9}$$

Fuzzy set $\tilde{A}_e$ is embedded in set $\tilde{A}$, and there are uncountable numbers of embedded sets. In an interval T2FLS, an embedded interval type-2 fuzzy set (IT2FS) $\tilde{A}_e$ has N elements $u_1, u_2, \dots, u_N$ with secondary membership equal to 1 as,

$$\tilde{A}_e = \int_{x_i \in X} [1/(u_i)]/x_i \qquad u_i \in J_{x_i} \subseteq U = [0,1] \tag{4.10}$$

An embedded IT2FS $\tilde{A}_e$is also represented using an embedded type-1 fuzzy set A_e , as follows:

$$\tilde{A}_e = 1/A_e \tag{4.11}$$

This study takes the advantages of interval type-2 fuzzy sets to tackle the problem of uncertainty in the training datasets. An interval type-2 fuzzy set with a Gaussian primary membership function is used for generating membership functions and rules in Chapter 5 and 6. The following examples show how uncertainty in the dataset is modelled through Gaussian IT2FSs.

Example 1: Gaussian type-2 primary membership function

Gaussian primary membership functions are one of the frequently applied membership functions in the majority of applications. The maximum likelihood of the measurements in an input dataset approximately follows a normal distribution. However, the normality of an input (x) in a training dataset can be investigated using statistical methods. In this study, the normality of the input features for candidate lung nodules training dataset and Wisconsin breast cancer diagnostic datasets were investigated using Jarque-Bera test (Jarque and Bera, 1987) and "jbest" function in Matlab. A Gaussian primary membership function with fixed standard deviation σ and uncertain mean in $[m_1, m_2]$ is given by (Mendel, 2001):

$$\mu_{\tilde{A}}(x) = \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right] \qquad m \in [m_1, m_2] \tag{4.12}$$

Each value of m results in a different Gaussian membership function. The union of all possible values of $m \in [m_1, m_2]$ constructs the footprint of uncertainty for a Gaussian type-2 membership function. As depicted in Figure 4.5, the shaded area shows the footprint of uncertainty.

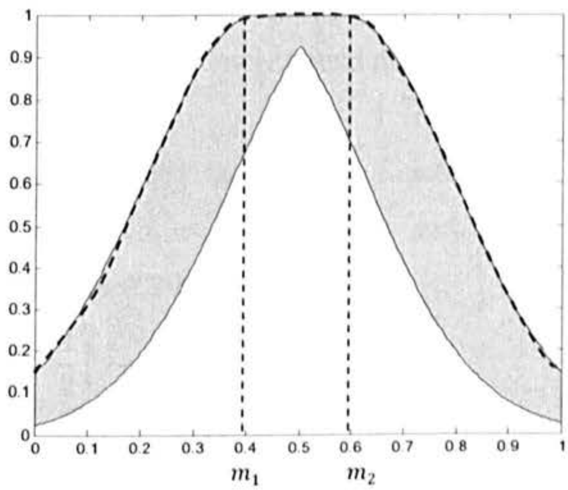


Figure 4.5: Gaussian primary membership function with fixed standard deviation and uncertain mean

A Gaussian primary membership function with fixed mean m and uncertain standard deviation $\sigma \in [\sigma_1, \sigma_2]$ is given by (Mendel, 2001):

$$\mu_{\tilde{A}}(x) = \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right] \qquad \sigma \in [\sigma_1, \sigma_2] \tag{4.13}$$

In a Gaussian primary membership function with fixed standard deviation and uncertain mean, the upper bound function is as follows:

$$\overline{\mu}_{\tilde{A}}(x) = \begin{cases} \exp\left[-\frac{1}{2}\left(\frac{x-m_2}{\sigma}\right)^2\right] & x < m_1 \\ 1 & m_1 < x < m_2 \\ \exp\left[-\frac{1}{2}\left(\frac{x-m_1}{\sigma}\right)^2\right] & x > m_2 \end{cases} \quad (4.14)$$

The dashed line in Figure 4.5 shows the upper bound membership function. The lower bound is as follows (Mendel, 2001):

$$\underline{\mu}_{\tilde{A}}(x) = \begin{cases} \exp\left[-\frac{1}{2}\left(\frac{x-m_1}{\sigma}\right)^2\right] & x \leq (m_1+m_2)/2 \\ \exp\left[-\frac{1}{2}\left(\frac{x-m_2}{\sigma}\right)^2\right] & x > (m_1+m_2)/2 \end{cases} \quad (4.15)$$

The solid line in Figure 4.5 shows the lower bound membership function.

Example 2: In a Gaussian primary membership function with fixed mean (m) and uncertain standard deviation (σ_1, σ_2), the upper and lower bounds membership functions are as follows (Mendel, 2001):

$$\overline{\mu}_{\tilde{A}}(x) = \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma_2}\right)^2\right] \quad (4.16)$$

$$\underline{\mu}_{\tilde{A}}(x) = \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma_1}\right)^2\right] \quad (4.17)$$

where $\sigma_1 < \sigma_2$. The dashed line in Figure 4.6 shows the upper bound membership function whilst the solid line shows the lower bound membership function.

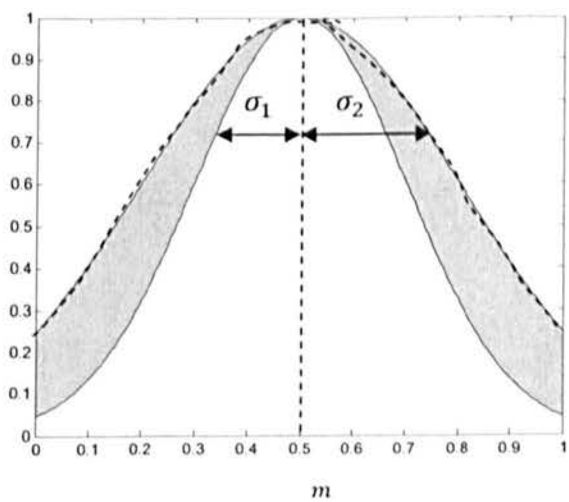


Figure 4.6: Lower bound and upper bound of a Gaussian primary membership function

The next section briefly provides an overview of the set theory operations on type-2 fuzzy sets as well as interval type-2 fuzzy sets. It also provides examples of applying the operations on type-2 fuzzy sets.

4.4 Set theoretic operations on type-2 fuzzy sets

Zadeh presented extension principle for computing set operations on type-2 fuzzy sets in which intersection, union and complement of type-1 fuzzy sets are extended to type-2 fuzzy sets (Zadeh, 1975). Set operations for type-1 fuzzy set are computed using minimum, product, maximum, and negation on crisp numbers while in type-2 fuzzy sets, they are computed on the membership functions which are type-1 fuzzy sets. The operations from $f(x_1, x_2, ..., x_n)$ to $f(A_{x1}, A_{x2}, ..., A_{xn})$ are transferred as follows (Mendel, 2001):

$$f(A_{x_1}, A_{x_2}, ..., A_{x_n}) =$$

$$\int_{x_1 \in X_1} \cdots \int_{x_n \in X_n} \mu_{\tilde{A}_{x_1}}(x_1) \star ... \star \mu_{\tilde{A}_{x_n}}(x_n) \,/\, f(x_1, x_2, ..., x_n) \quad (4.18)$$

Representation theorems: An interval type-2 fuzzy set with discrete X and U can be represented as the union of all of the embedded interval type-2 fuzzy sets as (Mendel, 2001),

$$\tilde{A}_e = \sum_{i=1}^{n_A} \tilde{A}_{e_i} \quad (4.19)$$

where

$$\tilde{A}_{e_i} = \sum_{j=1}^{N} [1/u_{ij}]/x_j \qquad u_{ij} \in J_{x_j} \subseteq U = [0,1] \quad (4.20)$$

and

$$n_A = \prod_{i=1}^{N} M_i \quad (4.21)$$

where M_i is the level of discritization of u_{ij} for all N points, x_j.

The ***join*** $\coprod_{i=1}^{n} \tilde{A}_{x_i}$, of interval type-2 fuzzy sets $\tilde{A}_{x_1}, \tilde{A}_{x_2}, ..., \tilde{A}_{x_n}$ with domains

$[l_1, r_1], [l_2, r_2], \dots, [l_n, r_n]$ is an interval type-2 fuzzy set with domain $[(l_1 \vee l_2 \vee \dots \vee l_n), (r_1 \vee r_2 \vee \dots \vee r_n)]$ where $\vee$ is a maximum (Mendel, 2001).

The ***meet*** $\prod_{i=1}^{n} \tilde{A}_{x_i}$, of interval type-2 fuzzy sets $\tilde{A}_{x_1}, \tilde{A}_{x_2}, \dots, \tilde{A}_{x_n}$ with domains $[l_1, r_1], [l_2, r_2], \dots, [l_n, r_n]$ is an interval type-2 fuzzy set with domain $[(l_1 \wedge l_2 \wedge \dots \wedge l_n), (r_1 \wedge r_2 \wedge \dots \wedge r_n)]$ where $\wedge$ is the product t-norm or minimum (Mendel, 2001).

Using Zadeh's extension principle (Zadeh, 1975), the union, intersection and negation of type-2 fuzzy sets $\tilde{A}$ and $\tilde{B}$ with the membership grades $\mu_{\tilde{A}}(x)$ *and* $\mu_{\tilde{B}}(x)$ for a given x, $\mu_{\tilde{A}}(x) = \sum_i f_x(u_i)$ *and* $\mu_{\tilde{B}}(x) = \sum_j g_x(w_j)$ are defined as follows (N. Karnik and M. Mendel, 2001):

$$\tilde{A} \cup \tilde{B} \Leftrightarrow \mu_{\tilde{A} \cup \tilde{B}}(x) = \mu_{\tilde{A}}(x) \sqcup \mu_{\tilde{B}}(x)$$
$$= \sum_{i,j} \left(f_x(u_i) \star g_x(w_j)\right)/(u_i \vee w_j) \tag{4.22}$$

For interval type-2 fuzzy sets $\tilde{A}$ and $\tilde{B}$, union is defined as follows (Mendel, John and Liu, 2006),

$$\tilde{A} \cup \tilde{B} = 1/[\, \underline{\mu}_{\tilde{A}}(x) \vee \underline{\mu}_{\tilde{B}}(x),\ \overline{\mu}_{\tilde{A}}(x) \vee \overline{\mu}_{\tilde{B}}(x)\,] \quad \forall x \in X \tag{4.23}$$

The intersection of two type-2 fuzzy sets $\tilde{A}$ and $\tilde{B}$ is defined as follows (Mendel, John and Liu, 2006)

$$\tilde{A} \cap \tilde{B} \Leftrightarrow \mu_{\tilde{A} \cap \tilde{B}}(x) = \mu_{\tilde{A}}(x) \sqcap \mu_{\tilde{B}}(x)$$
$$= \sum_{i,j} (f_x(u_i) \star g_x(w_j))/(u_i \wedge w_j) \tag{4.24}$$

The intersection of two *interval type-2 fuzzy sets $\tilde{A}$ and $\tilde{B}$* is:

$$\tilde{A} \cap \tilde{B} = 1/[\, \underline{\mu}_{\tilde{A}}(x) \wedge \underline{\mu}_{\tilde{B}}(x),\ \overline{\mu}_{\tilde{A}}(x) \wedge \overline{\mu}_{\tilde{B}}(x)\,] \quad \forall x \in X \tag{4.25}$$

The complement of a type-2 fuzzy set $\tilde{A}$ is given by:

$$\bar{\tilde{A}} \Leftrightarrow \mu_{\bar{\tilde{A}}}(x) = \sim\mu_{\tilde{A}}(x) = \sum_i f_x(u_i)\,/(1 - u_i) \tag{4.26}$$

The complement of *an interval type-2 fuzzy set $\tilde{A}$* is as,

$$\bar{\tilde{A}} = 1/\left[1 - \overline{\mu}_{\tilde{A}}(x), 1 - \underline{\mu}_{\tilde{A}}(x)\right] \quad \forall x \in X \tag{4.27}$$

where $\vee$ represents the maximum t-conorm, $\wedge$ indicates product t-norm, and $\star$ is the t-norm (minimum or product) and $\sum$ indicates logical union. The operation $\sqcup$ is join, $\sqcap$ is meet and ~ is negative (Mendel, John and Liu, 2006). For a continuum fuzzy set, the operation is the same and the sum is replaced by an integral (Mendel, John and Liu, 2006). An example of the union and intersection of two type-2 fuzzy sets $\tilde{A}$ and $\tilde{B}$ is shown in Figure 4.7

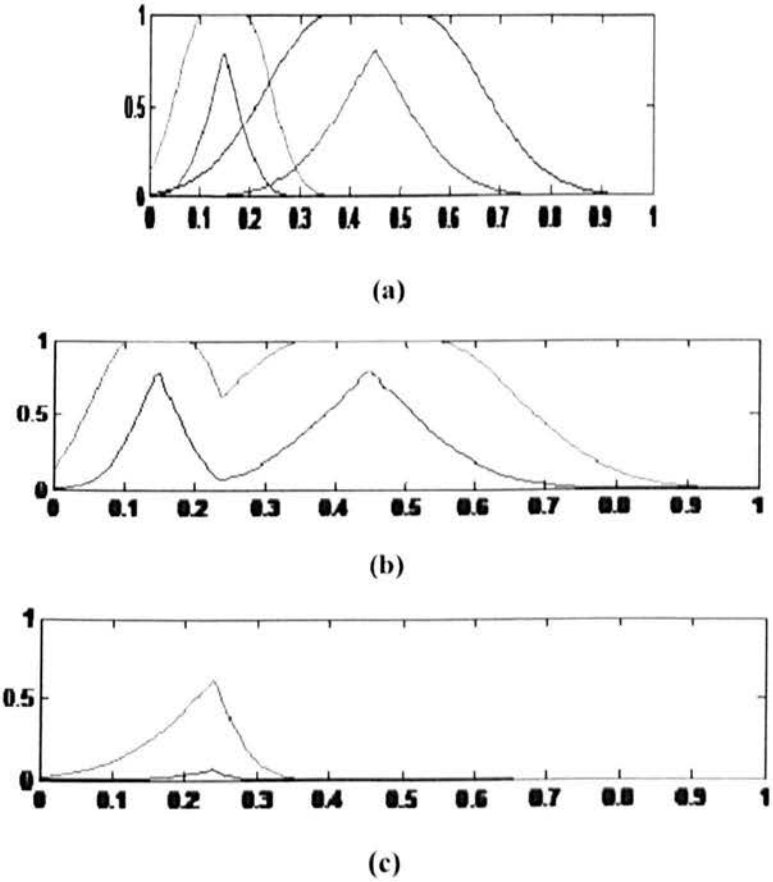


Figure 4.7: Union and Intersection of type-2 fuzzy sets; (a) two type-2 fuzzy sets, (b) union (c) intersection

Example 1: To illustrate the details of set operations on interval type-2 fuzzy sets, consider we are given two interval type-2 fuzzy sets with lower and upper bound values, $\tilde{A}$ = {(1/0.1 + 1/0.3)/ 0.1 , (1/0.1 + 1/0.5)/ 0.2}, and $\tilde{B}$ ={(1/0.1 + 1/0.3)/ 0.2 , (1/0.5 +

1/0.7)/ 0.4 , (1/0.1 + 1/0.4) / 0.6}. The lower and upper bound type-1 membership functions are as follows:

$$\underline{\mu}_{\tilde{A}}(x) = (0.1 / 0.1) + (0.1/ 0.2)$$

$$\overline{\mu}_{\tilde{A}}(x) = (0.3/ 0.1) + (0.5/ 0.2)$$

$$\underline{\mu}_{\tilde{B}}(x) = (0.1 / 0.2) + (0.5/ 0.4) + (0.1 / 0.6)$$

$$\overline{\mu}_{\tilde{B}}(x) = (0.3 / 0.2) + (0.7/ 0.4) + (0.4 / 0.6)$$

Example 2: Consider for an optional point x =0.2, the secondary membership functions are $\mu_{\tilde{A}}(x)$ = (1/0.1 + 1/0.5), and $\mu_{\tilde{B}}(x)$ = (1/0.1 + 1/0.3). The intersection of these two interval type-2 fuzzy sets using equation (4.24) is as follows:

$$\begin{aligned}\mu_{\tilde{A}\cap\tilde{B}}(x) &= \mu_{\tilde{A}}(x) \sqcap \mu_{\tilde{B}}(x)\\ &= [(1/0.1 + 1/0.5) \sqcap (1/0.1 + 1/0.3)]\\ &= \left(\frac{1\wedge 1}{0.1\wedge 0.1} + \frac{1\wedge 1}{0.1\wedge 0.3} + \frac{1\wedge 1}{0.5\wedge 0.1} + \frac{1\wedge 1}{0.5\wedge 0.3}\right)\\ &= 1/0.1 + 1/0.1 + 1/0.1 + 1/0.3\\ &= 1/0.1 + 1/0.3\end{aligned}$$

The intersection of these *two interval type-2 fuzzy sets $\tilde{A}$ and $\tilde{B}$* using equation (4.25) is as follows:

$$\begin{aligned}\tilde{A}\cap\tilde{B} &= 1/[\,\underline{\mu}_{\tilde{A}}(x) \wedge \underline{\mu}_{\tilde{B}}(x),\ \overline{\mu}_{\tilde{A}}(x) \wedge \overline{\mu}_{\tilde{B}}(x)\,]\\ &= 1/\,[0.1\wedge 0.1, 0.5\wedge 0.3] = 1/\,[0.1, 0.3]\end{aligned}$$

From the above result, $\mu_{\tilde{A}\cap\tilde{B}}(x) = 1/0.1 + 1/0.3$. The result of the general intersection equation (4.24) and computing the intersection of IT2FSs using equation (4.25) are the same and is the interval 1/ [0.1, 0.3]. This example illustrates how the computational complexity decreases using IT2FS equations.

Example 3: For the same point x =0.2 and interval type-2 fuzzy set in Example 2, the union of interval type-2 fuzzy sets $\tilde{A}$, and $\tilde{B}$ from (4.23) is as follows:

$$\begin{aligned}\tilde{A}\cup\tilde{B} &= 1/[\,\underline{\mu}_{\tilde{A}}(x) \vee \underline{\mu}_{\tilde{B}}(x),\ \overline{\mu}_{\tilde{A}}(x) \vee \overline{\mu}_{\tilde{B}}(x)\,]\\ &= 1/[0.1\vee 0.1, 0.5\vee 0.3] = 1/\,[0.1, 0.5]\end{aligned}$$

The complement of interval type-2 fuzzy sets $\tilde{A}$ using (4.27) is:

$$\bar{\tilde{A}} = 1/\left[1 - \bar{\mu}_{\tilde{A}}(x), 1 - \underline{\mu}_{\tilde{A}}(x)\right]$$
$$= 1/[1 - 0.5\ ,\ 1 - 0.1] = 1/[0.5, 0.9]$$

The next section gives an overview of the fuzzy reasoning and implications for type-2 fuzzy sets using fuzzy set operations explained in this section.

4.5 Type-2 fuzzy sets relation and implication

Rules in a type-2 FLS denote type-2 fuzzy relations between antecedent and consequent fuzzy sets. A rule can be represented as a fuzzy implication. The general form of a rule in a T2FLS is considered as follows, which is a Mamdani's rule type (Mamdani, 1977) as,

$$R_i : If\ x_1\ is\ \tilde{A}_{i1}\ and\ \dots and\ x_n\ is\ \tilde{A}_{in}\ then\ y\ is\ \tilde{G}_i$$

where $\tilde{A}_{i1}, \dots, \tilde{A}_{in}$ are antecedent type-2 fuzzy sets and $\tilde{G}_i$ is a consequent type-2 fuzzy set, $\tilde{A}_{i1} \in \tilde{A}$ and $\tilde{G}_i \in \tilde{G}$, and $\tilde{A}$ and $\tilde{G}$ are type-2 fuzzy sets, and R_i is the ith rule *(i = 1, 2, ..., M)*. By using Mamdani's implication (Mendel, John and Liu, 2006), a type-2 fuzzy relation maps the input space $X_1 \times X_2 \times \dots \times X_n$ to the output space Y as follows:

$$R_i: \tilde{A}_{i1} \times \tilde{A}_{i2} \times \dots \times \tilde{A}_{in} \rightarrow \tilde{G}_i \tag{4.28}$$

$$R_i: \tilde{A}_{i1} \times \tilde{A}_{i2} \times \dots \times \tilde{A}_{in} \longrightarrow \tilde{G}_i \equiv \tilde{A}_i \longrightarrow \tilde{G}_i \tag{4.29}$$

The Cartesian product of n type-2 fuzzy sets, $\tilde{A}_{i1} \times \tilde{A}_{i2} \times \dots \times \tilde{A}_{in}$, is a type-2 fuzzy set with the following membership function (Mendel, John and Liu, 2006):

$$\mu_{\tilde{A}_{i1} \times \tilde{A}_{i2} \times \dots \times \tilde{A}_{in}}(x_1, \dots, x_n)$$
$$= \mu_{\tilde{A}_{i1}}(x_1) \sqcap \mu_{\tilde{A}_{i2}}(x_2) \sqcap \dots \sqcap \mu_{\tilde{A}_{in}}(x_n) \tag{4.30}$$

where $\sqcap$ is the meet operation. A relation is presented using Cartesian product: thus, rule R_i is a relation between antecedent fuzzy sets $\tilde{A}_{i1} \times \tilde{A}_{i2} \times \dots \times \tilde{A}_{in}$ and the consequent fuzzy set $\tilde{G}_i$ .The membership function of the rule R_i, $\mu_{R_i}(x, y)$, is as (Mendel, 2001; Mendel and John, 2002):

$$
\begin{aligned}
\mu_{R_l}(\boldsymbol{x},y) = \mu_{\tilde{A}_l \to \tilde{G}_l} &= \mu_{\tilde{A}_{l1} \times \tilde{A}_{l2} \times \ldots \times \tilde{A}_{ln} \to \tilde{G}_l}(\boldsymbol{x},y) \\
&= \mu_{\tilde{A}_{l1} \times \tilde{A}_{l2} \times \ldots \times \tilde{A}_{ln}}(\boldsymbol{x}) \sqcap \mu_{\tilde{G}_l}(y) \\
&= \mu_{\tilde{A}_{l1}}(x_1) \sqcap \ldots \sqcap \mu_{\tilde{A}_{ln}}(x_n) \sqcap \mu_{\tilde{G}_l}(y) \\
&= [\sqcap_{j=1}^{n} \mu_{\tilde{A}_{lj}}(x_j)] \sqcap \mu_{\tilde{G}_l}(y) \qquad (4.31)
\end{aligned}
$$

where the meet operation, $\sqcap$, is either the minimum or product t-norms. For a given n-dimensional input $\boldsymbol{x}'$, the membership function $\mu_{\tilde{A}_x}(\boldsymbol{x}')$ is defined in (Mendel, 2001) as,

$$\mu_{\tilde{A}_x}(\boldsymbol{x}') = \mu_{\tilde{A}_{l1}}(x'_1) \sqcap \ldots \sqcap \mu_{\tilde{A}_{ln}}(x'_n) = [\sqcap_{j=1}^{n} \mu_{\tilde{A}_{lj}}(x'_j)] \qquad (4.32)$$

The $\mu_{\tilde{A}_x}(\boldsymbol{x}')$ is a type-1 fuzzy set which is called the *firing set*. The sup-star composition (also called the max-min or max-product composition) of a fired type-2 fuzzy set $\tilde{A}_x$ and a type-2 relation $R_l\ (\boldsymbol{x},y)$ is as (Mendel, 2001),

$$\mu_{\tilde{A}_x \, o \, R_l}(y) = \sqcup_{x \epsilon X} [\mu_{\tilde{A}_x}(x) \sqcap R_l(x,y)] \qquad (4.33)$$

The sup-star composition is used in fuzzy inference engine (Figure 4.8). The composition of fuzzy set $\tilde{A}_x$ and rule R_l determines a fuzzy set $\tilde{B}_l = \tilde{A}_x \, o \, R_l$ (Mendel, 2001; Mendel, John and Liu, 2006), in which

$$
\begin{aligned}
\mu_{\tilde{B}_l}(y) = \mu_{\tilde{A}_x \, o \, R_l}(y) &= sup_{x \epsilon X} \, [\, \mu_{\tilde{A}_x}(x') \star \mu_{\tilde{A}_l \to \tilde{G}_l}(x,y)] \qquad y \in Y \\
&= sup_{x \epsilon X} \, [\mu_{\tilde{A}_{l1}}(x'_1) \sqcap \ldots \sqcap \mu_{\tilde{A}_{ln}}(x'_n) \star [\mu_{\tilde{A}_{l1}}(x_1) \sqcap \ldots \sqcap \mu_{\tilde{A}_{ln}}(x_n)] \star \mu_{\tilde{G}_l}(y)] \\
&= sup_{x \epsilon X} \, [\prod_{j=1}^{n} \mu_{\tilde{A}_{lj}}(x'_j) \star [\prod_{j=1}^{n} \mu_{\tilde{A}_{lj}}(x_j)] \star \mu_{\tilde{G}_l}(y)] \\
&= sup_{x \epsilon X} [\prod_{j=1}^{n} \mu_{\tilde{A}_{lj}}(x'_j) \star \mu_{\tilde{A}_{lj}}(x_j)] \star \mu_{\tilde{G}_l}(y) \\
&= \sqcup_{x \epsilon X} [[\sqcap_{j=1}^{n} \mu_{\tilde{A}_{lj}}(x'_j) \star \mu_{\tilde{A}_{lj}}(x_j)] \star \mu_{\tilde{G}_l}(y) \qquad (4.34)
\end{aligned}
$$

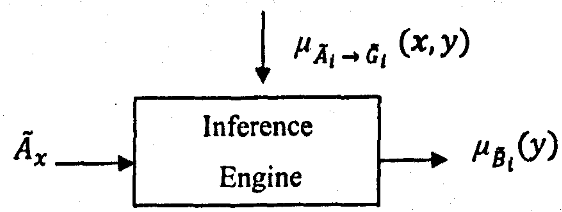


Figure 4.8: Fuzzy inference engine

The type-2 fuzzy set $\tilde{B}_i$ is also called the *fired rule output set* (Mendel and Liu, 2007). The next section explains a method for calculating centroid of the fired rule output type-2 fuzzy set.

4.6 Centroid of a type-2 fuzzy set

The centroid of a type-1 fuzzy set A, C_A, is a crisp number (explained in section 3.8) which is computed as follows:

$$C_A = \frac{\sum_{i=1}^{N} x_i \, \mu_A(x_i)}{\sum_{i=1}^{N} \mu_A(x_i)} \tag{4.35}$$

Using Zadeh's extension principle (explained in section 4.4), the centroid of a type-2 fuzzy set $\tilde{A}$, $C_{\tilde{A}}$, is a type-1 fuzzy set which can be calculated using the centroid of all of its embedded type-1 fuzzy sets as follows (Karnik and Mendel, 2001):

$$C_{\tilde{A}} = \int_{\theta_1 \in J_{x_1}} \cdots \int_{\theta_N \in J_{x_N}} [f_{x_1}(\theta_1) \star \dots \star f_{x_N}(\theta_N)] \Big/ \frac{\sum_{i=1}^{N} x_i \, \theta_i}{\sum_{i=1}^{N} \theta_i} \tag{4.36}$$

This means in order to compute the centroid, the domain of a type-2 fuzzy set is discretized into N points (x_1, ..., x_N) and then the centroid is computed for all the possible combinations of $\{\theta_1, \dots, \theta_N\}$.To every possible combinations of $\{\theta_1, \theta_2, \dots, \theta_N\}$, a membership grade equal to t-norm membership grade of all selected $\theta_i \in J_{x_i}$ is assigned. Each combination demonstrates a type-1 fuzzy set and its centroid is calculated using equation (4.35) where $\mu_A(x_i) = \theta_i$. If two or more combinations of the θ_is give the same point on the centroid, the one with the greater membership degree is kept.

The next section explains the architecture and main components of type-2 fuzzy logic while it utilizes various operations on type-2 fuzzy set explained in Sections 4.3 to 4.6.

4.7 Type-2 fuzzy logic

The general architecture of a T2FL is illustrated in Figure 4.9. There are four main components in a T2FLS (Mendel, 2001): 1) fuzzifier, 2) inference engine, 3) rules, and

4) output producer which includes a type reducer and a defuzzifier. A FLS is called a type-2 fuzzy logic (T2FLS) with at least one type-2 fuzzy set in the antecedent or consequent of its rule structure. When all the antecedent and consequents of the rules are interval type-2 fuzzy sets, the T2FLS is called interval type-2 fuzzy logic system (IT2FLS) which is the focus of this study. The rest of this section explains the role of these components in an IT2FLS.

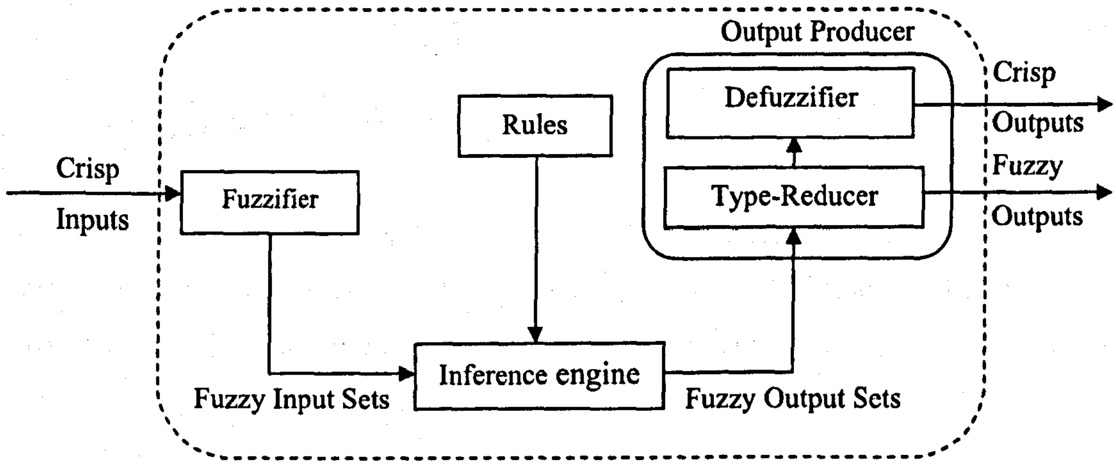


Figure 4.9: Type-2 fuzzy logic architecture

4.7.1 Fuzzifier

The fuzzifier converts crisp inputs into a type-2 fuzzy set, i.e., mapping a crisp point $x_i = (x_{i1}, x_{i2,}, \dots, x_{im})$ into a type-2 fuzzy set $\tilde{A}$ in the universe of discourse X.

4.7.2 Rules and inference engine

Rules represent experts' knowledge and are usually expressed using linguistic terms. The general form of If-Then rules in the type-2 fuzzy Mamdani model is as follows (Mendel, 2001):

$$R_i\text{: } IF\ x_1\ is\ \tilde{A}_{i1}\ and\ \ x_2\ is\ \ \tilde{A}_{i2}\ and\ \ \dots\ x_l\ is\ \tilde{A}_{in}\ then\ y\ is\ \tilde{G}_i \qquad (4.37)$$

where $i = 1, \dots, M$ is the number of rules, $x_i = (x_1, x_{2,}, \dots, x_n)$ is a given sample, $\tilde{A}_{i1}, \tilde{A}_{i2}, \dots, \tilde{A}_{in}$ are antecedent type-2 fuzzy sets, y is an output variable, and $\tilde{G}_i$ is a consequent type-2 fuzzy set.

In a type-2 FLS, the inference engine combines rules and gives a mapping from type-2 fuzzy input sets to type-2 fuzzy output sets. Multiple antecedents in the rules are connected using the t-norm equations (4.30) and (4.32). The membership values in the input sets are combined with those in the output sets using the sup-star composition, explained in section 4.5 equations (4.33) and (4.34) (Mendel, 2007a). Multiple rules may be combined using the t-conorm operation and the union of the firing rule output sets ($\tilde{B}_i$) using equation (4.34) as follows (Mendel, 2007a),

$$\tilde{B} = \bigcup_{i=1}^{M} \tilde{B}_i \tag{4.38}$$

where $\mu_{\tilde{B}} = \sqcup_{j=1}^{M} \mu_{\tilde{B}_i}(y)$ for $\forall y \in Y$.

Example of an IT2FLS Implication: For illustration, consider an interval type-2 fuzzy logic system with two features: (1) contrast ranging from -300 Hounsfield unit (HU) to +300 HU and (2) sphericity ranging from 0 to 1. There are three rules in this IT2FL as follows:

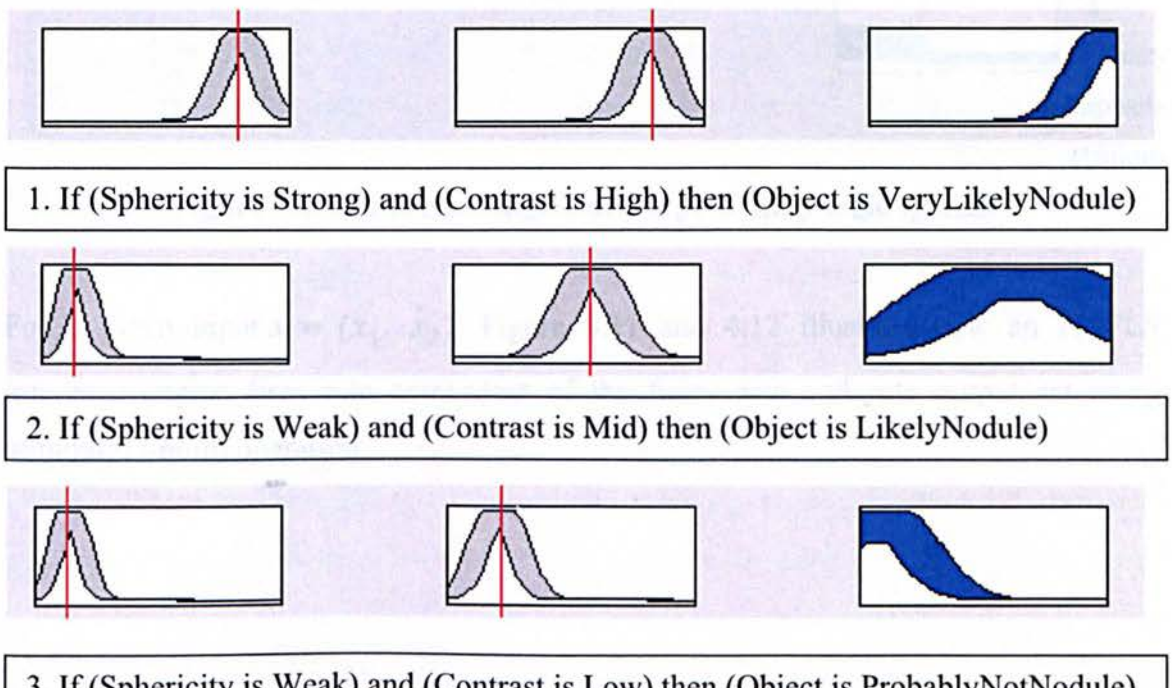


Figure 4.10: Rules in an IT2FL and corresponding antecedent and consequent type-2 fuzzy sets

For a given two-dimensional input sample $\boldsymbol{x} = (0.252, -293)$, the type-2 fuzzy inference first fuzzifies the inputs and defines their membership degrees at $x_1 = 0.252$ and $x_2 = -293$ to the sphericity and contrast fuzzy sets. Then it computes the fired sets using minimum t-norm in equation (4.32). The sup-star composition combines fired input sets with output sets using (4.33) and (4.34). The fired rule output sets are combined using the union operation in equation (4.38).

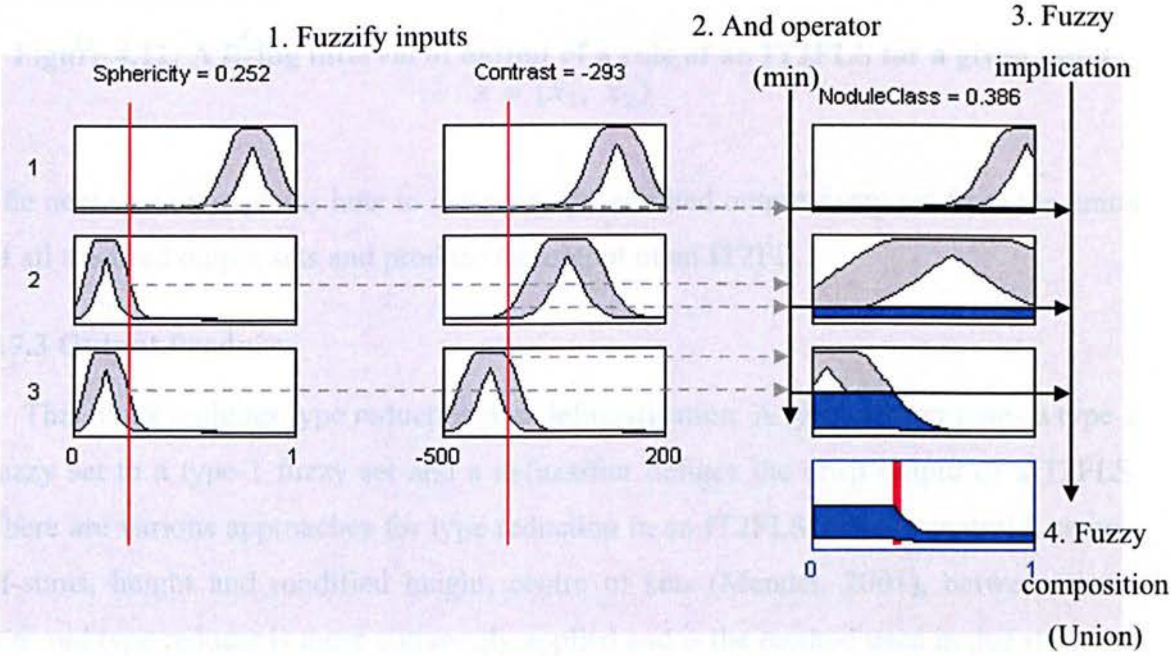


Figure 4.11: Inference engine of a type-2 fuzzy logic system

For a given input $\boldsymbol{x} = (\acute{x}_1, \ \acute{x}_2)$, Figure 4.11 and 4.12 illustrate how an IT2FLS inference engine fires rule antecedent of the fuzzy sets and rule output set using minimum t-norm operation.

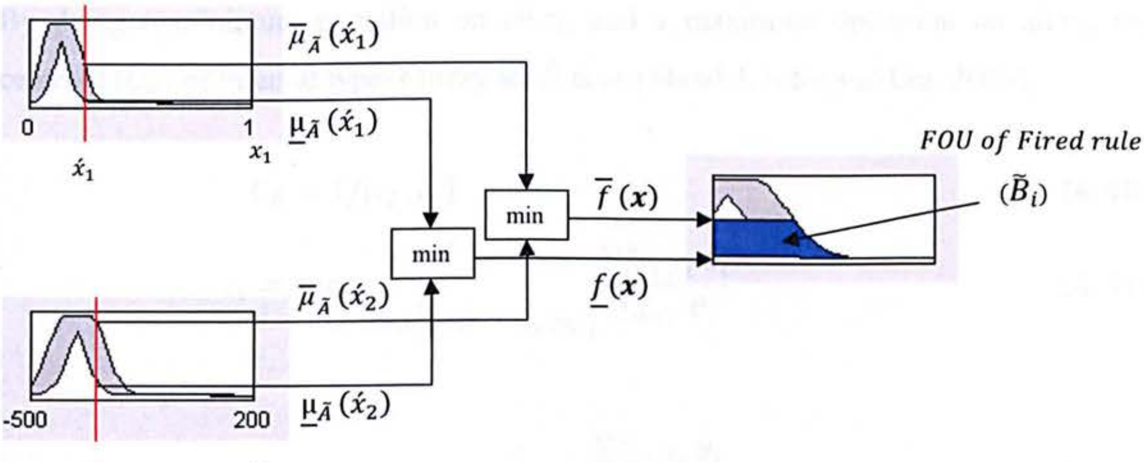


Figure 4.12: A firing interval of output of a rule of an IT2FLS for a given input $x = (\acute{x}_1, \acute{x}_2)$

The next section explains how to defuzzify the resulted output fuzzy set from the union of all the fired output sets and produce the output of an IT2FL.

4.7.3 Output Producer

This block includes type reduction and defuzzification. A type reducer maps a type-2 fuzzy set to a type-1 fuzzy set and a defuzzifier defines the crisp output of a T2FLS. There are various approaches for type reduction in an IT2FLS such as centroid, centre-of-sums, height and modified height, centre of sets (Mendel, 2001), between them, centroid type reducer is more commonly applied and is the method used in this study.

According to (N. Karnik and M. Mendel, 2001; Mendel, 2001), type-reduction is the first task in the process which converts a type-2 fuzzy set obtained from the union of the fired rule output sets, $\tilde{B}$, using equation (4.38), to a type-1 fuzzy set. It follows by applying a defuzzification method on the obtained type-1 fuzzy set.

Using representation theorem (Mendel, 2001), the centroid ($C_{\tilde{B}}$) of a type-2 fuzzy set $\tilde{B}$ is the union of the centroid of all its embedded type-2 fuzzy sets. For an IT2FS with secondary grade 1, the centroid is the union of all n_B embedded type-1 fuzzy sets. These n_A centroids provide a set of crisp numbers with a minimum value c_l and a maximum value c_r and to each centroid crisp number a secondary grade 1 is associated (Mendel, John and Liu, 2006), i.e.,

$$C_{\tilde{B}} = 1/\{c_l, \ldots, c_r\} \tag{4.39}$$

By doing a minimum operation on all c_l, and a maximum operation on all c_r, the centroid ($C_{\tilde{B}}$) of interval type-2 fuzzy set $\tilde{B}$ is as (Mendel, John and Liu, 2006),

$$C_{\tilde{B}} = 1/[c_l, c_r] \tag{4.40}$$

$$c_l = \min_{\forall \theta_i \in [\underline{\mu}_{\tilde{A}}(y_i),\ \bar{\mu}_{\tilde{A}}(y_i)]} \frac{\sum_{i=1}^{N} y_i\, \theta_i}{\sum_{i=1}^{N} \theta_i} \tag{4.41}$$

$$c_r = \max_{\forall \theta_i \in [\underline{\mu}_{\tilde{A}}(y_i),\ \bar{\mu}_{\tilde{A}}(y_i)]} \frac{\sum_{i=1}^{N} y_i\, \theta_i}{\sum_{i=1}^{N} \theta_i} \tag{4.42}$$

The c_l and c_r for an interval type-2 fuzzy sets can be computed as follows:

$$c_l = \frac{\sum_{i=0}^{L} y_i \bar{\mu}_{\tilde{B}}(y_i) + \sum_{i=L+1}^{N} y_i \underline{\mu}_{\tilde{B}}(y_i)}{\sum_{i=0}^{L} \bar{\mu}_{\tilde{B}}(y_i) + \sum_{i=L+1}^{N} \underline{\mu}_{\tilde{B}}(y_i)} \tag{4.43}$$

$$c_r = \frac{\sum_{i=0}^{R} y_i \bar{\mu}_{\tilde{B}}(y_i) + \sum_{i=R+1}^{N} y_i \underline{\mu}_{\tilde{B}}(y_i)}{\sum_{i=0}^{R} \underline{\mu}_{\tilde{B}}(y_i) + \sum_{i=R+1}^{N} \bar{\mu}_{\tilde{B}}(y_i)} \tag{4.44}$$

where L and R are different switch points which are iteratively defined by the KM algorithm (Karnik and Mendel, 2001). The proof of (4.43) and (4.44) and the detail of implementation of the algorithm are explained in (Mendel, John and Liu, 2006). Computational complexity in type reduction can be decreased using approaches which bypass the type-reducer task (Mendel, 2007a) and (Wu and Mendel, 2002). Karnik-Mendel (KM) algorithm is widely used for computing the centroid of interval type-2 fuzzy set (Karnik and Mendel, 2001). This general form of the centroid type-reduction algorithm for a type-2 fuzzy set includes the following steps (Karnik and Mendel, 2001):

1. Compute the membership function of the union of fired output set of rules, $\mu_{\tilde{B}}(y)$.
2. Discretize the domain of $\mu_{\tilde{B}}, y$ into N points $y_1, ..., y_N$.
3. Discretize each primary membership function, J_{y_i}, into M_i points for $i = 1,2, ..., N$.
4. Enumerate all embedded type-1 fuzzy sets of the $\tilde{B}$, there is $\prod_{i=1}^{N} M_i$.

5. Compute the centroid of type-reduced set using the centroid of all the enumerated embedded type-1 fuzzy sets. Assign a membership grade equal to t-norm of the related secondary membership grades to each embedded type-1 fuzzy sets.

However, this algorithm has a high computational complexity but converges rapidly (Mendel, John and Liu, 2006). The customized version of this algorithm for computing the centroid type-reduced set for an interval type-2 fuzzy set using equations (4.43), and (4.44) has considerably fewer computations by assigning secondary membership 1 to all embedded enumerated systems and computing the centroid for the lower and upper bound membership functions.

The output of the defuzzifier is a crisp number which is the centre of the type-reduced set. Defuzzification follows type reduction defines the output by averaging c_l and c_r as follows (Mendel, 200; Mendel, John and Liu, 20067a):

$$y(x) = \frac{1}{2}\left[c_l(x) + c_r(x)\right] \tag{4.45}$$

Example of output producer: Consider for a given input x, the firing output consequent set using product t-norm of the rules is an interval type-2 fuzzy set shown in Figure 4.13. The membership grade consists of two primary membership functions (two embedded type-2 fuzzy set) corresponding to the lower bound and upper bound membership functions. The secondary grade for each embedded type-1 fuzzy set is equal to 1. There are in total four (2×2) combinations of embedded type-2 fuzzy sets. The centroid can be calculated for each embedded type-1 fuzzy set to obtain a point in the type-reduced set. For example considering one of the four possible combinations with primary membership grades 0.4 and 0.2 and secondary grade 1, a corresponding point in type-reduced set is computed as,

$$= (1 \times 1) \,/\, \frac{0.1 \times 0.4 + 5 \times 0.2}{0.4 + 0.2} = 1/1.73$$

which is a type-1 fuzzy set. The type-reduce set completes when all the points are computed. Defuzzification follows type-reduction to obtain a crisp output.

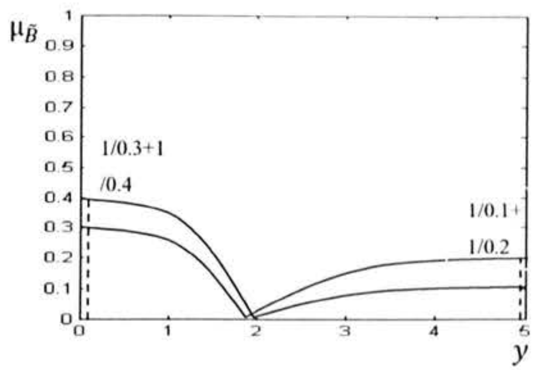


Figure 4.13: Fired consequent sets for three consequent of three rules

4.8 Summary

This chapter provided an overview of the theory and mathematical model of type-2 fuzzy sets. This included the main concepts which are applicable to model uncertainty in the training dataset, the focus was on interval type-2 FLS. The main aim of this chapter was to explore the necessary concepts and theory for designing an IT2FLS and illustrate them through a simple classification example for medical diagnosis. The main drawbacks of type-2 fuzzy set theory is its high computational complexity and the effort needed for learning type-2 fuzzy set concepts and operations on the three dimensional membership functions. However, interval T2FS ameliorates the problem by providing fuzzy operations on type-1 lower and upper bound membership functions which is the reason they have been used in this study for modelling uncertainty in the classification.

The next chapter presents the proposed approach in this study for automatic generation of IT2FMFs from imprecise and incomplete training dataset. The goal is to model uncertainty in the design of an IT2FLS classifier for CAD applications.

Part II

Chapter 5: An Approach for Modelling Uncertainty through Gaussian IT2FMFs for Nodule Classification in a Lung CAD Application

The goal of this chapter is to take advantage of type-2 fuzzy logic for managing high levels of uncertainty in the subjective knowledge of experts or in numerical information in the input of a classification for a lung CAD application. The uncertainties associated with the input of the classifier are modelled using Gaussian interval type-2 fuzzy membership functions (IT2FMFs) and their footprint of uncertainty (FOU). While one of the main challenges in the design of a type-2 fuzzy logic system is how to estimate the parameters of the FOU from imperfect and imprecise training datasets, this chapter presents an automatic approach for learning Gaussian interval type-2 fuzzy membership functions with application to multi-dimensional pattern classification problems. The IT2FMFs and their FOUs are estimated according to the training dataset using a genetic algorithm (GA). This approach is applied to nodule classification component in a lung CAD application. Furthermore, performance comparison (using a ten-fold cross-validation technique) of interval type-2 fuzzy logic with type-1 fuzzy logic for nodule classification in a lung CAD application is presented and the results are discussed in details.

5.1 Introduction

A fuzzy logic system (FLS) is a rule-based classifier with the ability to model a human's subjective decision making model through its linguistic rules. Type-2 fuzzy logic has capability to model uncertainty issues such as inter and intra observer variability, word perception, and uncertainty in imprecise and inadequate input datasets which cannot be managed using ordinary fuzzy sets. The footprint of uncertainty (FOU) in a T2FMF can model uncertainties in linguistic terms and numerical measurements in input of a classification. A type-2 fuzzy logic system (T2FLS) deals with fuzzy inference based on type-2 fuzzy sets. An interval T2FLS (IT2FLS) that is the practical extension of a general T2FLS and simplifies the computational complexity using the lower and upper bound membership functions, has been applied in this study. Integrating a type-2 fuzzy logic model for object classification in CAD architecture allows us to model the accumulative effect of uncertainty sources in the input of a classifier.

An essential factor in design of a nodule classifier for a lung CAD application (as a second reader) is interpretability. The classifier needs to be sufficient comprehensive to interact with medical experts such as radiologists. Fuzzy modelling using linguistic terms is the most comprehensive rule type with a high degree of interpretability between the various rule types explained in Section 3.4 (Castro *et al.*, 2007). A Mamdani rule type in which the antecedent and consequent of rules are linguistic terms is used in this study. On the other hand, interpretability and accuracy are two contradictory factors in the design of a T2FLS (Casillas, 2003). Learning and tuning fuzzy logic system parameters is one of the most frequently applied techniques for improving the accuracy of a classifier and predicting future observations. However, compared to type-1 fuzzy membership functions (T1FMFs), the three-dimensional membership functions (MFs) of a T2FLS give more freedom to the design of a classifier but it has more parameters to adjust, which makes the process of learning and tuning more complicated.

Several approaches have been reported for tuning and learning an IT2FLS in recent years. Most of the proposed methods take advantage of hybrid techniques such as incorporating neural networks or evolutionary approaches (Casillas, 2003) and (Casillas *et al.*, 2005). These learning and tuning IT2FMFs methods are reviewed in the following two main categories:

1) *Based on neural networks:* a Neuro-fuzzy learning method was proposed to learn the parameters of the fuzzy control system which utilizes back-propagation with recursive least squares and square-root filter methods (Mendez *et al.*, 2006). A method for computing derivatives for T2FMF, which are required for back propagation algorithms, was presented in (Mendel, 2004). A Gaussian interval type-2 FLS in a type-2 fuzzy neural network (T2FNN) was demonstrated in (Lee

et al., 2003). This approach utilizes a genetic learning algorithm for optimizing the type-2 FNN parameters. An interval type-2 fuzzy neural network consisting of an IT2FLS in the antecedent part and a two-layer interval neural network in the consequent part was proposed in (Wang, Cheng and Lee, 2004). In this optimal dynamic training method based on GA, learning rates for an T2FNN was suggested to maximize the error reduction during the back propagation process (Wang, Cheng and Lee, 2004).

2) Based on evolutionary techniques: Wu presented a GA learning approach for an IT2FLS that tunes all the parameters of the IT2FMFs without the help of an existing T1FLS (Wu and Wan Tan, 2006). Wagner (Wagner and Hagras, 2007) presented a GA tuning algorithm for evolving type-2 fuzzy logic controllers for autonomous mobile robots. The idea of using a human evolutionary model for optimization of an IT2FMF was discussed in (Sepulveda et al., 2007). An evolutionary approach for the optimal design of the type-2 fuzzy membership functions based on the level of uncertainty was presented in (Hidalgo, Melin and Castillo, 2010). A Hierarchical genetic algorithm (HGA) was suggested to solve the problem of defining the structure of a FLS in (Martinez, Castillo and Garcia, 2008). A fuzzy genetic architecture (FGA) for optimizing the MF parameters of a type-2 Mamdani FLS was presented in (Cazarez-Castro, Aguilar and Castillo, 2010). A particle swarm (PS) optimization technique was applied for the design of type-1 and type-2 Takagi-Sugeno (TKS) FLS for an autonomous mobile robot in (Martínez-Marroquín, Castillo and Soria, 2009). Recently, an optimization method for tuning MF parameters of a type-2 Takagi-Sugeno FLS was proposed which applies ant colony (AC) and particle swarm techniques for optimization of the MF parameters (Castillo et al., 2010). A Takagi–Sugeno interval type-2 fuzzy logic controller for an autonomous mobile robot which utilizes GA for optimization of the membership function generation was presented in (Martínez, Castillo and Aguilar, 2009).

Most abovementioned approaches have been proposed for a TKS FLS with a function in the consequent part of the rules which are not applicable for interaction with system users (Wang, Cheng and Lee, 2004), (Wagner and Hagras, 2007), (Sepulveda *et al.*, 2007), (Martínez-Marroquín, Castillo and Soria, 2009) and (Martínez, Castillo and

Aguilar, 2009). The TKS FLS with a function in consequent part of the rule is more accurate than Mamdani FLS with linguistic term in consequent part but it suffers from interpretability issues. In addition, these approaches initialize the IT2FMF parameters using knowledge of experts. Moreover, they have been applied for a T2FLS with only a few inputs. However, when dealing with a multi-dimensional IT2FLS with large number of inputs, as it is the case in many classification problems, designing the initial IT2FMF parameters and their FOUs for each input using several experts' knowledge is a complicated and expensive process. This fact justifies the need for an automatic IT2FMF learning approach. This chapter presents an automatic approach for learning IT2FMFs which aims at overcoming these issues. For evaluation purpose, the approach has been applied to candidate nodule classification in a lung CAD application.

The rest of this chapter is organized as follows: a learning method for modelling uncertainty in input training dataset using an interval type-2 MF is described in Section 5.2. A genetic approach for parameter estimation of the FOU in an IT2FMFs is presented in Section 5.3. Section 5.4 describes the nodule classification problem in a lung CAD application. The results of applying the proposed interval type-2 MF tuning and learning approach for managing uncertainty sources in the input of a nodule classification for a lung CAD application are demonstrated and discussed in details in Section 5.5; lastly, this chapter is concluded in Section 5.6.

5.2 Learning Gaussian interval type-2 fuzzy membership function

The approach proposed in this study for learning a Gaussian interval type-2 fuzzy membership function (IT2FMF) and its footprint of uncertainty is based on the two following methods:

(1) Learning an IT2FMF based on an T1FLS

(2) Learning an IT2FMF based on the training dataset

The first method is applicable when there is enough expert knowledge for designing T1FMFs. The advantage of this method is that because of the simplicity in the design of the T1FMF, this method has lower complexity and is easier to develop. Moreover, the IT2FLS parameters can be more intelligently estimated based on the T1FLS parameters. The second method is beneficial when there is difficulty in the design of the T1FLS because of the knowledge acquisition issues associated with extracting intuitive information from different experts with different levels of experience. The fact is that in most pattern recognition problems, we rarely have the luxury of access to several experts with different levels of experience.

In this study, a Gaussian IT2FLS is proposed for modelling uncertainties in the input of classifier using an approach for learning IT2FMF. The proposed IT2FLS learning algorithm for the membership function is described in the following subsections.

5.2.1 Fuzzy partitioning of input sets for generating type-1 membership functions

If the first method (in Section 5.2) is chosen, then the T1FLS, including the membership functions, are generated according to the intuitive expert knowledge. Otherwise, a Gaussian interval type-2 fuzzy model is imposed regarding the training dataset. For each of the features in the pattern classification, the training dataset is clustered and fitted into several normal Gaussian functions equivalent to each of the linguistic terms, such as low, middle, and high. For this reason, standard fuzzy clustering methods can be applied (Chi, Yan and Phạm, 1996).

5.2.2 Modeling the footprint of uncertainty

One of the crucial tasks in learning the T2FMF is to model the FOU. For this reason, Genetic algorithms have frequently been used for tuning the lower and upper bound parameters of the FOU (Cazarez-Castro, Aguilar and Castillo, 2010; Wu and Wan Tan, 2006; Castillo *et al.*, 2010). In these GA-based approaches, all the parameters of the membership function are incorporated into the structure of the chromosome. This method is straightforward, although there are some complexities in designing a multi-dimensional classification system. For illustration, consider a multidimensional pattern recognition problem with nine input features and three linguistic terms (low, middle, and high). If each linguistic term is modelled with a Gaussian interval type-2 fuzzy membership function with uncertain means (m_1, m_2) and standard deviations (s_1, s_2), there are 4 parameters (m_1, m_2, s_1, s_2) in each membership function. The total number of parameters of the membership function in the structure of the chromosome is 4*3*9= 108, as shown in Figure 5.1.

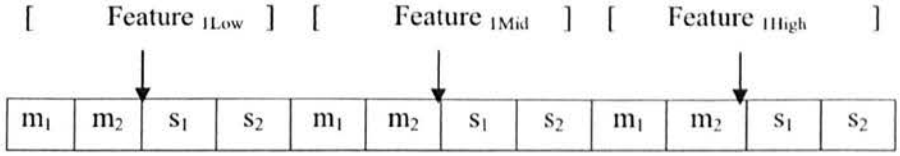


Figure 5.1: Typical structure of the chromosome for a feature

In such a system, approximation of the initial value of the parameters, (lower and upper bound of the means and standard deviations), and grammatical correctness, (e.g. m_1 must be less than m_2), of the genes in the chromosome can be difficult. Moreover, without these considerations, the tuned parameters may result in an inaccurate IT2FLS or one that is divergent to the optimum solution. This fact becomes clear by analyzing the shape and tuned parameters of the IT2FMFs. To tackle these problems, the following approach is proposed for modelling the FOU in a Gaussian IT2FMF.

If the Gaussian T1FMFs are available, then their means and standard deviations can be used; otherwise, a clustering method is used for partitioning the data into three clusters. For each of the T1FMFs, the mean and the standard deviation (STD) are determined from the clustering result. In order to construct the FOU of an IT2FS, the Gaussian distribution properties have been taken into account. In a Gaussian distribution with mean m and standard deviation s, 99.7% of the probability distribution is in the interval $[m - 3s, m + 3s]$ (Zeng, Xie and Liu, 2008). For each of the inputs, three linguistic terms with a Gaussian distribution are considered (as shown in Figure 5.2.). Therefore, for each linguistic term the k_m is considered in interval [0.1, 1] and k_v in interval [0.1, 1]. The lower and upper bound parameters of a Gaussian IT2FMF with mean m and standard deviation s are considered as follows:

$$\overline{m} = m + k_m s \,,\ \underline{m} = m - k_m s, \qquad k_m \in [0.1, 1] \tag{5.1}$$

$$\underline{s} = s \times k_v \,,\ \overline{s} = s \,/\, k_v, \qquad k_v \in [0.1, 1] \tag{5.2}$$

where the k_m and k_v are parameters for tuning the FOU, $\overline{m}$, and $\underline{m}$ are the upper and lower bound of the mean and $\overline{s}$ and $\underline{s}$ are the upper and lower bound of the standard deviation of the IT2FMF. The minimum k_m is considered 0.1 because when the k_m is equal to 0 then $\overline{m} = \underline{m} = m$ which means there is no uncertainty in the system. The bigger k_m and k_v parameters constructs the larger FOU of T2FMS which implies more uncertainty in a T2FLS. These parameters model the FOU of a Gaussian IT2FMF with uncertain mean and standard deviation (Hosseini *et al.*, 2012; Hosseini *et al.*, 2010b).

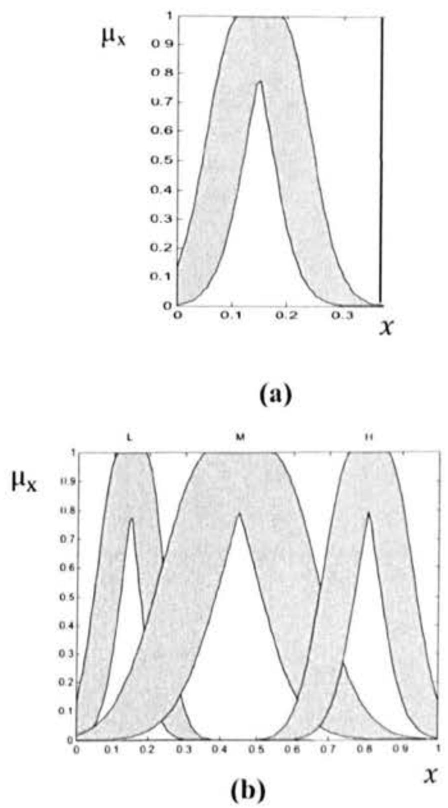


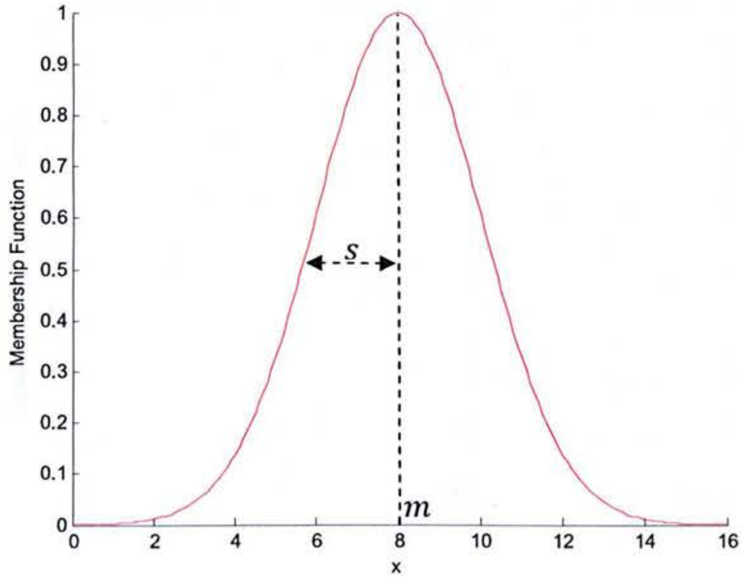
Figure 5.2: Type-2 fuzzy Gaussian membership function of a typical feature; (a) A linguistic term, (b) Three linguistic terms

For illustration of this model, consider a Gaussian membership function with mean $m = 8$ and STD as $s = 2$. In a Gaussian type-2 fuzzy membership function with uncertain mean with the $k_m = 0.9$, the lower and upper bound membership function parameters are as follows:

$$\underline{\mathrm{m}} = 8 - 0.9 \times 2 = 6.2$$

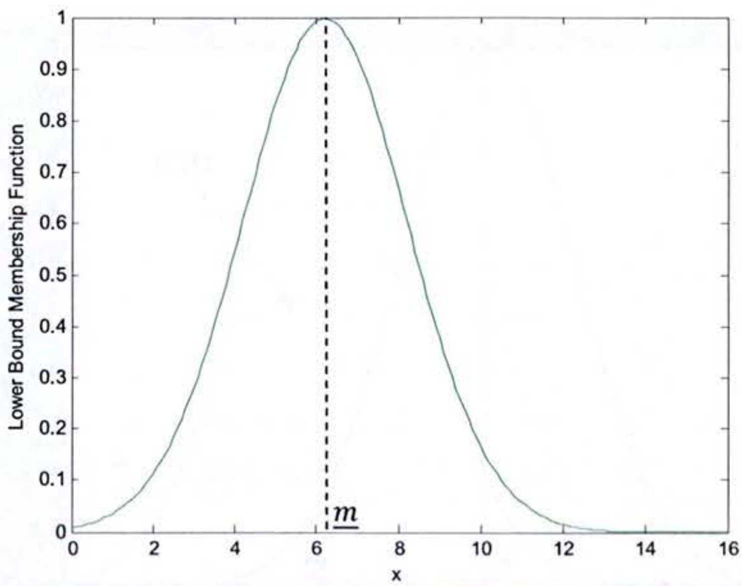
$$\overline{\mathrm{m}} = 8 + 0.9 \times 2 = 9.8$$

Figure 5.3 (a) to (c) illustrate the original Gaussian membership function and lower and upper bound membership functions. The shaded area between lower and upper bound membership function represents the FOU (see Figure 5.3 (d)).



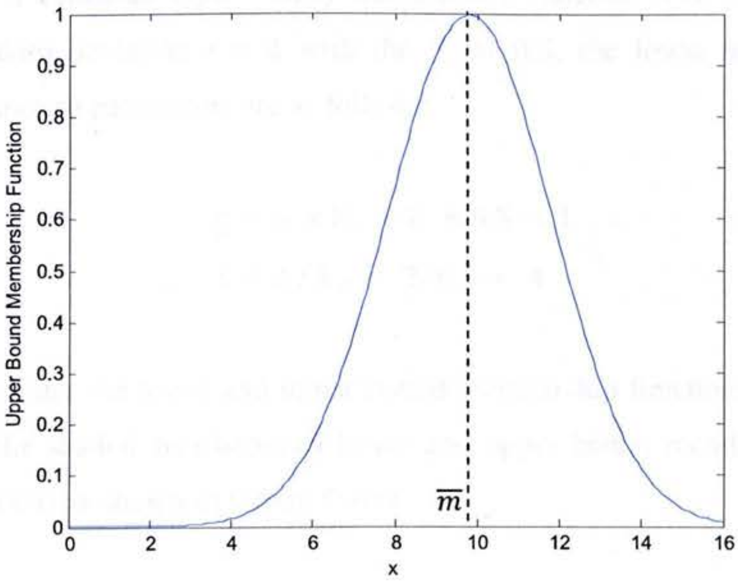
Membership Function
s
m
x
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
14
16

(a)

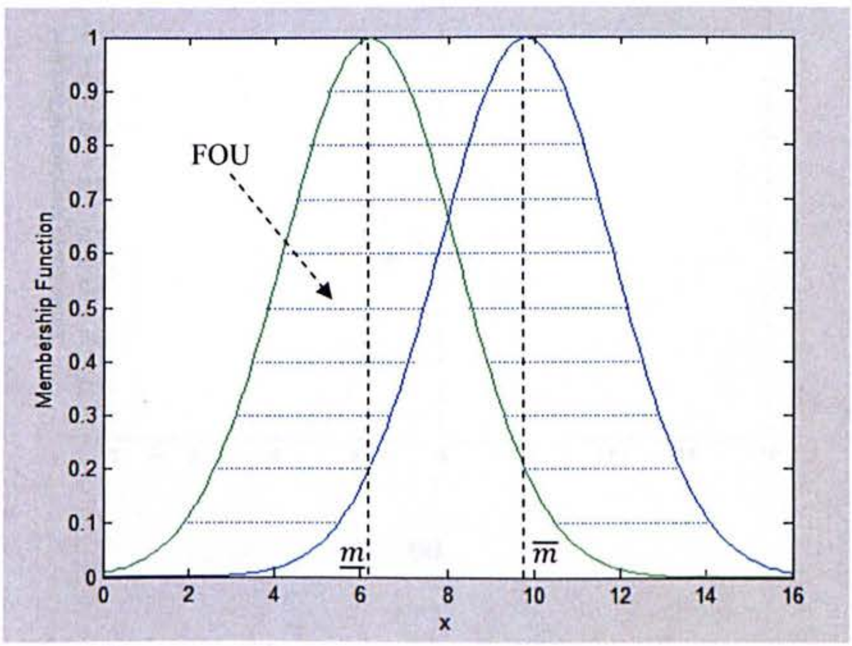


Lower Bound Membership Function
$\underline{m}$
x
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
14
16

(b)



(c)



(d)

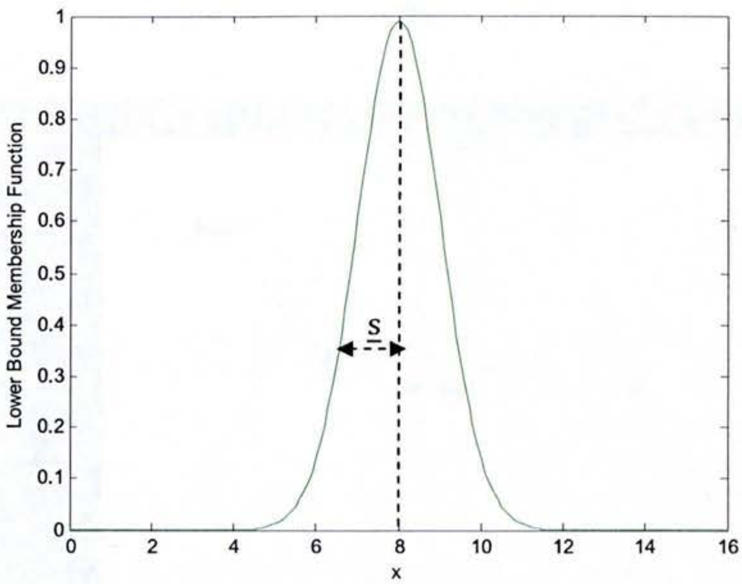
Figure 5.3: An example of a Gaussian membership function with uncertain mean (a) Original membership function with (m, s)=(8, 2); (b) Lower bound membership function with $\underline{\mathrm{m}}$ = 6.2; (c) Upper bound membership function with $\overline{\mathrm{m}}$ = 9.8; (d) The shaded area shows the FOU

Now consider a Gaussian type-2 fuzzy membership function with mean $m = 8$ and uncertain standard deviation $s = 2$ with the $k_v = 0.5$, the lower and upper bound membership function parameters are as follows:

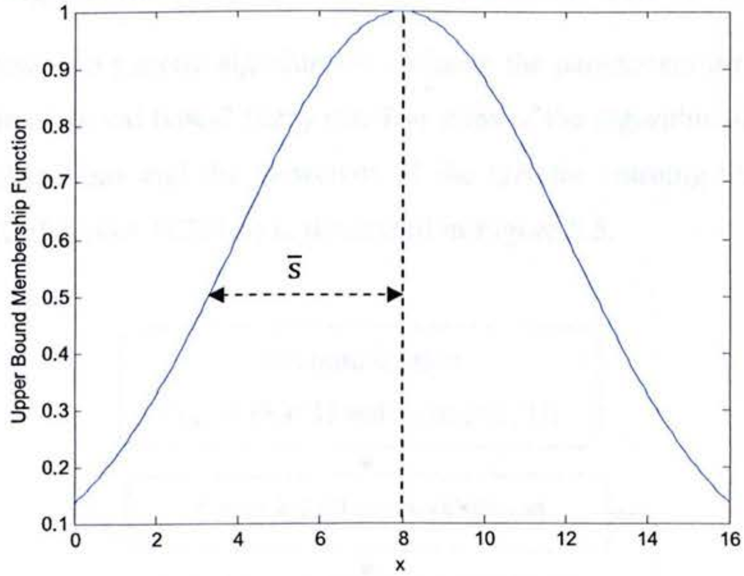
$$\underline{\mathrm{s}} = \mathrm{s} \times \mathrm{k_v} = 2 \times 0.5 = 1 \ ,$$

$$\overline{\mathrm{s}} = \mathrm{s} \,/\, \mathrm{k_v} = 2/0.5 = 4$$

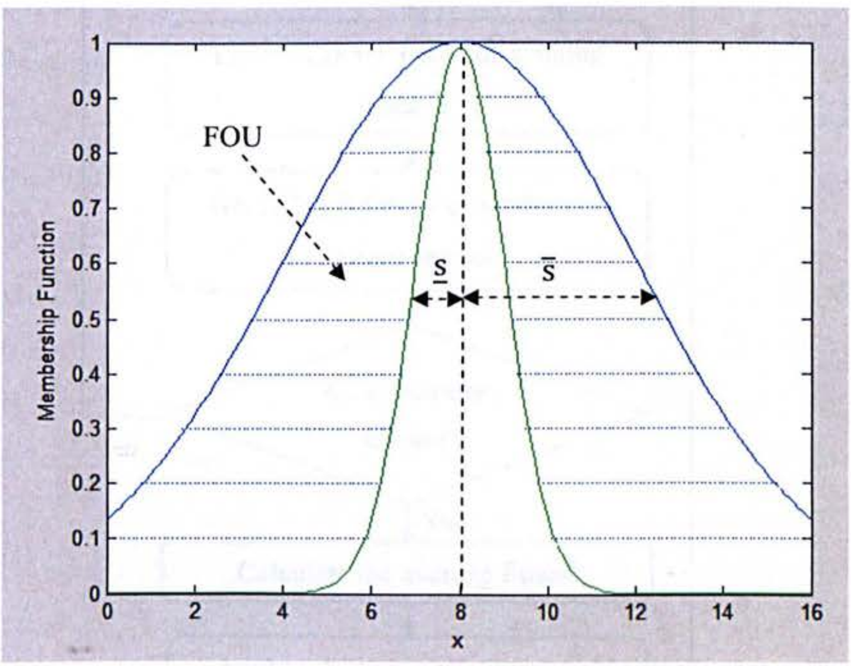
Figure 5.4 illustrates the lower and upper bound membership functions (see Figure 5.4 (a) and (b)). The shaded area between lower and upper bound membership functions represents the FOU as shown in Figure 5.4(c).



(a)



(b)



(c)

Figure 5.4: An example of a Gaussian membership function with uncertain STD; (a) Lower bound membership function with $\underline{s} = 1$; (b) Upper bound membership function with $\overline{s} = 4$; (c) The shaded area shows the FOU

5.3 A genetic algorithm for parameter estimation of the FOU

This section presents a genetic algorithm to estimate the parameters for modelling the FOU of a Gaussian interval type-2 fuzzy set. The steps of the algorithm are described in the following subsections and the flowchart of the GA for learning the membership function of an IT2FLS (GA IT2FLS) is illustrated in Figure 5.5.

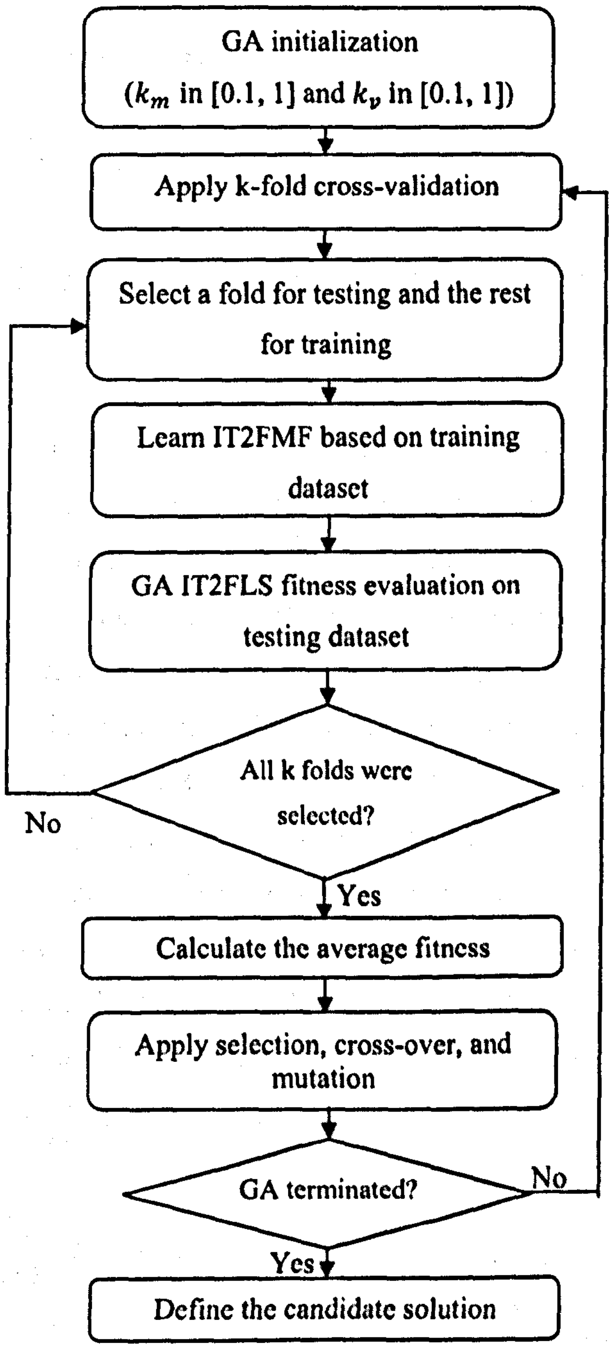


Figure 5.5: Flowchart of the GA learning MFs method for an IT2FLS

5.3.1 GA initialization

In this approach a real-valued genetic algorithm is used to represent genes in the chromosome (Corcoran and Sen, 1994). A pair of (k_m, k_v) can be used to model the FOU for all the features if uncertainty is monotonic; otherwise, for each feature's linguistic terms, a pair of (k_m, k_v) are considered, in the following chromosome structure (Figure 5.6).

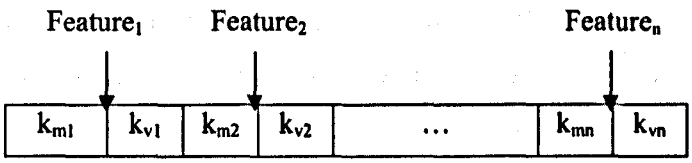


Figure 5.6: The structure of the chromosome

The genetic algorithm initializes the chromosome randomly in the k_m interval [0.1, 1] and the k_v interval [0.1, 1]. Next, an IT2FLS is developed according to the selected k_m and k_v values by applying the learning membership function algorithm proposed in Section 5.3, and the pre-defined set of rules (identified by experts or a learning algorithm). The IT2FMFs are learned using the training dataset and the fitness function is evaluated on the testing dataset. Training and testing datasets are chosen by a cross-validation technique (Picard and Cook, 1984). In the genetic algorithm, chromosomes use real-valued coding representing the k_m and k_v parameters of the membership function for each linguistic term of the corresponding fuzzy sets. Standard tournament selection, single point crossover, and uniform mutation operations are applied (Herrera, 2008). The advantage of the Genetic IT2FMF learning approach is that it has fewer parameters to estimate and defines the initial values of the parameters more precisely. This makes the process of optimization more accurate and easier to implement.

5.3.2 Cross-validation

In order to have a consistent and unbiased view of the performance of a designed IT2FLS classifier, k-fold cross-validation has been applied (Picard and Cook, 1984). This method partitions the dataset into k equal size subsamples randomly. In each algorithm run, $k-1$ folds are considered for training and one fold for testing. Therefore, all the partitions will have a chance to be included in the training and testing process. The k results from the partitions can then be averaged to produce the final estimate. This study applies a ten-fold cross-validation technique ($k = 10$) as it is common in classification applications. In each GA generation, a ten-fold cross-validation is applied. Then, the average fitness of a ten-fold run is considered as the fitness of that generation.

Leave-one-out is another form of cross-validation method which keeps one sample each time for testing reason and the rest of samples for the training and repeats this process until all samples are considered once for testing. Whilst leave-one-out cross-validation (LOOCV) produces more accurate results by giving each sample the chance to get used for validation, repeating the algorithm for the number of samples in the dataset but it is expensive and has high computational complexity. Moreover, employing LOOCV technique for several runs of the GA with a large number of samples deteriorates computational complexity of the cross-validation technique. This is the reason that this study applies a class labl

ten-fold cross-validation.

5.3.3 Fitness evaluation

The IT2FLS is evaluated on the test dataset. In the GA, the area under the ROC curve (AUC) has been used to represent accuracy for the purpose of fitness evaluation. The ROC curve, which depicts all the tradeoffs between the true-positive (tp) and false-positive (fp) rates of the classifier is applied to estimate the cost of classification (explained in Section 1.3). The genetic algorithm selects the best solution of k_m and k_v values for an IT2FLS with the maximum AUC.

5.3.4 Comparison of the GA IT2FMF learning approach with other GA-based methods

This section compares the structure of the chromosome for learning a Gaussian IT2FMF with two major GA based techniques for learning and tuning an IT2FMF in a Mamdani inference model. An n-input one-output IT2FLS classifier with Gaussian

IT2FMF with uncertain mean and standard deviation, with three linguistic terms (L) and a number of rules (R) have been considered in all of the approaches.

The genetic algorithm for IT2FLS begins after generating type-1 fuzzy membership functions based on a training dataset selected by a ten-fold cross-validation technique (explained in Section 5.3.2) and defining the system rules by experts. The GA IT2FMF learning method is described as follows:

1. The GA algorithm initializes the chromosome randomly, with $k_m \in [0.1,\ 1]$ and $\mathrm{k_v} \in [0.1,\ 1]$ for each linguistic input and output type-2 fuzzy set. There are 3 linguistic terms (Low, Middle and High) for each input and output interval type-2 fuzzy set for a classifier with two class labels. For the given IT2FLS, the number of genes in each chromosome is $R \times 3(L) \times n(k_m) + R \times 3(L) \times n(k_v) + R \times 3(L) \times 1(k_m) + R \times 3(L) \times 1(k_v)$.
2. For each generation, the FOU of a Gaussian IT2FMF is estimated using equations (5.1) and (5.2).
3. The fitness of each chromosome is evaluated for IT2FLS using the area under the ROC curve.
4. This procedure (1-3) is repeated in each generation for each fold in the cross-validation technique.
5. The average fitness of the ten-fold cross-validation is considered as the fitness of each chromosome in the current generation.
6. Standard tournament selection, single point crossover, and uniform mutation operations are applied to the individuals in the current generation with a mutation and crossover probability (Herrera, 2008).
7. The GA terminates after reaching the maximum number of generations.

The GA learning of an interval type-2 fuzzy neural network (FNN) begins by initializing a type-1 FNN system (Lee *et al.*, 2003). Then, a back-propagation algorithm trains a type-1 FNN to obtain a set of Gaussian functions (mean, variance) and a weighting vector and adds uncertainty to these parameters. The genes of each chromosome include the MF parameters and the weighting vectors. Each MF contains two $mean$ values (for upper MF and lower MF), two standard deviations (STD), and the weighting vector (W). Therefore, for a given n-input one-output type-2 FNN with R

rules and three linguistic terms (L), the number of genes for each chromosome is $R \times 2(mean) \times 3(L) \times n + R \times 2(STD) \times 3(L) \times n + 2 \times R(W)$ (Lee *et al.*, 2003).

In comparison, *the fuzzy GA learning methodology to evolve the MF parameters of T2FLC* starts by designing the structure of the FLS, including the input and output, the rules, the type of MF, and the parameters of the MF (Cazarez-Castro, Aguilar and Castillo, 2010). Then a chromosome is constructed using the parameters of the MF of each input-output variable. Therefore, for the given n-input one-output Gaussian IT2FLS with R rules and three linguistic terms (L), the number of genes for each chromosome is $R \times 2(mean) \times 3(L) \times n + R \times 2(STD) \times 3(L) \times n + R \times 3(L) \times 2(mean) + R \times 3(L) \times 2(STD)$.

Table 5.1 Comparison of the chromosome structure in the GA IT2FMF learning and tuning methods

No	Method	Number of genes in Chromosome	Initialization
1	GAIT2FLS (this work)	R×3×n + R×3×n+ R×3×2	Automatic
2	IT2FNN (Lee *et al.*, 2003)	R×2×3×n+ R×2×3×n+ R×2	Subjective
3	IT2FGA(Cazarez-Castro, Aguilar and Castillo, 2010)	R×2×3×n+ R×2×3×n+ R×3×2×2	Subjective

Table 5.1 summarizes the comparison results of the structure of chromosome in these three methods. For illustration, consider the lung nodule classification problem with 7 inputs (features) and one output (class) with three linguistic terms (low, middle and high) and 14 rules, we have $n = 7, R = 14, and\ L = 3$. The chromosome structure of the IT2FNN approach includes $14 \times 2 \times 3 \times 7 + 14 \times 2 \times 3 \times 7 + 14 \times 2 = 1204$ genes, the IT2FGA approach includes $14 \times 2 \times 3 \times 7 + 14 \times 2 \times 3 \times 7 + 14 \times 3 \times 2 \times 2 = 1344$ genes whilst the proposed GAIT2FLS membership function learning approach consists of $14 \times 3 \times 7 + 14 \times 3 \times 7 + 14 \times 3 \times 2 = 672$ genes, about half of the two other methods. Moreover, in the proposed method, the initial values of parameters k_m and k_v are automatically defined using equations (5.1) and (5.2), while

the two other approaches require expert subjective knowledge for initialization of the parameters and their grammatical correctness. This can be expensive to acquire and in some circumstances, such as the lung nodule classification problem, may be impossible to achieve as it requires agreement between the experts.

5.4 Nodule classification problem for a lung CAD application

Lung cancer is the most common diagnosed cancer and general cause of cancer deaths in males across the world in 2008 (Ferlay *et al.*, 2011). It is also fourth most diagnosed cancer among females with the second mortality rate (Ferlay *et al.*, 2011). Age-Standardized lung cancer rates (ASR) compared to other cancers by sex is illustrated in Figure 5.7.In this figure, the worldwide incidence and mortality rates of different diagnosed cancers in different aging range up to 60 years in 2008 is depicted according to (Ferlay *et al.*, 2011). As shown in this figure, lung cancer has the largest incidence and mortality rate between men compared to the other cancer types.

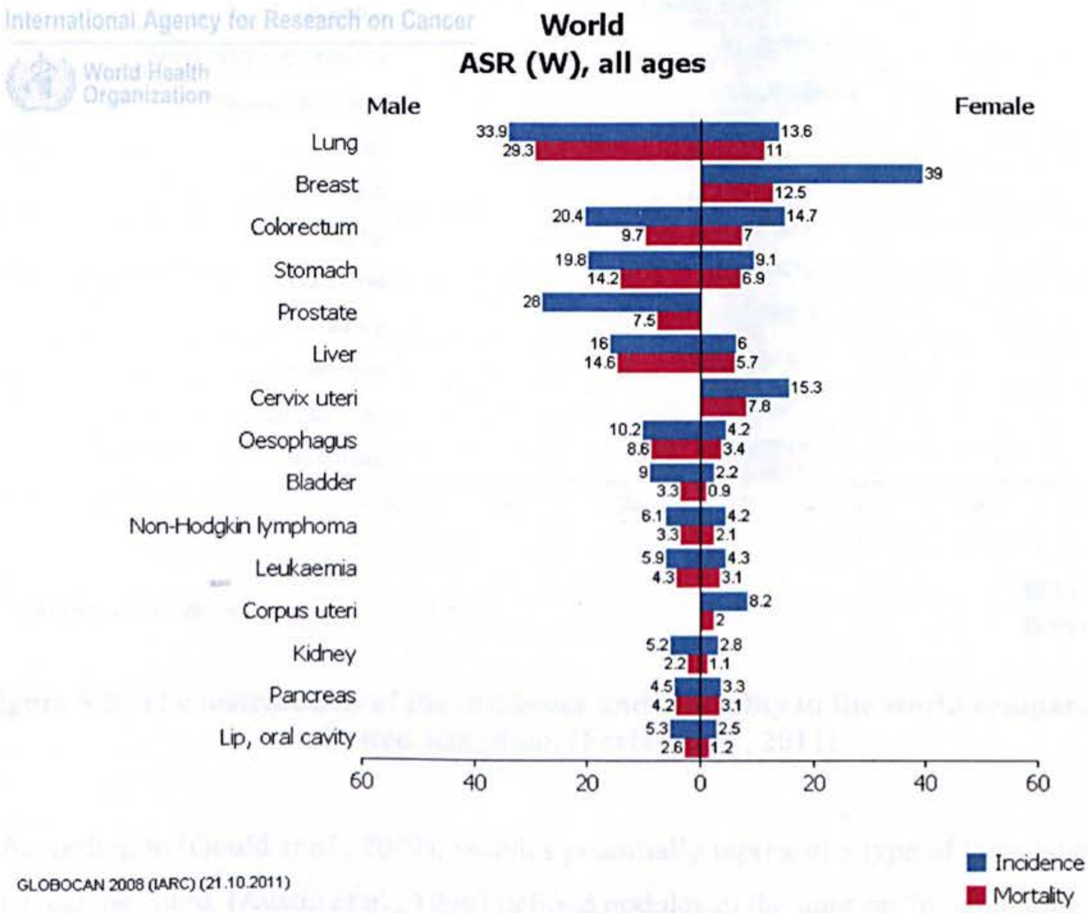


Figure 5.7: Worldwide age-standardized cancer rate for both male and female per 100,000, (Ferlay *et al.*, 2011)

Lung cancer is also the leading cause of cancer death in the United Kingdom (UK) in 2008 with the most mortality rate for both males and females. It is the most common diagnosed cancer among males and second highest among females (after breast cancer) (Ferlay *et al.*, 2011) as shown in Figure 5.8. Detection of the lung cancer in early stages increases the survival rate to 49% or more (Armato *et al.*, 2002), cited in (White *et al.*, 2007). About 38,000 people are diagnosed with lung cancer each year in the UK whilst 33,000 people died as a direct consequent. Smoking is the main cause of lung cancer, which leads to 90% of these deaths (Oversight, 2011).

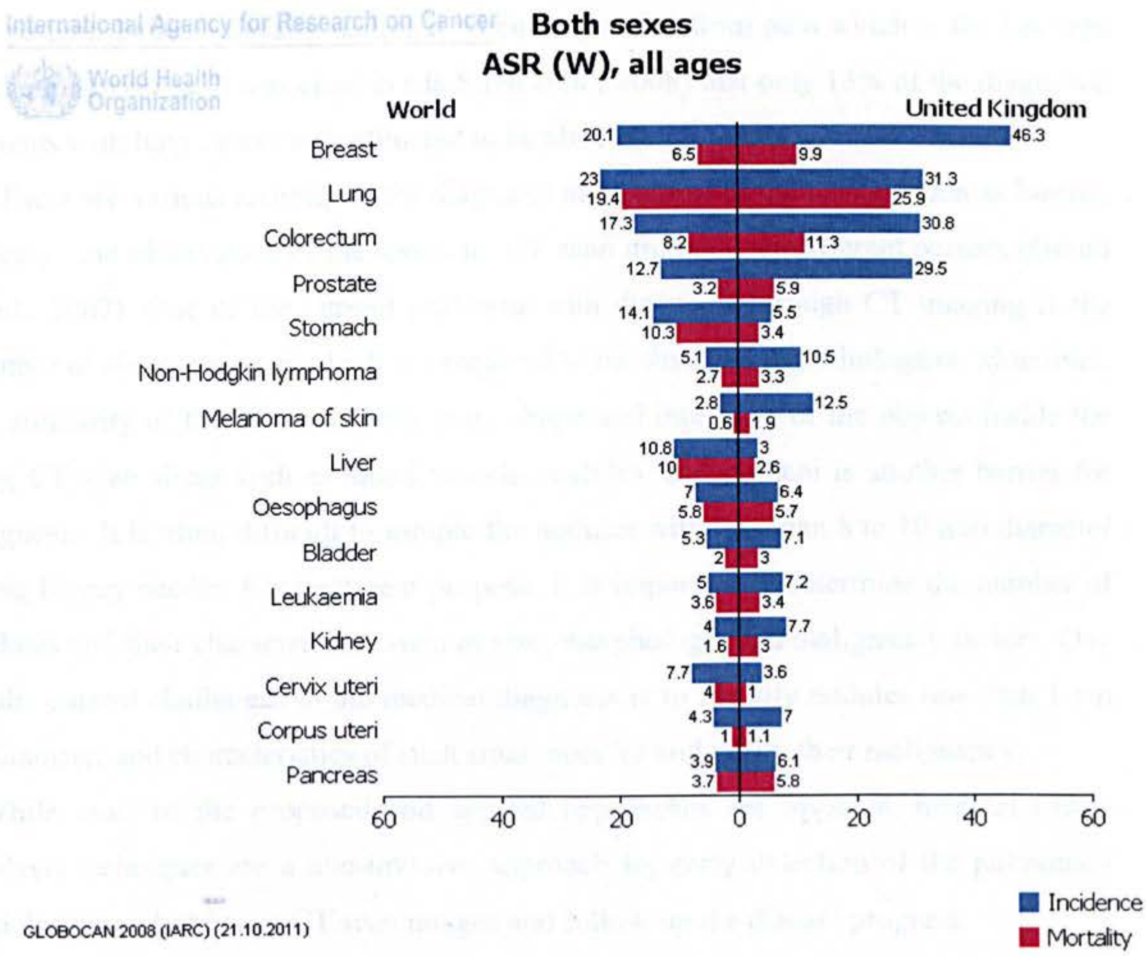


Figure 5.8: The distribution of the incidence and mortality in the world compare to the United Kingdom, (Ferlay *et al.*, 2011)

According to (Gould *et al.*, 2007), nodules potentially represent a type of lung cancer which can be cured. (Austin et al., 1996) defined nodules in the lung as: "in pathology, a nodule refers to a small approximately spherical circumscribed focuses of abnormal tissue"; in radiology, it is defined as spherical opacity, at least moderately well marginated with a maximum diameter not greater than 3 cm". Patients with lung cancer do not have any symptoms at the early stage; however all the non-calcified pulmonary nodules are considered potentially malignant and need monitoring until proven stable over a period of two years (MacMahon *et al.*, 2005). Nodules that are stable for a period of two years are characterized as benign nodules. CT Screening of the population with high risk factors of lung cancer (e.g., age, smoking history) is an approach for the early detection of the lung cancer. Although advances in the diagnosis and treatment can help to save more lives, most patients are diagnosed when the cancer is in advanced stages and difficult to treat. Diagnosis is usually performed after appearance of symptoms in the patient. Patients usually realize it when they suffer from pain which is the last sign of the lung cancer. It was cited in (da Silva *et al.*, 2008) that only 13% of the diagnosed patients with lung cancer are estimated to be alive after 5 years.

There are various techniques for diagnosis and control of the disease such as biopsy, surgery, and observation of the lesion in CT scan images over different periods (Gould *et al.*, 2007). One of the current problems with diagnosis through CT imaging is the number of slices per scan which are required to be observed by radiologists. Moreover, the similarity of the characteristics (e.g., shape and intensity) of the objects inside the lung CT scan slices such as blood vessels, nodules, and bronchi is another barrier for diagnosis. It is often difficult to sample the nodules with less than 8 to 10 mm diameter using biopsy needle. For treatment purpose, it is important to determine the number of nodules and their characteristics such as size, morphology, and malignancy factors. One of the current challenges in the medical diagnosis is to identify nodules less than 1 cm in diameter and characteristics of such small nodules and verify their malignancy.

While most of the proposed and applied approaches are invasive, medical image analysis techniques are a non-invasive approach for early detection of the pulmonary nodules through thoracic CT scan images and follow-up the disease progress.

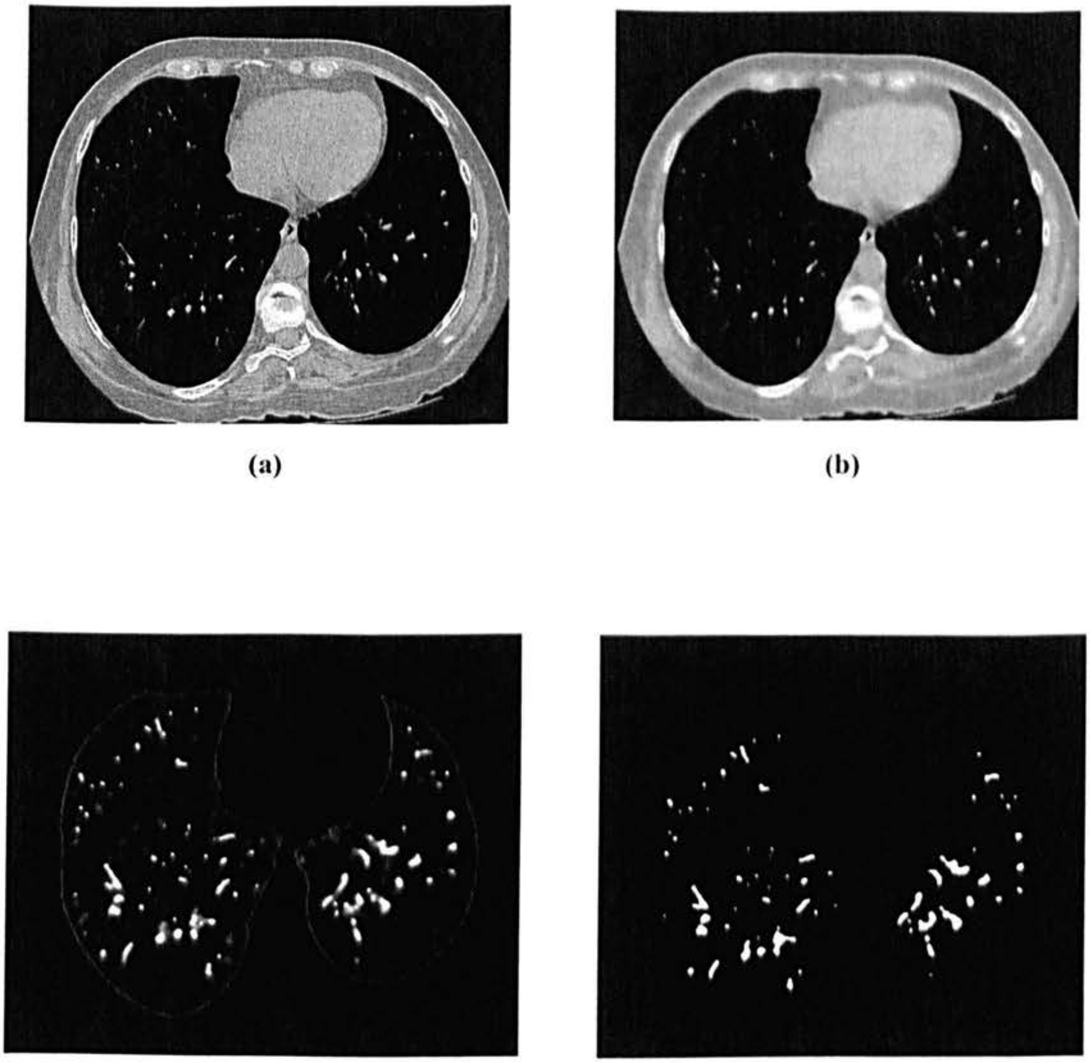
5.4.1 Overview of the lung CAD system

The aim of a lung CAD system is to automatically detect pulmonary nodules presented in computed tomography (CT) scans. Figure 5.9 depicts an example of nodule annotated in a CT scan image of a patient. Although, the development of the CAD system is not within the scope of this study, a summary of the lung CAD system is given in this section and the results of the IT2FLS classifier applied to the lung dataset are presented in the next section. A more detailed description of the lung CAD system can be found in (Dehmeshki *et al.*, 2007; Dehmeshki *et al.*, 2006).

Figure 5.9: Sample of a nodule presented in a CT image slice and manually annotated by a radiologist

The lung CAD system consists of several major components which are common to many CAD systems (Ye *et al.*, 2009; Keserci and Yoshida, 2002):

1. *Image pre-processing:* this component is used to minimize image noise and eliminate unwanted structures, see Figure 5.10 (b).
2. *Lung segmentation*: a segmentation algorithm is applied to isolate the lung region from the image data and to identify large structures, see Figure 5.10 (c).
3. *Object detection*: this stage detects all possible objects inside the lung; see Figure 5.10 (d). This step is expected to detected all potential nodules. However, it might detect many non-nodule objects.
4. *Feature extraction*: based on the radiologists' input, a set of features have been identified that differentiate nodules from other structures. These features are extracted for each of the potential nodules found in the object detection stage.
5. *Object classification*: classification of all objects detected in Step 3 into nodule or non-nodule classes, according to the extracted feature characteristics previously obtained in Step 4.



(a) (b)

(c) (d)

Figure 5.10: Major components of a lung CAD system: (a) Original image and (b) Pre-processing of the image to smooth the lung area; (c) Lung segmentation, applied to the smoothed image; (d) Object detection applied to the lung region

5.5 Experimental results and performance evaluation

To assess the strength of the proposed IT2FLS method, particularly in comparison to the T1FLS, both methods were integrated into the object classification component of a lung CAD system. The dataset of thoracic CT scans was provided by Lausanne Hospital, Switzerland. The dataset contains 81 nodules positively identified by the radiologists across the 40 patient scan datasets. Each scan consists of about 300 image slices in DICOM (digital imaging and communication in medicine) format with slice thicknesses ranging from 0.5 mm to 2 mm and X-ray tube currents ranging from 80 to 300 mA. The nodule diameter varies from 5mm to 20mm. To define the ground truth, the nodules were visually identified and manually annotated in the CT images (shown in Figure 5.9) by two radiologists and the results were considered as the gold standard. The radiologist's decision about each nodule candidate in the training dataset (A_{TS_i}) is a deterministic number which defines whether an object is a nodule (1) or a non-nodule (0). This result is compared to the IT2FLS output to estimate the accuracy of the classifier. An in-house lung CAD application was used to detect candidate objects and extract their characteristics (values of the features).

The IT2FLS was integrated into the lung CAD application for classification of candidate objects. In the lung nodule classification, a set of 14 rules were compiled based on the criteria adopted by the radiologists for differentiating between nodules or non-nodule samples. A total of 300 potential nodules were detected by the CAD system over the 40 scan dataset. The actual number of nodules in this datasets was 81. For the lung CAD, 7 features were extracted for each object. The features are described in Table 5.2. A ten-fold cross-validation technique was used for partitioning training and testing datasets. The nodule classification rules are listed in Table 5.3. The membership functions are considered as a Gaussian IT2FMF with fixed standard deviation and uncertain mean. The IT2FLS classification method was implemented in Matlab using the IT2FLS toolbox (Castro *et al.*, 2007; Castro, Castillo and Melin, 2007; Castro *et al.*, 2008). The details of the components of the software developed in this study for IT2FL membership function generation are presented in Appendix A.1. The membership function of the sphericity feature for low, middle and high type-2 fuzzy sets is shown in Figure 5.11(a). All the membership functions produced by the IT2FMF generation approach for seven features of the nodule candidates are presented in Figure 5.11(b). The interval type-2 fuzzy inference system is illustrated in Figure 5.11(c). The rules (14 rules) in the fuzzy logic are demonstrated in Figure 5.11 (d), where each row represents a classifer rule, each column is related to one of the seven type-2 fuzzy sets of features, and the last column is the output type-2 fuzzy set nodule possibility (NP). For a given sample X, the output is defined by calculating the type-2 implication membership function (explained in Chapter 4, Section 4.5) followed by applying centroid type-reduction and defuzzification, (see Chapter 4, Section 4.6).

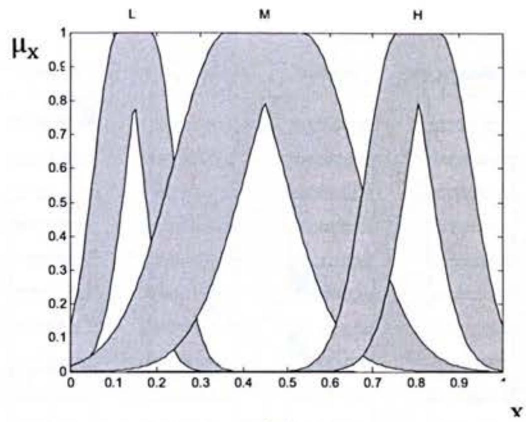
Table 5.2 Lung nodule candidate features

No	Feature	Definition
1	Volume	The size of the nodule
2	Diameter	The effective diameter of a nodule. It is defined as the diameter of a sphere with the same volume of the candidate object.
3	Sphericity	The degree of being spherical. It is is defined as the ratio between the object volume size and the minimum enclosing sphere volume size.
4	Mean HU- Spherical	The mean of the HU of the spherical part of the nodule
5	Elongation	The degree of elongation of a nodule
6	Mean HU-Elongated	The mean of the HU of the elongated part of a nodule
7	Distance	The distance of a nodule to the thoracic wall

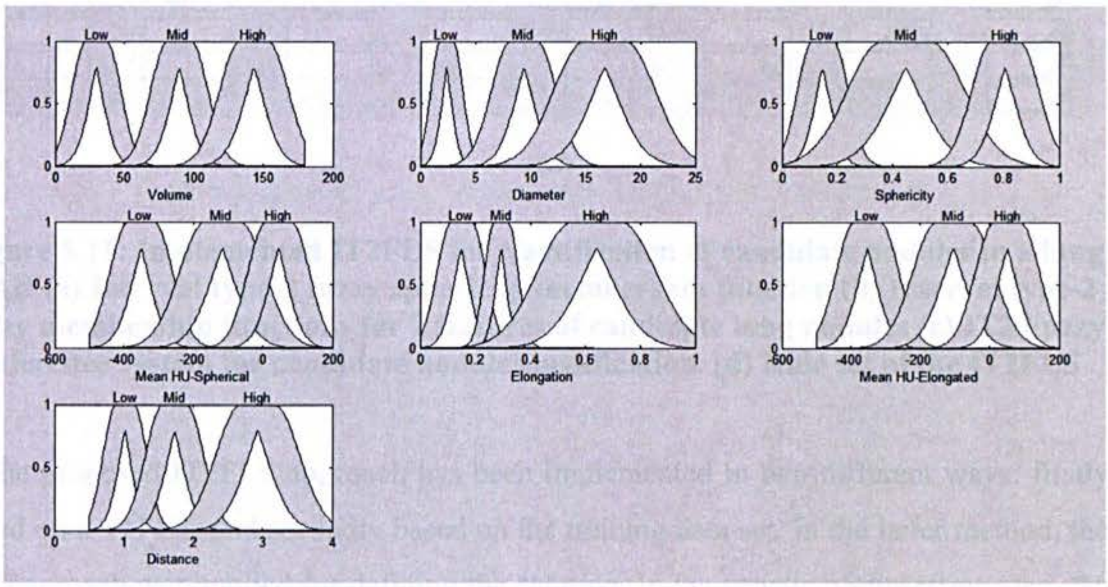
Table 5.3 List of nodule classification rules

1. If (Volume is High) and (Sphericity is High) then (NP* is High)
2. If (Volume is High) and (Distance is Mid) then (NP is High)
3. If (Volume is High) and (Diameter is High) then (NP is High)
4. If (Sphericity is High) and (Volume is Mid) then (NP is High)
5. If (Sphericity is High) and (Volume is High) and (Distance is High) then (NP is High)
6. If (Sphericity is High) and (Elongation is not Low) then (NP is High)
7. If Sphericity is High and MeanHU_Spherical is High then (NP is High)
8. If (Elongation is not Low) then (NP is High)
9. If (Sphericity is not Low) then (NP is High)
10. If (Distance is High) and (Diameter is High) then (NP is High)
11. If (Elongation is High) and (Sphericity is High) and (Diameter is Mid) then (NP is High)
12. If (Elongation is High) and (Diameter is not Mid) then (NP is High)
13. If (Distance is High) then (NP is High)
14. If (MeanHU_Elongated is not High) and (MeanHU_Spherical is not Low) then (NP is High)

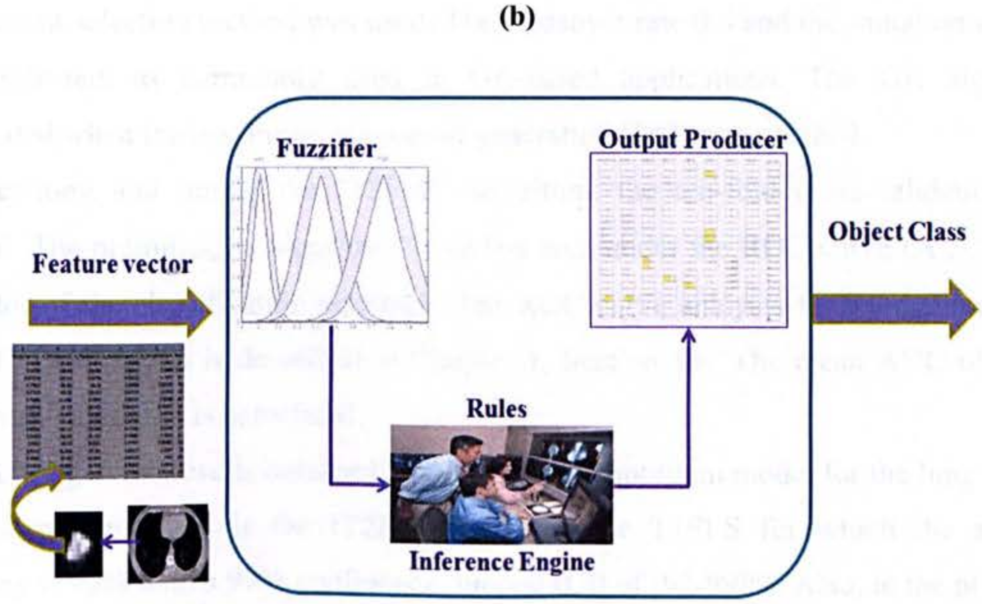
*: Nodule Possibility


L
M
H
μ_x
x

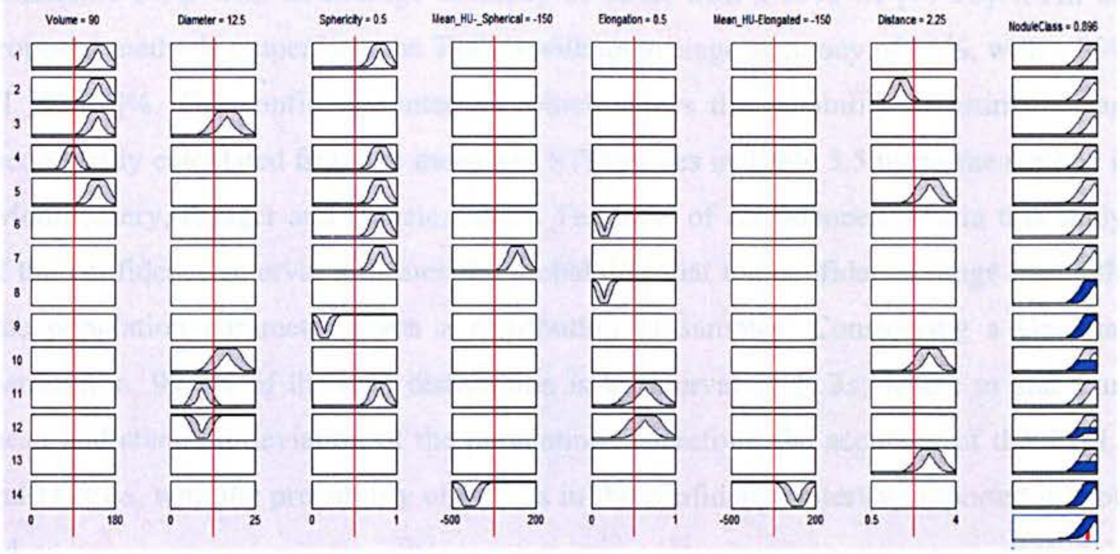
(a)


Low
Mid
High
Volume
Diameter
Sphericity
Mean HU-Spherical
Elongation
Mean HU-Elongated
Distance

(b)


Fuzzifier
Output Producer
Object Class
Feature vector
Rules
Inference Engine

(c)



(d)

Figure 5.11: Implemented IT2FLS for classification of candidate nodules in a lung CAD (a) Interval type-2 fuzzy sphericity membership function (b) Interval type-2 fuzzy membership functions for 7 features of candidate lung nodules (c) IT2 Fuzzy inference system for candidate nodule classification (d) Rule set of the IT2FLS

The proposed IT2FLS approach has been implemented in two different ways: firstly based on a T1FLS, and secondly based on the training data set. In the latter method, the GA approach was applied for defining the solution. In the genetic optimization step, the tournament selection method was used. The crossover rate 0.5 and the mutation rate 0.1 were selected, as commonly used in GA-based applications. The GA algorithm terminated when the maximum number of generation (100) was reached.

In learning and tuning the IT2FMF algorithm, the ten-fold cross-validation was applied. The optimization algorithm considers area under the ROC curve (AUC) as an indicator of the classification accuracy. The ROC curve analysis for estimation of the cost of classification is described in Chapter 1, Section 1.6. The mean AUC of a ten-fold cross-validation is calculated.

According to the results obtained in Table 5.4, the optimum model for the lung nodule classification problem is the IT2FLS based on the T1FLS for which the average accuracy is 95% with a 99% confidence interval (CI) of [92-99]%. Also, in the proposed method based on the training dataset using three linguistic terms (Low, Middle, High), when a GA tuning approach for non-monotonic FOU parameters was applied, the average accuracy is 83%, with 99% CI [71-96]%. This outperforms the method using a monotonic FOU with an average accuracy of 82%, with a 99% CI [71-96]%. All the proposed methods outperform the T1FLS with an average accuracy of 65%, with a 99% CI [52-77]%. The confidence intervals which shows the reliability of estimation are theoretically calculated from the mean and STD values in Table 5.5 using the method in (Montgomery, Runger and Hubele, 2009). The level of confidence (99% in this study) of the confidence interval indicates the probability that the confidence range meets the true population parameter given a distribution of samples. Considering a Gaussian distribution, 99.7% of the data distribution is in interval $m \pm 3s$, where m and s are mean and standard deviation of the population. Therefore, the accuracy of the T1FLS and IT2FLs, with the probability of 99% is in the confidence intervals reported in table 5.4.

Table 5.4 Analysis of the results of the T1FLS and the IT2FLSs

	Accuracy	
	Mean	CI*
T1FLS	65%	[52%-77%]
IT2FLS (Based T1FLS)	95%	[92%-99%]
IT2FLS (Monotonic FOU)	82%	[66%-96%]
GA IT2FLS (non-Monotonic)	83%	[71%-96%]

*CI (Confidence Interval) is the 99% confidence intervals

In order to compare the efficiency of the proposed method, a two-sample left tailed t-test has been applied. By comparing the t-test results, we are able to determine if the proposed IT2FLS classifiers are statistically significantly different from the T1FLS classifier. We performed the t-test with the null hypothesis that the mean of the ith method is greater than the mean of the jth method, against the alternative hypothesis that the mean of the ith method is less than mean of the jth method. The null hypothesis has been defined as,

$$H_0: \mu_i > \mu_j$$

$$H_1: \mu_i < \mu_j$$

where μ_i and μ_j are the means of the accuracy of the ten-fold cross-validation as

$$\mu_i = \frac{1}{10}\sum_{k=1}^{10} AUC_i \tag{5.3}$$

Table 5.5 Two sample left-tailed t-test analysis of the results of the T1FLS and proposed IT2FLSs for lung nodule dataset

Fold #	T1FLS	IT2FLS based T1	IT2FLS Monotonic FOU	GA IT2FLS
1	0.5	0.92	0.75	0.752
2	0.67	0.89	0.62	0.66
3	0.68	0.98	1	1
4	0.69	0.91	0.63	0.64
5	0.69	1	0.62	0.64
6	0.58	0.96	1	1
7	0.68	0.97	0.81	0.8
8	0.46	0.94	0.85	0.86
9	1	1	1	1
10	0.52	0.96	0.96	0.96
Mean	**0.65**	**0.95**	**0.82**	**0.83**
STD	0.15	0.04	0.16	0.15
Decision/p-value (T1FLS)		R 4.8574e-005	R 0.0241	R 0.0071
Decision/p-value (GA IT2FLS)	FR 0.9929	R 0.0166	FR 0.6081	
Decision/p-value (IT2FLS based T1FLS)	FR 1		FR 0.9796	FR 0.9834

FR: Fail to Reject null hypothesis, R: rejection of null hypothesis at α=0.05 significance level. The number below each decision (FR/R) shows *p*-value which is the probability, under the null hypothesis, of observing a value as extreme or more extreme of the test statistic.

The two sample left-tailed t-test results in Table 5.5 (the last three rows) reveals the superiority of the proposed learning and tuning IT2FMF in comparison to the T1FLS for nodule classification in the lung CAD application. The test failed to reject (FR) the null hypothesis, H_0 (the mean of the IT2FLS is greater than the mean of the T1FLS), with a high probability. Furthermore, we have applied the t-test for the following null hypothesis:

$$H_0: \mu_{IT2FLS} - \mu_{T1FLS} > 30\%$$

where μ_{T1FLS} and μ_{IT2FLS} are the means of the accuracy of ten-fold cross-validation for the T1FLS and the IT2FLS based T1FLS, calculated using equation (5.3). The t-test result failed to reject the null hypothesis with 95% confidence interval (at p-value=0.86) which indicates that the IT2FLS performs 30% better than the T1FLS.

The graph in Figure 5.12 compares the overall ROC curves of the T1FLS classifier and the proposed IT2FLS methods. The area under the ROC curve is considered for the evaluation of the cost of classification. The ROC curve of each classifier is plotted by setting a threshold value ranging from a minimum to maximum degree of membership of the classified objects. The fuzzy logic classifier provides a membership degree to which an object belongs to a class. Using different degrees of membership of the object to a class fuzzy set as a threshold provides different points with different true positive (*tp*) and false positive (*fp*) rates on the ROC curve. The IT2FLS based T1FLS curve is superior in general. Although the two T2FLS designed based on the training datasets begin to rise slowly, they have a similar trend throughout the maximum sensitivity. Compared with the T1FLS curve, the IT2FLS methods have better ROC curves with a larger area under the curve (the larger the area under the curve the better the classification results). The nodule classifica\tion results of the T1FLS and the IT2FLS are illustrated in the example thoracic CT scan images in Figure 5.13.

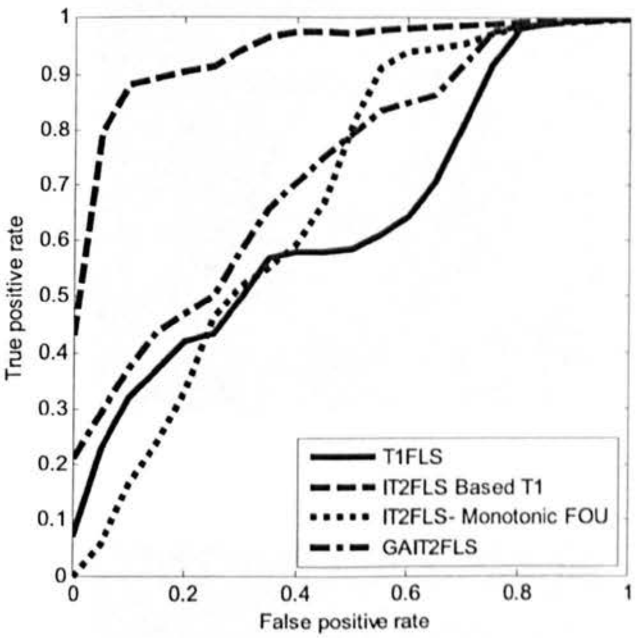
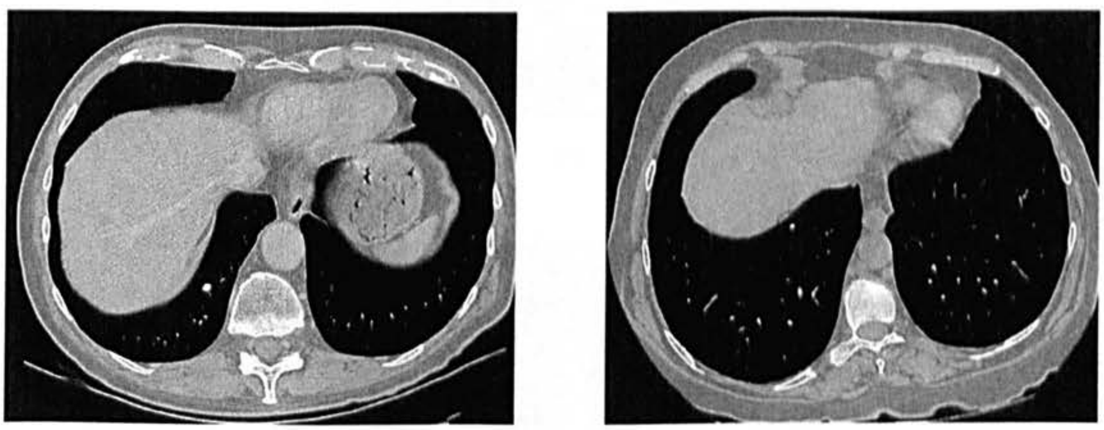
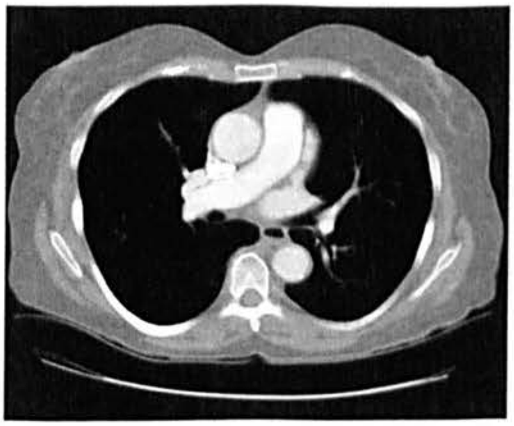


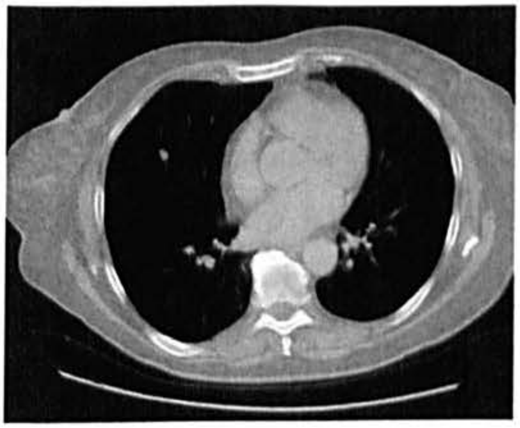
Figure 5.12: The ROC curve of the T1FLS compared to the proposed IT2FLSs



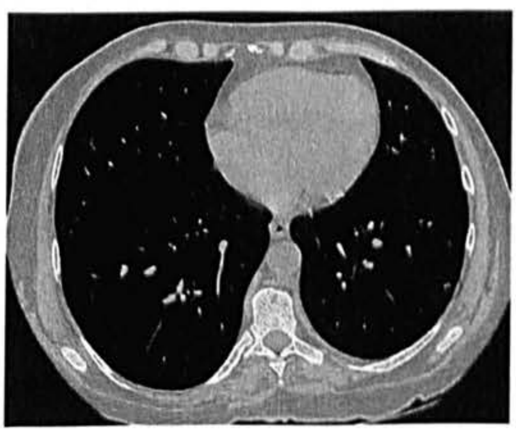
(a)



(b)



(c)



(d)

Figure 5.13: Examples of the T1FLS and the IT2FLS nodule classification: detected nodules are circled in red (a) True Positive (TP) detected by the T1FLS and the IT2FLS (nodules), (b) False Positive (FP) detected by the T1FLS and the IT2FLS, (c) FP detected by the T1FLS but not the IT2FLS (d) TP detected by the IT2FLS but not by the T1FLS

5.6 Summary

This chapter presented an approach for learning and tuning a Gaussian interval type-2 membership function to manage the problem of uncertainty in the input of a classifier. This method generates the FOU of a Gaussian IT2FMF based either on an uncertain T1FLS or an imperfect training dataset. The GA algorithm searches to find an optimum set of parameters of the IT2 membership functions. The FOU of the IT2FMF was learned using the genetic algorithm, the new chromosome structure and initialization technique. The FOU in an IT2FLS is modelled using the k_m and k_v parameters. In the GA-based approach, the structure of the chromosome consists of FOU parameters for each feature linguistic terms and has fewer genes (about half of the other GA based methods) than other GA based methods. In the GA IT2FMF generation approach for estimating membership function parameters, chromosome structure includes the FOU parameters (k_ms and $k_v s$). This method initializes chromosome more precisely. This property is significant for medical classification problems with large number of inputs where it is difficult or expensive to initialize membership function parameters using knowledge of experts.

This method was applied to classify candidate nodules detected by a lung CAD application. The improvement achieved in nodule classification application is considerably significant whilst lung cancer is the leading cause of cancer death with the most mortality rate for both males and females. Integrating interval type-2 fuzzy logic classifier into the lung CAD application enables us to model the uncertainties in input training sets such as inter- and intra uncertainties, word perception, and numerical measurements. These uncertainty sources are modelled using the FOU in an IT2FMF. In order to have an unbiased view of the classifier performance with the minimum dependency to the dataset, the training and testing datasets were randomly selected using the ten-fold cross-validation technique. Then, the average performance of ten runs was used to represent the final accuracy of the method.

The accuracy of each IT2FLS classifier is measured using the area under ROC curve. The maximum accuracy is achieved for the IT2FLS based on the uncertain T1FLS defined by experts, with an average accuracy result of 95%. The three IT2FLS methods described in this chapter all outperform the T1FLS counterpart (with an average accuracy of 65%). Analysis of the results reveals that the IT2FLS is more capable of capturing the uncertainty sources in the model and achieving a better performance results. For the lung nodule classification, the IT2FLS performance is 30% better than the T1FLS. In addition, in order to compare the efficiency of the proposed method compared to each other, a two-sample left tailed t-test was conducted. According to the t-test results, the proposed IT2FLS classifiers are statistically significantly different from the T1FLS classifier. The t-test result failed to reject the null hypothesis with 95% confidence interval (at p-value=0.86) which indicates that the IT2FLS performs 30% better than the T1FLS.

The result of the proposed methods for automatic learning and tuning of the IT2FMFs based on the training dataset is more promising for multidimensional classification problems that lack expert knowledge. Furthermore, the result of the proposed methods for automatic learning and tuning of the IT2FMFs based on the training dataset with an average accuracy 82% is more promising for multidimensional problems that lack expert knowledge. These findings suggest that, IT2FLS is capable to provide more efficient performance results compared to T1FLS for classification of candidate nodules in a lung CAD application.

Chapter 6: A Genetic Type-2 Fuzzy Approach for Rule Extraction for Uncertain Pattern Classification Problems

Fuzzy rule-based classification provides a rule set with more interpretability than the other proposed methods. While most fuzzy classifiers need expert knowledge for initializing the fuzzy set's membership functions, and the initial rule set, extracting this knowledge from medical experts is a difficult, time consuming and expensive process. Furthermore, a FLS for classification in medical applications is expected to model uncertainty sources in the input training datasets as well as system rules. While, several methods have been proposed for modelling uncertainty in the input of a classifier using type-2 fuzzy membership functions, no attempt has been reported for extracting rules for a type-2 fuzzy logic using an uncertain training dataset. In chapter 5, an approach for modelling uncertainty in the input of the classifier using IT2FMFs was presented where rules are available from experts. This chapter extends this work (Hosseini et al., 2012) for circumstances where the knowledge about the rules is not available. For this purpose, the idea of uncertain rule-based fuzzy pattern classification is introduced. An automatic approach is presented which extracts rules from imprecise and imperfect training dataset and learns them for future observations. This model takes advantage of type-2 fuzzy sets for tackling the problem of uncertainty in classification. A genetic algorithm with a punishment-reward scheme is applied for learning rules. For evaluation purposes, the approach has been applied to the anonymous Wisconsin breast cancer diagnostic (WBCD) dataset and the results are provided. The aim is to show that modelling uncertainty in the rule set provides a classifier for the WBCD dataset which competes with the best results of other classification methods for this dataset.

6.1 Introduction

Type-2 fuzzy sets have the potential for managing uncertainties in linguistic terms in vague environments (i.e., with subjective or imprecise and imperfect knowledge). A type-2 fuzzy logic system (T2FLS) utilizes a three-dimensional type-2 fuzzy membership function (T2FMF) and the footprint of uncertainty (FOU) for managing the problem of uncertainty in classification (see Section 4.3 Chapter 4). Uncertainty sources in a linguistic fuzzy rule-based classifier can be modelled through two important components: the rule set and the database. The rule set contains a collection of linguistic rules (extracted from experts or defined by a learning algorithm) whilst the database includes linguistic term sets in the rules and their associated MFs. The accuracy of a fuzzy rule-based classifier is affected by these two major components.

Despite the fact that there are several methods for learning and tuning membership functions in a T2FLS rule-based classifier, to my present knowledge, no effort has been made for automatically extracting rules for a medical imaging classification applications taking account of uncertainty sources in noisy images and imperfect datasets. However, imperfect and imprecise datasets (including features characterizing object of interest) influence the rule extraction process as well as membership function generation. Uncertainty issues are major hidden barriers to better performance of the pattern classification. Moreover, most learning and tuning MF approaches for an IT2FLS are applied to rules with fuzzy sets (Mamdani rule), a linear function or a real number (TKS rule types) in the consequent part of the rule (different type of rules are explained in Chapter 3 Section 3.3). Whilst the rule type with a class label and a degree of certainty of the classifier decision (Ishibuchi and Nojima, 2008; Ishibuchi, Nakashima and Murata, 1996) is more applicable for complex pattern classification problems and has higher interpretability than the other types of the rules and is the focus of this chapter.

The idea of uncertain rule-based fuzzy pattern classification and the related mathematical model is introduced in this chapter. In such a system, antecedent of the rules take advantage of type-2 fuzzy sets for tackling the problem of uncertainty and consequent of the rules defines the class label with a degree of certainty. A Genetic algorithm (GA) in combination with a punishment-reward scheme is applied for learning rules and the degree of certainty. The proposed approach improves the Ishibuchi fuzzy rule-based pattern classification model (Ishibuchi and Nojima, 2008) for extracting rules from imprecise and imperfect datasets. The proposed model presents a new rule structure for uncertain rule-based pattern classification using type-2 fuzzy set theory and provides its mathematical framework. This type of rule has the capacity to capture the uncertainties in the training dataset such as uncertainties in the mathematical models applied for measuring the features (different methods with different error rates may lead to different results) in the antecedent part of the rule. The main application of the proposed rule learning algorithm is the capacity of automatically extracting rules for multi-dimensional pattern classification problems with high degree of uncertainty in numerical measurements and linguistic terms, such as medical diagnosis applications.

The rest of this chapter is organized as follows: The next three sections (6.2 to 6.4) describe the proposed approach for automatically constructing two major components of an IT2FLS while the aim is to model uncertainties: (1) interval type-2 fuzzy membership function generation, and (2) interval type-2 fuzzy GA (IT2F GA) rule learning approach for pattern classification applications. For evaluation purpose, the proposed approach has been applied to the popular WBCD classification problem. The breast cancer diagnosis problem using a mammography CAD application is explained in Section 6.5. Then, an overview of current rule extraction methods proposed for the WBCD classification problem is presented in Section 6.6. The result of the IT2F GA rule learning method applied to the WBCD is demonstrated in Section 6.7 and this chapter is concluded in Section 6.8.

6.2 Interval type-2 fuzzy membership function generation

Uncertainty sources in the input of a fuzzy classification can be managed through type-2 fuzzy set membership functions and their FOU. The membership function of a type-2 fuzzy set can be defined using knowledge of experts or through an expert system. Several approaches have been presented for automatic learning and tuning of interval type-2 fuzzy membership functions in (Hosseini et al., 2010a; Mitchell, 2005; Hidalgo, Castillo and Melin, 2008; Castro et al., 2007; Hosseini et al., 2012). In order to design type-2 fuzzy sets, expert knowledge can be used for defining the number of linguistic terms and parameters of the membership function for each input feature in the pattern space. However, in circumstances where we cope with a lack of expert knowledge or expensive knowledge extraction processes, such as for most pattern classification for medical diagnosis applications, automatic learning and tuning methods are vital for defining fuzzy membership function for linguistic terms using on a training dataset. For illustration of the problem, assume a 2-dimensional pattern classification system, $X = (x_1, x_2)$, with three fuzzy sets associated to three linguistic terms (low, middle and high), the total number of possible combinations of fuzzy sets in the antecedent of the rules is $(3)^2 = 9$ (see Figure 6.1), and if considering "don't care" as a fuzzy set, it is $(3+1)^2 = 16$.

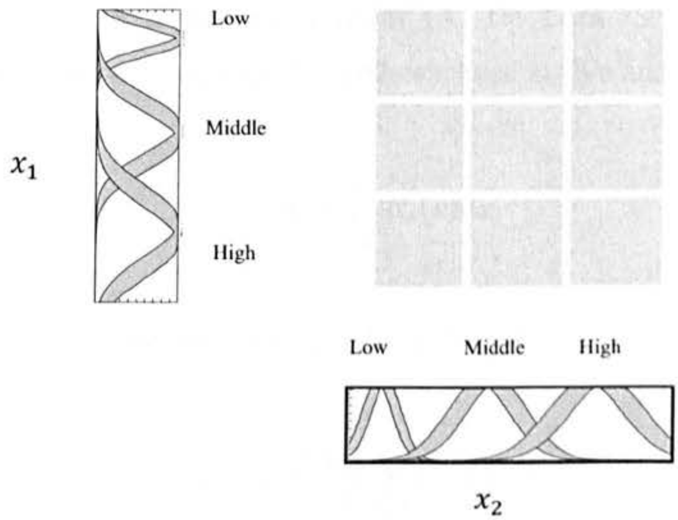


Figure 6.1: Type-2 fuzzy partitioning of the pattern space

In this study an interval type-2 fuzzy set with a primary Gaussian membership function (called an interval type-2 Gaussian membership function) is imposed. For each feature, the corresponding values in the training dataset is clustered using standard c-mean clustering method (Chi, Yan and Phạm, 1996) into three Gaussian functions equivalent to the low, middle, and high linguistic terms. For modelling the footprint of uncertainty, genetic algorithms have been frequently used to tune the lower and upper bound parameters of a type-2 MF (Cazarez-Castro, Aguilar and Castillo, 2010; Wu and Wan Tan, 2006; Martinez, Castillo and Garcia, 2008). However, the proposed GA learning and tuning type-2 membership function approaches are applicable when a fuzzy logic with initial values of the membership function parameters and a set of rules exists. These methods are not applicable for circumstances with no prior expert knowledge or when the initial fuzzy membership functions and rule set are not available. The classification problem for the WBCD breast cancer diagnosis is an instance of such a problem with only a training dataset available.

In order to generate the interval type-2 fuzzy MFs for the breast cancer features using the model described in Chapter 5, Section 5.2.2, first it is needed to estimate the FOU parameters, k_m and k_v. For this, an uncertainty measure is calculated for each possible pair of (k_m, k_v) and interval type-2 fuzzy sets related to a linguistic term of an input feature. Wu and Mendel suggested cardinality of an IT2FS as a measure of intra-uncertainty between all embedded type-1 fuzzy sets (Wu and Mendel, 2007; Wu and Mendel, 2009). De Luca and Termini (A. De Luca, S. Termini, 1972) defined cardinality of a type-1 fuzzy set A as follows, cited in (Wu and Mendel, 2009):

$$P_k(A) = \int_x \mu_A(x)dx \tag{6.1}$$

The normalized cardinality of a type-1 fuzzy set A is defined as follows:

$$p(A) = \frac{|X|}{N}\sum_{i=1}^{N} \mu_A(x_i) \tag{6.2}$$

where $|X| = |x_N - x_1|$ is the length of universe of discourse where x_i $(i = 1 \dots N)$ are chosen with equal space in the domain of x. Cardinality of interval type-1 fuzzy set is used to calculate cardinality of an interval type-2 fuzzy set (Szmidt and Kacprzyk, 2001). The cardinality of an IT2FS $\tilde{A}$ is the union of cardinality of all its embedded type-1 fuzzy sets, A_e, is defined as follows in (Wu and Mendel, 2009):

$$P(\tilde{A}) = \bigcup_{\forall A_e \in FOU(\tilde{A})} p(A_e) = [\min_{\forall A_e} p(A_e), \max_{\forall A_e} p(A_e)] \tag{6.3}$$

$$P(\tilde{A}) = [p(\underline{\mu}_{\tilde{A}}(x)),\ p(\overline{\mu}_{\tilde{A}}(x))] \tag{6.4}$$

The length of interval indicates uncertainty in an IT2FS while a larger interval represents more uncertainty. The average cardinality of $\tilde{A}$ is defined as follows:

$$P_{avg}(\tilde{A}) = \frac{p(\underline{\mu}_{\tilde{A}}(x)) + p(\overline{\mu}_{\tilde{A}}(x))}{2} \tag{6.5}$$

The average cardinality, $P_{avg}(\tilde{A})$, is calculated for each possible pair of k_m and k_v and corresponding interval type-2 fuzzy sets of the input. The IT2FL with minimum average cardinality of its MFs shows the minimum uncertainty in the system. For all possible pairs of k_m and k_v, a ten-fold cross-validation is applied which considers one fold for testing and the rest for training and repeats this process until all folds are selected once for validation. The average results of a ten-fold cross-validation run of the GA learning rule method represents performance of each k_m and k_v pairs in terms of average cardinality and average error rate on training and testing datasets. The k_m and k_v with minimum overall classification error rate (on training and testing datasets) and minimum average cardinality is selected for generating membership functions in the GA rule learning algorithm.

6.3 Rule extraction for interval type-2 fuzzy pattern classification

In this section, an automatic rule learning approach for a type-2 fuzzy pattern classification is presented. This approach extracts rules from an uncertain and imperfect training dataset. The approach proposed for learning linguistic rules from training data extends the method presented in (Ishibuchi and Nojima, 2008; Ishibuchi, Nakashima and Murata, 1996) and improves it for uncertain rule-based pattern classification. The Ishibuchi et. al. method is proposed for extracting rules from a training dataset using type-1 fuzzy logic and the genetic algorithm. However, this study extends this method and its mathematical model for uncertain pattern classification using type-2 fuzzy logic. The footprint of uncertainty in type-2 fuzzy sets allows us to capture more uncertainties in membership functions for classification of the patterns compared to the type-1 fuzzy equivalent. This is the case for most classification problems in medical applications such as CAD applications. Moreover, in their method the fuzzy sets in the antecedent of the rule is manually defined by a human expert while the proposed approach in this study automatically generates fuzzy sets using the model explained for partitioning the pattern space in Section 6.2 and without using the knowledge of experts.

This study is the first attempt toward automatically generating type-2 fuzzy logic including membership functions and rule set from imprecise and imperfect training dataset. However, some attempts have been reported in recent years for linguistic summarization of the database for natural language representation of the objects using type-2 fuzzy sets but these methods need user intraction for defining the type-1 fuzzy membership function and the FOU parameter estimation (Niewiadomski, 2008; Niewiadomski, 2007; Niewiadomski and Bartyzel, 2006; Niewiadomski and Szczepaniak, 2006; Niewiadomski, 2005). Recently a linguistic summarization method using interval type-2 fuzzy sets and If-then rules has been proposed for weighted rule extraction from a causal database (Wu, Mendel and Joo, 2010) using quality measures. This method also needs expert knowledge for initializing fuzzy membership function parameters. Furthermore, it requires knowledge of a group of experts to define the interval of the FOU.

6.3.1 Rule generation for IT2FL

A rule generation procedure generates linguistic classification rules for the uncertain rule-based fuzzy classification. The uncertainties in the rules are managed through type-2 fuzzy sets in the antecedent part of the rules. The structure of a rule in a type-2 fuzzy rule-based pattern classification for a given pattern $X_p = (x_{p1}, \dots, x_{pn})$ is as follows:

$$Rule\ R_i\text{: } If\ x_{p1}\ is\ \tilde{A}_{i1}\ and\ \dots and\ x_{pn}\ is\ \tilde{A}_{in}\ then\ Class\ C_i\ with\ GC_i$$

where $\tilde{A}_{i1}, \dots, \tilde{A}_{in}$ are interval type-2 fuzzy sets, $i=1..M$ is the number of rules, C_i is the class label and the GC_i is the grade of certainty of the rule. An example of such a rule is *"If Clump Thickness (CT) is Low and Marginal Adhesion (MA) is Low and* Bare Nuclei (BN) *is Low and Normal Nucleoli (NN) is Low then object is benign"*. In order to improve interpretability of an IT2FL, a maximum of three linguistic terms is considered during the rule generation method. The steps of the proposed method for the rule selection from an imperfect training dataset are explained as follows:

Step 1: Calculate the grade of compatibility of pattern X_p to the jth linguistic classification rule R_j as (Mendel, 2001),

$$\tilde{A}_j(X_p) = \tilde{A}_{j1}(x_{p1}) \cap \tilde{A}_{j_2}(x_{p2}) \cap \dots \cap \tilde{A}_{j_n}(x_{pn}) \tag{6.6}$$

where $\cap$ is the type-2 fuzzy intersection (see Section 4.5 in Chapter 4) which can be implemented by a meet operation (N. Karnik and M. Mendel, 2001).

The total grade of compatibility of the given patterns in training dataset in $Class\ k$ to the jth classification rule R_j, $(\tilde{\beta}_{Class\ k})$, is an interval type-2 fuzzy set $\tilde{\beta}_{Class\ k} = [\underline{\beta}_{Class\ k}\ \bar{\beta}_{Class\ k}]$ which is calculated as follows:

$$\tilde{\beta}_{Class\ k} = [\underline{\beta}_{Class\ k}\ \bar{\beta}_{Class\ k}] = \sum_{\forall X_p \in Class\ k} \mu_j(X_p) \tag{6.7}$$

where μ_j is the firing strength of the given pattern X_p for rule R_j as,

$$\mu_j(X_p) = \mu_{\tilde{A}_{j_1}}(x_{p1}) \cdot \mu_{\tilde{A}_{j_2}}(x_{p2}) \cdot \dots \cdot \mu_{\tilde{A}_{j_n}}(x_{pn}) \tag{6.8}$$

In the approach proposed in this chapter, $\mu_{\tilde{A}_{j_1}}(x_{p1})$ is the MF of the input x_{p1} to the interval type-2 fuzzy set $\tilde{A}_{j1}$ and $(\cdot)$ is a product t-norm (N. Karnik and M. Mendel, 2001).

$$\begin{aligned} \underline{\beta}_{Class\ k} &= \sum_{\forall X_p \in Class\ K} \underline{\mu}_j(X_p) \\ &= \sum_{\forall X_p \in Class\ k} \underline{\mu}_{\tilde{A}_{j_1}}(x_{p1}) \cdot \underline{\mu}_{\tilde{A}_2}(x_{p2}) \cdot \dots \cdot \underline{\mu}_{\tilde{A}_{jn}}(x_{pn}) \end{aligned} \tag{6.9}$$

$$\begin{aligned} \bar{\beta}_{Class\ k} &= \sum_{\forall X_p \in Class\ k} \bar{\mu}_j(X_p) \\ &= \sum_{\forall X_p \in Class\ k} \bar{\mu}_{\tilde{A}_{j_1}}(x_{p1}) \cdot \bar{\mu}_{\tilde{A}_{j2}}(x_{p2}) \cdot \dots \cdot \bar{\mu}_{\tilde{A}_{j_n}}(x_{pn}) \end{aligned} \tag{6.10}$$

where $\tilde{\beta}_{Class\ K}$ is calculated for each class (1...c) in the pattern space.

The centroid type reduction followed by defuzzification in (Karnik, Mendel and Liang, 1999) for an interval type-2 fuzzy set $\tilde{\beta}$ is used to calculate $\beta_{Class\ k}$ as,

$$\beta_{Class\ k} = \frac{\underline{\beta}_{Class\ k} + \bar{\beta}_{Class\ k}}{2} \tag{6.11}$$

For illustration, consider a given sample with input vector:

$$X_p = (x_{p1}, x_{p2}, \dots, x_{pN}) = (5, 1, 1, 1, 2, 1, 3, 2, 1)$$

The membership function values of the input X_p to the linguistic type-2 fuzzy sets in the antecedent part of the rule:

R_j: *"If x_{p1} is Mid and x_{p2} is Low and x_{p3} is Low and x_{p4} is Low and x_{p5} is Low and x_{p6} is Low and x_{p7} is Low and x_{p8} is Low and x_{p9} is Low "*, are computed using the model presented in Chapter 5 Section 5.2.2 and the method explained in Section 6.2 for the FOU parameter $(k_m, k_v) = (0.2, 0.7)$, as follows:

$\mu_{\tilde{A}_{j_1}}(x_{p1}) = [\underline{\mu}_{\tilde{A}_{j_1}}(x_{p1})\ \bar{\mu}_{\tilde{A}_{j_1}}(x_{p1})] = [0.6481\ 0.9846]$,

$\mu_{\tilde{A}_{j_2}}(x_{p2})$ =[0.9775 1], $\mu_{\tilde{A}_3}(x_{p3})$ =[0.9698 1], $\mu_{\tilde{A}_{j_4}}(x_{p4})$ = [0.9297 0.9787],

$\mu_{\tilde{A}_5}(x_{p5})$ =[0.9706 1], $\mu_{\tilde{A}_{j_6}}(x_{p6})$ =[0.9801 1], $\mu_{\tilde{A}_{j_7}}(x_{p7})$ = [0.0987 0.7409],

$\mu_{\tilde{A}_{j_8}}(x_{p8})$ = [0.5146 0.9603], $\mu_{\tilde{A}_{j_9}}(x_{p9})$ = [0.9666 1]

The grade of compatibility of X_p to the rule R_j using product t-norm is computed as follows:

$$\begin{aligned}\mu_{\tilde{A}_j}(X_p) &= [\underline{\mu}_{\tilde{A}_j}(X_p)\ \bar{\mu}_{\tilde{A}_j}(X_p)] \\ &= [\underline{\mu}_{\tilde{A}_{j_1}}(x_{p1}) \cdot \underline{\mu}_{\tilde{A}_2}(x_{p2}) \cdot \dots \cdot \underline{\mu}_{\tilde{A}_{jn}}(x_{pn}) \\ &\qquad \bar{\mu}_{\tilde{A}_{j_1}}(x_{p1}) \cdot \bar{\mu}_{\tilde{A}_{j_2}}(x_{p2}) \cdot \dots \cdot \bar{\mu}_{\tilde{A}_{jn}}(x_{pn})]\end{aligned}$$

where,

$$\underline{\mu}_{\tilde{A}_j}(X_p) = \underline{\mu}_{\tilde{A}_{j_1}}(x_{p1}) \cdot \underline{\mu}_{\tilde{A}_2}(x_{p2}) \cdot \dots \cdot \underline{\mu}_{\tilde{A}_{jn}}(x_{pn})$$

$$= prod(0.6481\ 0.9775\ 0.9698\ 0.9297\ 0.9706\ 0.9801\ 0.0987\ 0.5146\ 0.9666)$$
$$= 0.0267$$

and,

$$\bar{\mu}_{\tilde{A}_j}(X_p) = \bar{\mu}_{\tilde{A}_{j1}}(x_{p1}) \cdot \bar{\mu}_{\tilde{A}_{j2}}(x_{p2}) \cdot \ldots \cdot \bar{\mu}_{\tilde{A}_{jn}}(x_{pn})$$
$$= prod\ (0.9846\quad 1\quad 1\ 0.9787\quad 1\quad 1\quad 0.7409\quad 0.9603\quad 1)$$
$$= 0.6855$$

For example, the WBCD dataset with two classes "benign" and "malignant", ($c = 2$), includes 444 benign samples and 239 malignant samples. The total grade of compatibility, $\tilde{\beta}_{Class\,1} = \left[\underline{\beta}_{Class\,1}\ \bar{\beta}_{Class\,1}\right]$, of all benign samples ($Class$ 1) to the rule R_j: *"If* x_{p1} *is Mid and* x_{p2} *is Low and* x_{p3} *is Low and* x_{p4} *is Low and* x_{p5} *is Low and* x_{p6} *is Low and* x_{p7} *is Low and* x_{p8} *is Low and* x_{p9} *is Low"*, is computed as follows:

$$\beta_{Class\,1} = \frac{\underline{\beta}_{Class\,1} + \bar{\beta}_{Class\,1}}{2} = \frac{18.1640\ +\ 89.5836}{2}$$
$$= 53.8738$$

where as $\underline{\beta}_{Class\,1} = 18.1640$, and $\bar{\beta}_{Class\,1} = 89.5836$ are the lower and upper bound of the grade of compatibility for all benign samples, calculated using equations 6.9 and 6.10. The total grade of compatibility of all malignant samples ($Class$ 2) to the rule R_j is calculated in the same way as follows:

$$\beta_{Class\,2} = \frac{\underline{\beta}_{Class\,k} + \bar{\beta}_{Class\,k}}{2} = \frac{1.2654e-033\ +\ 1.2599e-006}{2}$$
$$= 6.2993e-007$$

where as $\underline{\beta}_{Class\,2} = 1.2654e-033$, and $\bar{\beta}_{Class\,2} = 1.2599e-006$ are the lower and upper bound of the grade of compatibility for all malignant samples, calculated using equations 6.9 and 6.10. This example shows that the benign samples in the WBCD dataset are more compatible to the rule R_j.

Step2: The consequent C_j of the rule R_j is the class with maximum total grade of compatibility ($\beta_{Class\,k}$) as (Mendel, 2001),

$$\beta_{Class\,\tilde{k}} = max\ \{\ \beta_{Class\,1},\quad \beta_{Class\,2}\,,\dots,\quad \beta_{Class\,c}\} \tag{6.12}$$

If $\beta_{Class\,\tilde{k}}$ is not unique, then two or more classes have the same $\beta_{Class\,\tilde{k}}$ then $\beta_{Class\,\tilde{k}}$ is assigned to $\emptyset$, the empty class. A rule with the empty class in the consequent part is called a dummy rule (Ishibuchi and Nojima, 2008). It is clear that the firing strength of the interval type-2 fuzzy sets in the antecedent of the rules has influence on defining the consequent of the rules.

For the above example for the WBCD classification, the consequent of the rule R_j is defined using,

$$\beta_{Classk} = max\ \{\ \beta_{Class\,1},\quad \beta_{Class\,2}\}$$

$$= max\,\{53.8738, 6.2993e - 007\} = 53.8738$$

Thus, $k = 1$ is selected, which means consequent C_j of the rule R_j is class "benign".

Step3: The grade of certainty GC_j of a dummy rule is assigned to 0. The grade of certainty of a non-dummy rule is defined as follows (Ishibuchi, Nakashima and Murata, 1996; Ishibuchi and Yamamoto, 2005):

$$GC_j = \frac{\beta_{Class\,k} - \acute{\beta}}{\sum_{l=1}^{c} \beta_{Class\,l}} \tag{6.13}$$

$$\acute{\beta} = \left(\sum_{k \neq l}^{c} \beta_{Class\,k}\right) \Big/ (c-1) \tag{6.14}$$

For the WBCD dataset example, the grade f certainty (GC_j) of the rule R_j is calculated as follows:

$$GC_j = \frac{\beta_{Class\,1} - \beta_{Class\,2}}{\sum_{l=1}^{2} \beta_{Class\,l}}$$

$$= \frac{53.8738\ - (6.2993e - 007)}{53.8738 + (6.2993e - 007)} = \frac{53.8738}{53.8738} = 1$$

This example shows that the degree of certainty of the rule R_j in classifying benign objects is 1.

This section presented the method for computing the consequent and the grade of compatibility of a rule in an IT2FLS using training datasets. The steps of the algorithm were explained with a simple example of the WBCD dataset. The next section considers that a set of rule for an IT2FLS has already been generated from a training dataset using the method explained in this Section, and it describes fuzzy reasoning method for classifying new patterns.

6.3.2 Fuzzy reasoning

A set of linguistic rules is selected using the rule generation method explained in the previous section. The process of fuzzy reasoning is applied for classifying new patterns. Consider we are given a subset S of all possible rule sets, a new pattern X_p is classified by the $Class\, k$ in the consequent of the rule $R_j \in S,\ (j = 1..M)$ which has the maximum $\tilde{\alpha}_j$ as,

$$\begin{aligned}\tilde{\alpha}_j = [\underline{\alpha}_j \quad \bar{\alpha}_j] &= [max\,\{\underline{\mu}_j(X_p).GC_j\} \quad max\,\{\bar{\mu}_j\,(X_p).GC_j\}] \\ &= [\,max\,\{\underline{\mu}_{\tilde{A}_{j1}}(x_{p1}) \cdot \underline{\mu}_{\tilde{A}_2}(x_{p2}) \cdot \dots \cdot \underline{\mu}_{\tilde{A}_{jn}}(x_{pn}).\,GC_j\} \\ &\quad max\,\{\bar{\mu}_{\tilde{A}_{j1}}(x_{p1}) \cdot \bar{\mu}_{\tilde{A}_{j2}}(x_{p2}) \cdot \dots \cdot \bar{\mu}_{\tilde{A}_{jn}}(x_{pn})\ .GC_j\}]\end{aligned} \tag{6.15}$$

The centroid type reduction following defuzzification for an interval type-2 fuzzy set $\tilde{\alpha}_j$ is used to calculate $\alpha_j(X_p)$ as follows,

$$\alpha_j(X_p) = \frac{\underline{\alpha}_j + \bar{\alpha}_j}{2} \tag{6.16}$$

Thus, the pattern X_p is classified by class label (k) in the consequent of the rule R_j with maximum $\alpha_j(X_p)$ (Ishibuchi and Nojima, 2008) as,

$$\alpha_{Class\,k} = max\,\{\,(\alpha_j(X_p)|\,Class_k = C_j\ and\ R_j \in S\} \tag{6.17}$$

For classifying a new pattern X_p, the value of $\alpha_j(X_p)$ is computed for each rule (R_j)in the rule set. The winner rule is the rule with maximum $\alpha_j(X_p)$. If two or more rules with different consequent classes take the maximum $\alpha_{Class\,k}$, then we assign -1 to $\alpha_{Class\,k}$ which means the given pattern X_p cannot be classified by rule set S and is an unclassified pattern (Ishibuchi and Nojima, 2008). The value of the $\alpha_{Class\,k}$ shows the strength of each rule (R_j) for assigning pattern X_p to its consequent class $(Class\ k)$. The interval of $[\underline{\alpha}_j \quad \bar{\alpha}_j]$ of the rule with maximum $\alpha_j(X_p)$ defines the lower and upper bounds of the strength of the rule.

For illustration, consider we want to classify the given pattern $X_p = (x_{p1}, x_{p2}, \dots, x_{pN}) = (5,1,1,1,2,1,3,2,1)$ and there are two rules in an IT2FLS as follows:

R_1: *If* x_{p1} *is Mid and* x_{p2} *is Low and* x_{p3} *is Low and* x_{p4} *is Low and* x_{p5} *is Low and* x_{p6} *is Low and* x_{p7} *is Low and* x_{p8} *is Low and* x_{p9} *is Low then it is benign with* $GC = 1$

and

R_2: *If* x_{p1} *is High and* x_{p2} *is Low and* x_{p3} *is Low and* x_{p4} *is Low and* x_{p5} *is Low and* x_{p6} *is Low and* x_{p7} *is Mid and* x_{p8} *is Low and* x_{p9} *is Low then it is benign with* $GC = 0.95$

The pattern X_p is classified by the $Class\ k$ in the consequent of the rule R_j (the winner rule) which has the maximum $\tilde{\alpha}_j$ as follows:

$$\begin{aligned}\tilde{\alpha}_1 = [\underline{\alpha}_1 \quad \bar{\alpha}_1] &= [\{\underline{\mu}_1(X_p).GC_1\} \quad \{\bar{\mu}_1(X_p).GC_1\}] \\ &= [\{0.0267 \times 1\} \quad \{0.6855 \times 1\}] \\ &= [0.0272 \ \ 0.6855]\end{aligned}$$

$$\begin{aligned}\tilde{\alpha}_2 = [\underline{\alpha}_2 \quad \bar{\alpha}_2] &= [\{\underline{\mu}_2(X_p).GC_2\} \quad \{\bar{\mu}_2(X_p).GC_2\}] \\ &= [\{0 \times 0.95\} \quad \{0.004 \times 0.95\}] \\ &= [0 \ \ 0.0038]\end{aligned}$$

then,

$$\alpha_1(X_p) = \frac{0.0272 + 0.6855}{2} = 0.3564$$

$$\alpha_2(X_p) = \frac{0 + 0.0038}{2} = 0.0019$$

and $\alpha_{Class\ k} = max\{\ 0.3564, 0.0019\} = 0.3564$. Thus, the strength of the rule R_1 is more than the rule R_2 which means the winner rule for classifying the pattern X_p is rule R_1. The next section describes an interval type-2 fuzzy GA approach for rule extraction and learning using an imprecise and imperfect training dataset.

6.4 An interval type-2 fuzzy GA approach for uncertain linguistic rules learning

The applied method for GA linguistic rule learning (explained in Section 6.3) extracts type-2 fuzzy rules for an IT2FLS. The proposed method in this section is based on the method presented by Ishibuchi et al. (Ishibuchi and Nojima, 2008) for extracting rules for a type-1 fuzzy logic and extends and improves it and its mathematical model for interval type-2 fuzzy logic. Furthermore, some changes have been made to the chromosome structure, and fitness evaluation of the proposed GA algorithm. The GA evolutionary algorithm follows the Michigan approach (Herrera, 2008, Pena-Reyes and Sipper, 1999) in which each individual represents a rule and a population in each run of the GA represents a candidate interval type-2 fuzzy logic. This evolutionary method has less computational complexity compared to the Pittsburg approach (Herrera, 2008, Pena-Reyes and Sipper, 1999) in which each individual in the population encodes a type-2 fuzzy logic and several candidate fuzzy logics are evaluated in each generation. The general steps of the type-2 fuzzy GA rule learning method are described in this section and demonstrated in Figure 6.2.

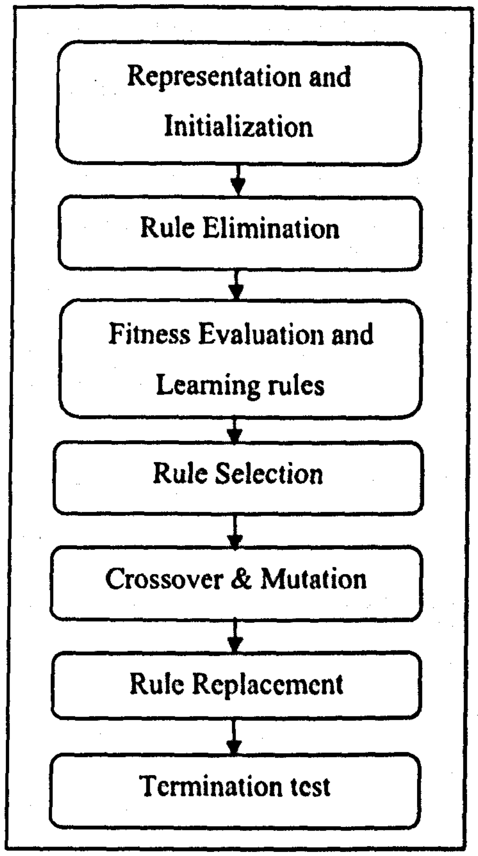


Figure 6.2: Genetic algorithm for learning IT2FL rules

Step1: *GA representation* – Real-valued codes in a chromosome are used to model rules in an IT2FLS. This structure is more applicable for pattern classification, easy to implement, has no length limitation and has more precision than bit string representation. It is also capable to model rules with a large number of antecedents (input features) with less number of genes compared to a bit string coding representation. The numerical structure of the chromosome for type-2 fuzzy rules consists of n genes, each showing presence or absence of a fuzzy linguistic term in a rule. Each genome has a value between 0 and the number of fuzzy sets associated to the linguistic terms of the feature. Figure 6.3 illustrates the chromosome structure, for an IT2FL with nine input features and three linguistic terms (low, middle and high). Each gene can have any values between 0 to 3; 0 means no contribution of the feature in the rule (don't care) and 1 to 3 represent three linguistic terms, respectively. It should be noted that the chromosome only models the antecedent part of the rule and the consequent part and grade of certainty of each rule is defined through the rule extraction method explained in the section 6.3.

1	0	0	2	1	3	0	0	0

Figure 6.3: Chromosome structure with numerical coding

***Step2**: Initialization* – In this structure, each rule is treated as an individual and a population (N_{pop}) consists of a fixed number of rules. The solution is a set of rules ($R_1 R_2$ ... R_n) which has the maximum fitness value. Each population represents a potential solution (rule set). To each genome related to a linguistic term, a random number between 0 to maximum number of linguistic terms (L) is associated. Then, the consequent and the grade of uncertainty of the rule are calculated using the method described in Section 6.3.

***Step3:** Rule elimination* - in this step, dummy rules (rules with the empty class in the consequent part) are first eliminated from the list of rules. Then the rules which do not classify any patterns (non-active rules) are removed from the rule set.

***Step4:** Fitness evaluation* - A fitness value is assigned to each linguistic classification rule R_j in the current generation as follows:

$$Fitness(R_j) = \left(W_{NCP} * N_{NCP}(R_j) - W_{NMP} * N_{NMP}(R_j)\right) \qquad (6.18)$$

where N_{NCP} is the number of patterns correctly classified by rule R_j, N_{NMP} is the number of misclassified patterns, and W_{NCP} and W_{NMP} are the weights of correctly classified and misclassified patterns, respectively.

Furthermore, during the process of classifying patterns in the training dataset by each rule, the initial grade of certainty (GC_j) of each rule is adjusted through a learning technique which employs a reward or punishment scheme (Ishibuchi and Nojima, 2008). For this, whenever a pattern is correctly classified then its grade of certainty is increased as,

$$GC_j^{new} = GC_j^{old} + W_{reward} * (1 - GC_j^{old}) \tag{6.19}$$

where W_{reward} is a positive learning weight. Otherwise, the grade of certainty of the rule is decreased (Ishibuchi and Nojima, 2008):

$$GC_j^{new} = GC_j^{old} - W_{punishment} * GC_j^{old} \tag{6.20}$$

where $W_{punishment}$ is the positive punishment weight.

***Step4:** Rule Selection* - This step applies a selection strategy which keeps the fittest chromosomes in the current population in the next generation. This selection method considers that the fitness of the individuals is positive. This step first calculates a probability of selection (P) for each rule in the current population by normalizing the fitness value of each rule compared to minimum fitness value of the rules in the current population, $f_{min}(S)$. Then it sorts the individuals based on their probability and selects 50% of the rules with higher selection probability. The probability of rule j in the gth generation is defined using a roulette wheel selection with a linear scale as follows (Ishibuchi, Nakashima and Murata, 2001):

$$P(R_j, g) = \frac{f(R_j) - f_{min}(S)}{\sum_{R_i \in S}\{f(R_i) - f_{min}(S)\}} \tag{6.21}$$

$$f_{min}(S) = min\{f(R_i) \mid R_i \in S\} \tag{6.22}$$

where $f(R_j)$ is the fitness of rule j, and $f_{min}(S)$ is the minimum fitness in the current generation (g).

***Step5:** Crossover and mutation* – crossover operation selects individuals in a population with a probability, $P_{Crossover}$ as parents and combines them to create a new offspring (Michalewicz, 1996). For each pair of selected individuals, a random number (r) in interval [0, 1] is generated. If $r < P_{Crossover}$, a single-point crossover is applied (See Figure 6.4). All the bits before and after that point in parent string are swapped. A position point, p, is randomly selected while $1 \leq p \leq n$, and n is the number of inputs (Michalewicz, 1996).

Parent 1:

a_1	a_2	...	a_{p-1}	a_p	a_{p+1}	...	a_{n-1}	a_n

Parent 2:

b_1	b_2	...	b_{p-1}	b_p	b_{p+1}	...	b_{n-1}	b_n

Child 1:

a_1	a_2	...	a_{p-1}	a_p	b_{p+1}	...	b_{n-1}	b_n

Child 2:

b_1	b_2	...	b_{p-1}	b_p	a_{p+1}	...	a_{n-1}	a_n

Figure 6.4: An example of crossover operation on two individuals at crossover point, p

The offspring are normally mutated after crossover combination. The mutation operation selects a gene in a chromosome with a mutation probability, P_m, and changes it to a new possible value (Michalewicz, 1996). The probability of mutation defines the expected number of genes in a chromosome undergoing mutation operation. The probability of the mutation, P_m, is inversely proportion of the length of the chromosome. Mutation is usually applied for each offspring in the current population and for each gene within them. It first generates a random number (r) in the range [0, 1] then compares it with the probability of mutation. The gene is mutated if $r < P_M$ (Michalewicz, 1996). The new value of the chosen gene (Shown with * in Figure 6.5) is replaced with a new uniform random value between 0 to maximum number of linguistic terms (L).

				*				
a_1	a_2	...	a_{r-1}	a_r	a_{r+1}	...	a_{n-1}	a_n

Figure 6.5: An example of mutation operation on an individual at random mutation point *

The mutation and crossover operation apply only to the antecedent part of a rule. Then the consequent and the grade of certainty of rules are defined according to the rule generation method in Section 6.3.

The crossover operation directly affects the convergence property of the GA while mutation provides variation for the GA (Koza and Poli, 2005). High probability of crossover and low probability of mutation can improve convergence of the GA. Premature convergence is a common problem in the genetic algorithm which means the algorithm is trapped into local optima rather than the global optima. Selection of the optimum individual in each population does not assure the optimum solution for the problem. However, large population of chromosomes can ameliorate this problem but does not guarantee all variations. Thus, mutation is a useful technique to improve the genetic diversity and avoid premature convergence to a local optimum solution (Koza and Poli, 2005).

Step6: *Rule Replacement* - the generated rules in the previous step are replaced with some of the rules in the current population. The replacement strategy is based on the fitness value of the rules which means rules with lower fitness values are replaced by new rules.

Step7: *Termination test-* the GA algorithm stops whenever it gets to the maximum number of generations.

6.4.1 The GA overfitting problem

One of the major problems in most machine learning algorithms evolving GA is overfitting. An algorithm overfits a dataset when it models the given samples very well but cannot predict non-observed samples. This problem occurs when the model overfits the train set but it does not generalize well for the new samples in the test set. This means, during the GA run the error on the training dataset is decreasing but it is increasing on a non-observed testing dataset. However, complex rules with several antecedents or large number of rules in a rule set improves the accuracy of classification over the training set, but this technique may affect the generalization property of the selected rule set which may cause the performance to deteriorate over the testing dataset.

Overfitting can be avoided if the number of samples is much more than the number of parameters in the GA. Another solution is to apply a reduce error pruning idea in a decision tree classifier to remove the rules with low classification rates (Witten and Frank, 2005). This approach splits the whole dataset into a training set (also called growing) and a testing set (pruning) and monitors the classification error on both sets. There are several techniques for splitting a dataset into training and testing sets. One of the common technique is to consider two thirds of the dataset for training and the rest for testing. The samples in each group are randomly selected. The reduced error pruning method first extracts a complex rule set with a high accuracy on the training set then it prunes the rule set to fit it for the testing dataset (Witten and Frank, 2005).

The proposed approach for learning rules in this study applies a post-pruning stage to remove rules, which ameliorate the classification rate on the training dataset but deteriorate the classification rate on the testing dataset. The next section describes the importance of the breast cancer diagnosis problem.

6.5 Breast cancer diagnostic problem

Breast cancer is the most common type of the cancer among women in the UK (Cancer Research UK, 2011). Each year about 45,000 female and 300 male are diagnosed with breast cancer and 12,000 women and 90 men die. In 2008, a total of 48,034 new cases were diagnosed (over 99% in women and less than 1% in men) (Cancer Research UK, 2011). The risk factor of breast cancer among women is 1 in 9 in 2008 and varies in different ages (Cancer Research UK, 2011). As shown in Figure 6.6 (Ferlay *et al.*, 2011), women (under the age of 75) are at high risk (incidence and mortality) of breast cancer than younger women (under the age of 25). Breast cancers in the early stages are more likely curable. Mammographic screening (X-ray imaging of the breast) is a way to find a lump in the breast and detect the cancer at the early stage. Detection of the lumps in the early stage improves the survival rate. Radiologists observe mammogram images to find abnormalities such as micro-classification (small deposits of calcium), masses (space-occupying lesion in two projections of the breast, see Figure 6.7), and structural disorders and characterization of them (shape or margin) (Sampat, Markey and Bovik, 2005). However, radiologist's diagnosis is associated with human errors. Biopsy is a sure way to identify the breast cancer; it is an invasive method which is performed through a tiny needle or surgery for taking sample tissue of the lump. Computer aided mammography, as an invasive method, helps radiologists to detect abnormalities in the breast images and improve diagnosis the breast cancer.

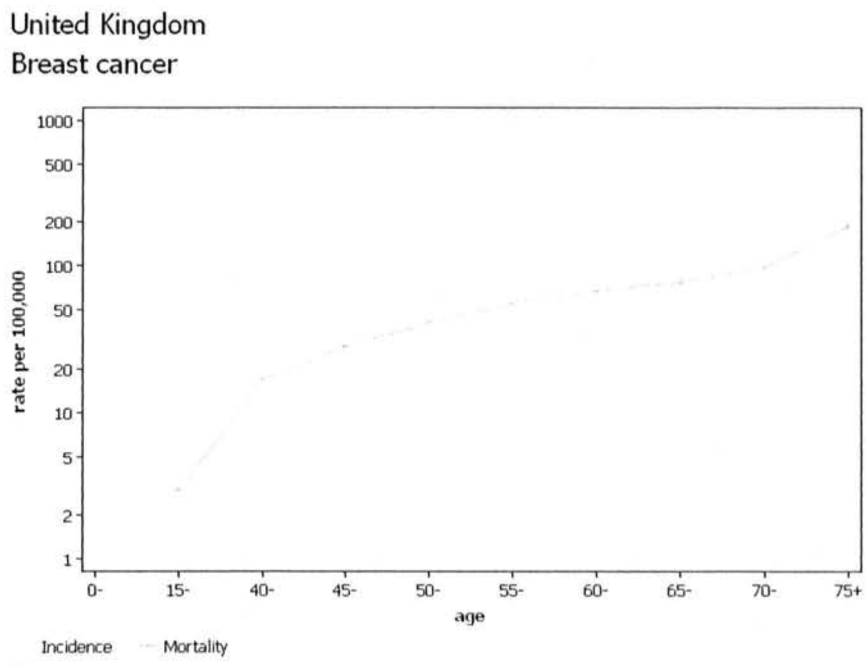


Figure 6.6: Incidence and mortality age-specific breast cancer rate for women per 100,000 in the UK in 2008 (Ferlay *et al.*, 2011)

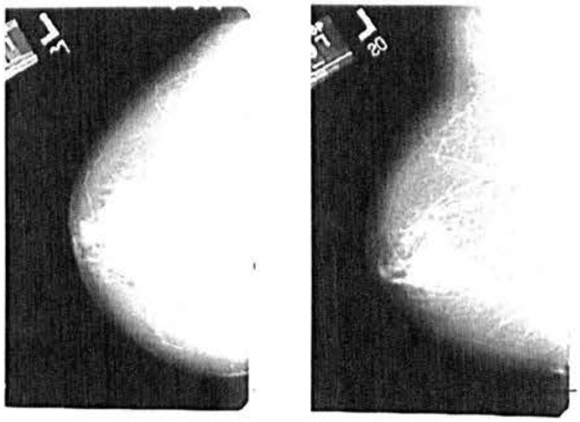


Figure 6.7: Two view projections of the breast through mammographic screening from (Sampat, Markey and Bovik, 2005)

6.6 Review of the classification methods for the WBCD classification problem

This section provides an overview of the previous rule extraction methods applied to the Wisconsin breast cancer diagnostic (WBCD) dataset. This dataset is a popular dataset for evaluation of the performance of classification. Various techniques have been reported in the literature to classify the WBCD, including naive Bayesian, neural networks, support vector machines, fuzzy approaches and a combination of learning and data mining approaches with these methods, such as neuro-fuzzy, fuzzy decision tree and genetic fuzzy approaches. A review of neural network classification methods can be found in (Abbass, 2002; Fogel and Wasson, 1995). A survey of improved naive Bayes methods for classification is presented in (Jiang et al., 2007). Fuzzy rule-based classification provides comprehensive rules with linguistic terms which are a suitable choice for interaction with medical experts. On the other hand, accuracy and interpretability are two important contradictory goals in fuzzy rule-based classification. Evolutionary learning techniques are one of the commonly applied methods for improving the tradeoffs between accuracy and interpretability in fuzzy rule-based classifiers. Several attempts have been reported to improve the trade-off using multi-objective evolutionary algorithms (Ishibuchi and Nojima, 2007; Ishibuchi and Yamamoto, 2003; Ishibuchi, Kaisho and Nojima, 2009; Alcalá et al., 2009). The rest of this section presents an overview of various rule extraction methods in the rule-based classifiers applied to the WBCD while the main focus is on a fuzzy-based approach.

6.6.1 Rule extraction methods based on deterministic approaches

Setiono introduced a rule extraction method from a pruned neural network for breast cancer diagnosis in (Setiono, 1996). In this method network pruning is used to remove the redundant connections and decrease the complexity of the network (Setiono, 1996). Their rules achieved more than 95% accuracy on training and testing datasets. In another work (Setiono, 2000), the samples with missing values were removed and feature extraction was applied as a pre-processing method on the dataset to improve the accuracy of the classifier.

Taha and Ghosh presented an approach for extracting rules from a trained neural network (Taha and Ghosh, 1996). In their method, the final decision is made with a confidence measure. They introduced three methods for extracting the rules:

1. A binarized input-output rule extraction (BIO-RE) which extracts binary rules from a neural network trained with binary inputs. If the inputs are not binary they need to be binarized.
2. A partial rule extraction (Partial-RE) which searches for the incoming link to activate a hidden or output node and calculates a measure of belief. This method has a lower computational cost and a lower number of premises per rule
3. A Full-RE generates intermediate rules by considering the effect of each of the inputs on the consequent using linear programming and an input discretization method.

The first method is not effective for a large number of inputs and binarizing input values can degrade the network performance. The Partial-RE methods tackle these issues by extracting certain rules with a small number of rule premises. This method needs all inputs to be continuous and has the same range while there is no restriction on input values in the Full-RE method. Furthermore, the first two rule extraction methods based on the trained neural network may need a default rule (rules that cover input-output samples and are not extracted by a neural network) to classify new samples and their accuracy drops significantly without the default rule (Taha and Ghosh, 1996).

Recently, support vector machines (SVM) have been applied to the classification of the WBCD (Akay, 2009). Akay presented a SVM technique combined with a feature selection method (Akay, 2009). This method provides a maximum 98.51% ROC accuracy. An overview of the SVM rule extraction methods are provided in (Martens et al., 2007). Martens et al. added comprehensibility to the SVM by extracting symbolic rules from the trained model in order to make it appropriate for medical diagnosis. This method provides an accuracy of 96.6% (Martens et al., 2007).

6.6.2 Rule extraction methods based on fuzzy techniques

Comprehensibility is an essential factor in medical imaging applications where the reader (such as a radiologist) requires interaction with the system for analysis of the result and making the final diagnosis. Fuzzy rule based methods for classification of the WBCD problem provide a rule set with more interpretability than the other proposed methods. Furthermore, they are capable of capturing the uncertainties in the system and provide a rule set which makes nondeterministic decisions that allows an object to belong to different classes with different degrees of membership. For these reasons, the focus of this study is on a fuzzy-based approach for classification of the WBCD problem.

A fuzzy genetic approach was reported in (Pena-Reyes and Sipper, 1999) for the classification of the tissue mass for the WBCD. However the output is more reliable than deterministic methods because of adding a confidence measure to the classifier decision (malignant or benign), but the classifier output is based on a threshold which is manually defined by the user. Moreover, it utilizes prior knowledge of the previous proposed rule sets to initialize the membership function parameters and rules of the fuzzy system in the genetic algorithm (Pena-Reyes and Sipper, 1999).

An evolutionary approach for designing the fuzzy classifier from a training dataset was presented by Chang and Lilly which uses parameters of a variable input spread inference training (VISIT) algorithm (Chang and Lilly, 2004). Although their algorithm employs GA for optimizing the fuzzy system parameters, this method needs initialization of the spread, alpha-cut values, and the degree of overlap between adjacent membership functions by a fuzzy expert before starting the algorithm and the GA optimization (Chang and Lilly, 2004).

Ishibuchi introduced an evolutionary multi-objective optimization (EMO) rule selection approach for a fuzzy rule-based system (Ishibuchi and Yamamoto, 2003). The algorithm considers three objectives in the fitness function: maximizing the number of correctly classified patterns, minimizing the number of rules, and minimizing the length of the rules. This algorithm first extracts a pre-specified number of candidate rules from all possible candidate rules using a heuristic rule extraction method and data mining criterions for rule evaluation (SLAVE criterion) and then an evolutionary method finds a rule set using a Genetic rule selection method. There are $(L+1)^n$ possible antecedents for a FLS with n inputs and L linguistic terms, considering "don't care" condition. A rule extraction algorithm removes unnecessary rules that do not classify any pattern (Ishibuchi and Yamamoto, 2003).

Wang presented a self-adaptive Neuro-fuzzy inference system (SANIF) for a self-organizing internal network structure and extracting the rules (Wang and Lee, 2002). This approach was implemented with three types of rules: output fuzzy sets, an output singleton and an output consisting of the linear combination of input variables. There are two major stages in this algorithm (Wang and Lee, 2002):

1. A clustering algorithm to identify the internal structure in which the number of fuzzy rules is close to the true number of clusters in the training data and each cluster represents a fuzzy rule; a mapping-constrained agglomerative (MCA) clustering algorithm was used for this problem
2. A recursive linear/nonlinear least-squares optimization algorithm to accelerate the convergence of the learning algorithm and tune the weights using a similarity measure.

However, this method initializes the fuzzy membership function parameters (mean and variance) in the antecedent and consequent of rules using the MCA clustering algorithm but needs expert knowledge for initializing the number of seed clusters and the number of hidden layers in the neural network. This algorithm is sensitive to the initial number of seed clusters which must be greater than the true number of clusters. However, knowledge about the true number of clusters is not always available.

Abonyi et al. presented a decision tree-based (DT based) initialization of a fuzzy classifier method in (Abonyi, Roubos and Szeifert, 2003). In their method, the GA is used to tune the fuzzy system parameters and improve classification performance (Abonyi, Roubos and Szeifert, 2003). However, this approach generates the most interpretable rule set among reported methods with comparable accuracy but it initializes the fuzzy classifier based on a crisp decision tree extracted from an expert's knowledge and pruning the initial tree which can be complex because of noise; moreover it inherits the complexity of the decision tree model.

Most of the presented fuzzy methods need expert knowledge for initializing the fuzzy set's membership functions, initial rule set or making the final classification decision. However, extracting this knowledge from a medical experts is challenging, time consuming and expensive process. The automatic approach presented in this chapter for generating fuzzy set membership functions and extracting rules overcomes these issues. The next section presents the results of applying the proposed IT2F GA rule extracting approach to the WBCD classification problem and a comparison of the results with other fuzzy approaches.

6.7 Experimental results and performance evaluation

The GA IT2FL rule learning approach for modelling the uncertainties in membership function and rules (explained in Section 6.4) is applicable for all uncertain classification problems with an incomplete and imprecise datasets. This approach has been applied to the popular Wisconsin breast cancer diagnosis (WBCD) database. The details of the components of the software developed in this study for IT2FL rules and membership function generation and implemented components are presented in Appendix A (Figures A.1 and A.2). The features in this database are computed using image processing techniques and are visually assessed by an expert. Furthermore, uncertainty in image processing techniques in addition to the noisy and variations in this dataset makes it a suitable choice for performance evaluation of the proposed approach.

The WBCD dataset was collected at the University of Wisconsin (Frank and Asuncion, 2010). The samples contain visual assessment of the nuclear features of fine needle aspirates (FNAs) taken from patients breasts by Dr. Wolberg and are available online at UCI (University of California at Irvine) machine learning depository (Frank and Asuncion, 2010). This dataset contains 699 samples. The 16 examples with missing features were removed, as has been done in previous studies (Abbass, 2002; Fogel and Wasson, 1995; Jiang *et al.*, 2007; Ishibuchi and Yamamoto, 2003; Setiono, 2000; Taha and Ghosh, 1996; Akay, 2009; Chang and Lilly, 2004; Wang and Lee, 2002; Abonyi, Roubos and Szeifert, 2003). There are nine integer features, each has an integer value between 1 and 10, value 1 corresponding to a normal state and 10 to the most abnormal state. The two output classes, indicating either a benign or malignant sample, are distributed as 444 benign (65%) and 239 malignant (35%). According to the (Wolberg and Mangasarian, 1990) features are described as the extent to which epithelial cell aggregates is considers as clump thickness, marginal adhesion is the cohesion of the peripheral cells of the epithelial cell aggregates, the single epithelial cell size considers as the diameter of the population of the largest epithelial cells relative to erythocytes, the proportion of single epithelial nuclei that were devoid of surrounding cytoplasm considers as bare nuclei, blandness of nuclear chromatin, normal nucleoli, infrequent mitoses, uniformity of epithelial cell size, and uniformity of cell shape as follows:

1. Clump Thickness (CT)
2. Uniformity of Cell Size (UC)
3. Uniformity of Cell Shape (UCS)
4. Marginal Adhesion (MA)
5. Single Epithelial Cell Size (SEC)
6. Bare Nuclei (BN)
7. Bland Chromatin (BC)
8. Normal Nucleoli (NN)
9. Mitoses (M)

All malignant samples were histologically confirmed whereas benign samples were biopsied only at the patient's request. The rest of benign samples were confirmed by clinical re-examination 3 and 12 months after the aspiration. Masses that grows unsatisfactory or suspicious FNAs were surgically biopsied (Wolberg and Mangasarian, 1990).The rest of this section provides the results of applying the GA IT2FL rule learning approach to the WBCD classification problem. It also presents a comparison of performance of this approach with the best results of the other type-1 fuzzy methods for this dataset.

6.7.1 Estimation of the FOU parameters

In order to estimate the footprint of uncertainty parameters (k_m and k_v), the method explained in Section 6.2 was employed. The average cardinality, $P_{avg}(\tilde{A})$, was calculated for each possible pair of k_m and k_v and for corresponding interval type-2 fuzzy sets of the linguistic terms related to the 9 input features of the WBCD dataset using a ten-fold cross-validation run of the algorithm.

Table 6.1 Estimation of the FOU parameters

No	k_m	k_v	Average Cardinality	Error rate	Sum
1	0.1	0.4	0.3693	0.1702	0.5395
2	0.1	0.5	0.1268	0.2611	0.3879
3	0.1	0.6	0.0423	0.1742	0.2165
4	0.1	0.7	0.0164	0.1838	0.2002
5	**0.1**	**0.8**	**0.0070**	**0.1339**	**0.1410**
6	**0.1**	**0.9**	**0.0012**	**0.1745**	**0.1757**
7	0.1	1	0.0012	0.2213	0.2225
8	0.2	0.4	0.3723	0.2955	0.6678
9	0.4	0.3	1.0000	0.2623	1.2623
10	0.2	0.5	0.1503	0.1414	0.2918
11	0.2	0.5	0.1398	0.1166	0.2563
12	0.2	0.6	0.0458	0.4203	0.4661
13	**0.2**	**0.7**	**0.0082**	**0.1344**	**0.1426**
14	0.3	0.5	0.1233	0.1654	0.2887
15	0.3	0.5	0.1116	0.1794	0.2910
16	0.3	0.6	0.0411	0.3036	0.3447
17	**0.3**	**0.7**	**0.0076**	**0.1849**	**0.1926**
18	0.3	0.9	0.0012	0.2074	0.2086
19	0.4	0.5	0.0857	0.2066	0.2923
20	0.4	0.6	0.0458	0.3273	0.3731
21	0.5	0.5	0.1122	0.2456	0.3577
22	0.5	0.7	0.0059	0.2327	0.2386
24	0.6	0.5	0.0464	0.2216	0.2679
25	0.6	0.7	0.0023	0.4250	0.4273
26	0.6	0.9	0.0018	0.4118	0.4136
28	0.8	0.4	0.2008	0.3849	0.5857
29	0.8	0.5	0.0675	0.1987	0.2662
30	0.8	0.7	0.0029	0.2909	0.2938
32	0.9	0.4	0.1985	0.1618	0.3603
33	0.9	0.6	0.0112	0.4532	0.4643
34	0.9	0.8	0.0018	0.4014	0.4031
35	0.9	1	0.0123	0.2127	0.2251
36	1	0.4	0.2208	0.3057	0.5265
37	1	0.6	0.0106	0.3408	0.3514
38	1	0.8	0.0023	0.7020	0.7044
39	1	1	0.0147	0.2022	0.2168

Selected k_m and k_v parameters for IT2FMF generation are bold. The IT2FLSs with the minimum average cardinality (<0.009) and minimum average error rate (<0.2), are heuristically selected for the GA IT2FL rule learning algorithm.

The k_m and k_v with the minimum average cardinality and minimum average error rates on the training and testing sets is selected for the IT2F membership function generation in the GA IT2FL rule learning algorithm.

Table 6.1 summarizes a sample of results obtained in terms of average cardinality and average error rate for different k_m and k_v values obtained for IT2FLS with three linguistic terms and nine input features of the WBCD dataset while the average cardinality and average error rate are all normalized. The IT2FLSs with the minimum average cardinality (<0.009) and minimum average error rate (<0.2), are heuristically selected for the GA IT2FL rule learning algorithm. Table 6.1 shows a sample of results close to the optimum results, after applying above criteria the best k_m ranging $[0.1, 0.3]$ and the best k_v ranging $[0.7, 0.9]$ represent the IT2FLS with minimum average cardinality and overall classification rate (sum<0.2) for the WBCD dataset and have been selected for the IT2FLS GA rule learning algorithm.

6.7.2 The GA IT2FL rule learning algorithm

The GA IT2FL rule learning approach was applied to automatically extract rules for the WBCD dataset by searching the solution space in an optimized way to find the optimum rule set for an IT2FLS. Different reasonable values for the GA algorithm parameters were examined using a ten-fold cross-validation run of the algorithm. The average of a ten-fold cross validation results in terms of classification rate and area under the ROC curve were considered.

Different mutation and crossover rates and weights were examined for the (k_m , k_v) = (0.1, 0.8) in the following experiments and the results are presented. According to the obtained results in Table 6.2, the mutation rate 0.1 and crossover rate of 1 provides maximum performance with a classification rate 95.64% and AUC ROC equal to 98.2% and were selected for the GA IT2FL rule learning algorithm. In this study, the effect of different fitness weights on the classification rates was investigated and the results are summarised in Table 6.3. Although the (W_{NCP}, W_{NMP}) equals to (1, 0) produces a better classification rate, but the fitness weights were considered (1, 5) in order to consider the effect of incorrect classification as well as correct classification during the learning process. The effect of different punishment and reward weights on the classification performance was investigated in Table 6.4. According to the obtained results, the

(W_{Reward}, $W_{Punishment}$) values of (0.01, 0.1) provides a better classification rate (95.64%) and AUC ROC (98.2%) compared to the other ratios. The obtained weight results for the proposed approach on the WBCD dataset is in agreement with the suggested learning weights in (Ishibuchi, Nakashima and Murata, 1996).

Table 6.2 Relation between different mutation and crossover rates and classification performance

Mutation Probability	0.01	0.1	0.15	**0.1**	0.2	0.2
Crossover Probability	0.5	0.5	1	**1**	0.5	1
*Classification rate (%)**	94.89	93.38	94.65	**95.64**	79.84	94.39
*ROC (AUC) (%)**	97.8	97.68	97.7	**98.2**	97.29	98.11

* The number shows average result of a ten- fold cross-validation run of the algorithm

Table 6.3 Relation between different values of the classification weights and classification performance

(W_{NCP}, W_{NMP})	**(1, 0)**	**(1, 1)**	**(1, 2)**	**(1, 5)**	**(1, 10)**
*Classification rate (%)**	96.54	89.02	93.25	**95.64**	90.78
*ROC (AUC) (%)**	98	97.61	97.33	**98.2**	98.72

* The number shows average result of a ten- fold cross-validation run of the algorithm

Table 6.4 Relation between different values of the learning weights and the classification performance

(W_{reward}, $W_{punishment}$)	**(0.01, 0.1)**	(0.1, 0.1)	(0.1, 0.01)	(0.1, 0.5)
*Classification rate (%)**	**95.64**	94.56	93.74	94.68
*ROC (AUC) (%)**	**98.2**	97.41	98.10	98.30

* The number shows average result of a ten- fold cross-validation run of the algorithm

After several experiments, it was observed that the GA converged to the best fitness after a maximum of 100 generations. Thus, the termination condition of the GA IT2FL rule learning algorithm was considered 100 runs. The learning rule approach generates the best set of rules according to the fitness function. The summary of the selected parameters of the GA learning algorithm are as follows:

$Crossover\ rate = 1$

$Mutation\ rate = 0.1$

$W_{NCP} = 1, and\ W_{NMP} = 5$

$W_{reward} = 0.01, and\ W_{punishment} = 0.1$

$Maximum\ number\ of\ generation = 100$

The GA IT2FL rule learning algorithm was performed for the selected (k_m, k_v) parameters in Section 6.7.1. In order to have a robust and unbiased view of the performance of the classification, the classifier performance is measured at each run of the ten-fold cross-validation using the following measures: area under the ROC curve, overall classification rates, error on training dataset and error on testing dataset, and the average result of a ten-fold is considered as performance measures.

For illustrating the type-2 GA rule learning approach, the rest of this section explains four examples of extracted IT2FLS rules for the selected (k_m, k_v) parameters in detail. Different IT2FLSs according to the selected (k_m, k_v) parameters were investigated for the WBCD dataset. For selected pairs of (k_m, k_v), four experiments have been performed: (1) a ten-fold cross-validation run, (2) 50% training and 50% testing, (3) 75% training and 25% testing, and (4) 100% for training and 0% for testing. The samples in the training and testing datasets were randomly selected. The post-processing reduce error pruning method (explained in section 6.4.1) was also applied after training to remove the rules with low classification rates (rules for which the number of misclassifications is greater than the number of classifications).

The IT2FLS_1 generated rules have three linguistic terms (low, middle and high). The (k_m, k_v) parameters were selected as (0.1, 0.8) using the method explained in 6.7.1. The generated IT2F Gaussian MFs are shown in Figure 6.8. Four rules are generated for IT2FLS_1; two rules for classifying malignant objects are as follows:

If MA *is High and* BN *is High then object is malignant with DC 0.92*

If BN is High and BC is High then object is malignant with DC 0.8

and two rules for classifying a benign object as follows:

If UCS is Low and BN is Low then object is benign with DC 0.82

If CT *is Mid and* UCS *is Low and* BN *is Low then object is benign with DC 1*

The classification accuracy of each rule is shown in Table 6.5 while 75% of the dataset was considered for training and 25% for testing. The average rule length is 2.25 in the antecedent part. The error on the testing dataset is 5.26% and the error on the training dataset is 2.73%. This set of rules provides a classifier with an overall classification rate 96.63%. The average accuracy of this system is measured using area under the ROC curve which is 98.04% with 95% CI [96.79 99.29] as shown in Figure 6.9.

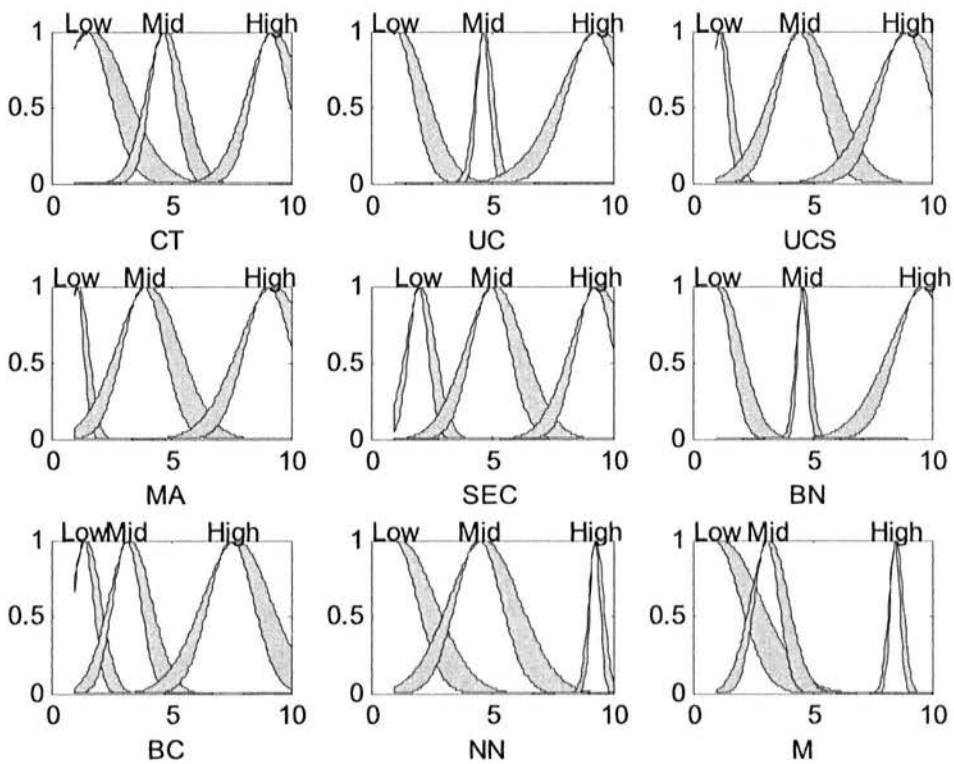


Figure 6.8: Membership function of IT2FLS_1, ($k_m = 0.1, k_v = 0.8$), for nine features of the WBCD

Table 6.5 Extracted rules of IT2FLS_1 for the WBCD dataset

No.	Antecedent	Class	Degree of certainty	NCP[1]	NMP[2]
R1	[0 0 1 0 0 1 0 0 0]	0	0.82	350	4
R2	[0 0 0 3 0 3 0 0 0]	1	0.92	60	3
R3	[2 0 1 0 0 1 0 0 0]	0	1	77	2
R4	[0 0 0 0 0 3 3 0 0]	1	0.8	173	14
			Sum =	660	23

*0, 1, 2, and 3 are equivalent to "don't care", Low, Middle and High linguistic terms

[1] Number of correctly classified patterns, [2] Number of misclassified patterns

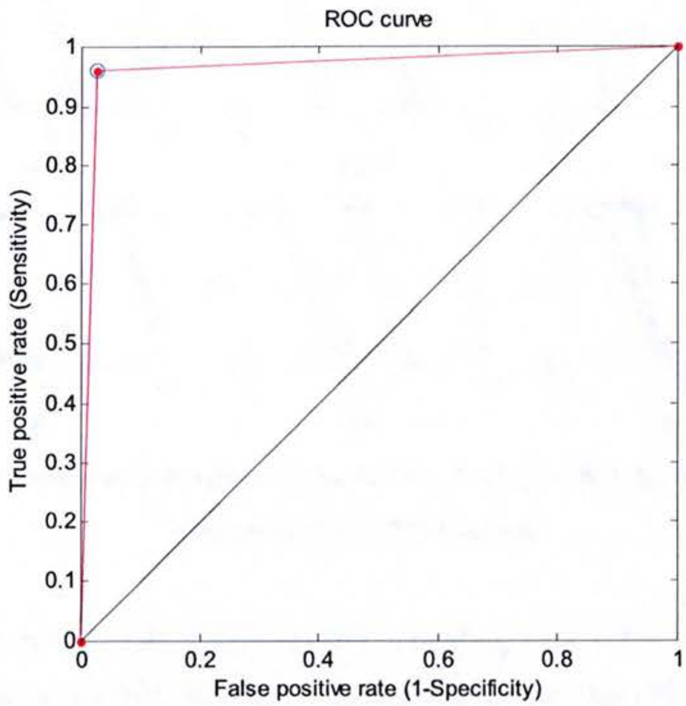


Figure 6.9: ROC curve of the IT2FLS_1 for the WBCD dataset

For the generated rules for the IT2FLS_2, the (k_m, k_v) parameters were selected as (0.1, 0.9). Figure 6.10 shows the IT2 Gaussian MFs generated for nine features of the WBCD dataset. Four rules with an average rule length of 1.5 variables per rule were extracted by the algorithm, shown in Table 6.6, three rules for classifying malignant samples as follows:

If MA is High and BN is High then object is malignant with DC 0.96

If CT is High and BC is High then object is malignant with DC 0.9

If BN is High then object is malignant with DC 0.65

and one rule for classifying benign objects as follows:

If UCS is Low then object is benign with DC 0.55

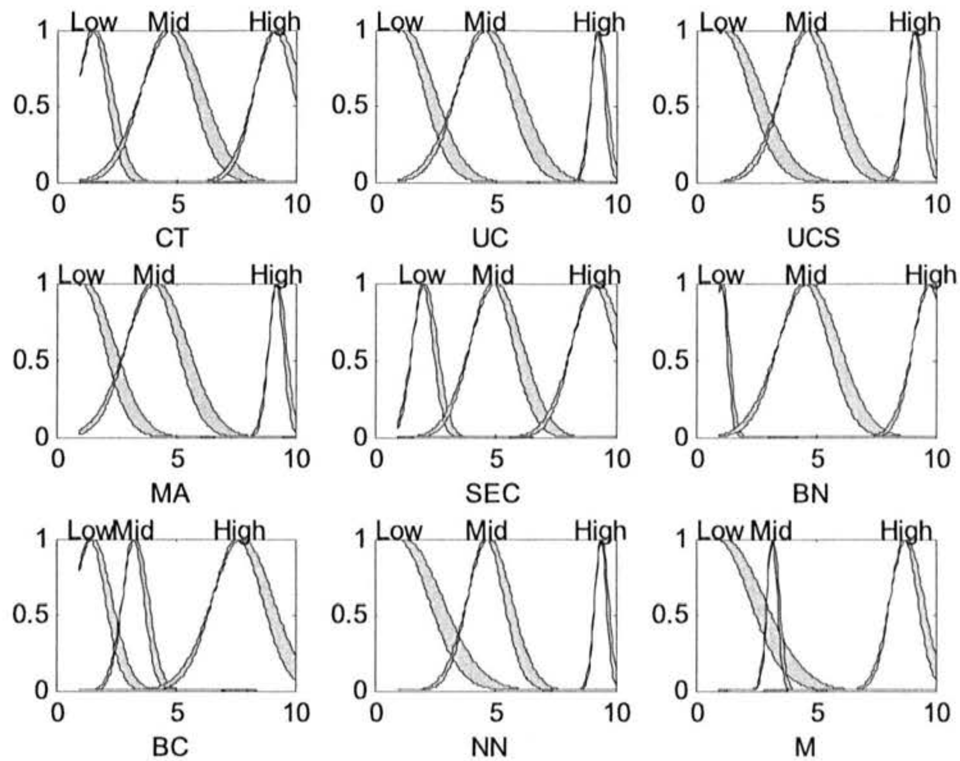


Figure 6.10: Membership function of the IT2FL_2, ($k_m = 0.1, k_v = 0.9$), for nine features of the WBCD dataset

This set of rules has a high interpretability including four rules and three linguistic terms and provides a classifier with the overall classification rate (95.75%) with an error rate on the training dataset of 3.71% and an error rate on the testing dataset of 5.85%. The performance of this classifier (75% of the dataset for training and 25% for testing) in terms of the AUC of the ROC is 97.45% with 95% CI [96.03 98.87] as shown in Figure 6.11.

Table 6.6 Extracted rules for IT2FLS_2 for the WBCD dataset

No	Antecedent	Class	Degree of certainty	NCP[1]	NMP[2]
R1	[0 0 1 0 0 0 0 0 0]*	0	0.55	422	7
R2	[0 0 0 3 0 3 0 0 0]	1	0.96	57	0
R3	[0 0 0 0 0 3 0 0 0]	1	0.65	150	22
R4	[3 0 0 0 0 0 3 0 0]	1	0.9	25	0
			Sum =	654	29

*0, 1, 2, and 3 are equivalent to "don't care", Low, Middle and High linguistic terms

[1] Number of correctly classified patterns, [2] Number of misclassified patterns

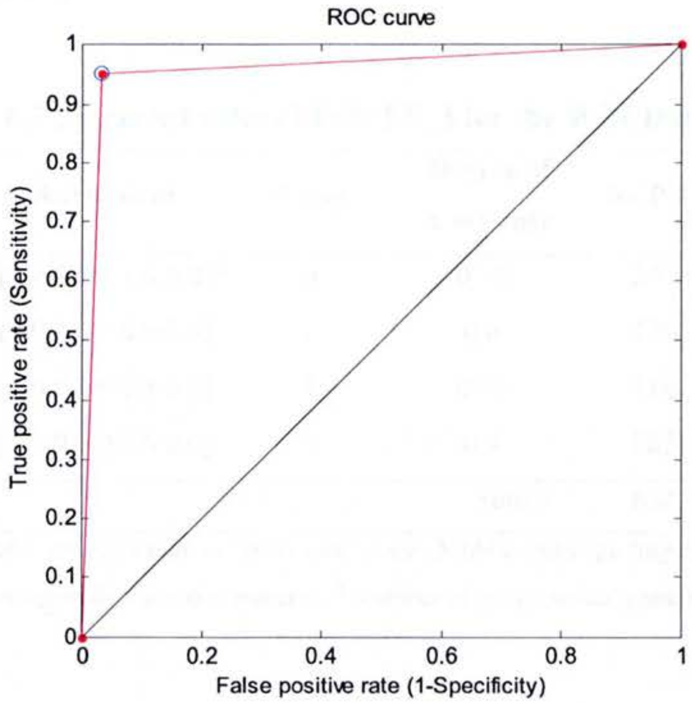


Figure 6.11: ROC curve of the IT2FLS_2 for the WBCD dataset

The third system, IT2FLS_3, generates rules with three term sets. The (k_m, k_v) parameters were selected as (0.2, 0.7). The generated IT2 Gaussian MFs of the IT2FLS_3 are shown in Figure 6.12. Four rules are generated for the IT2FLS_3 (while 75% of the dataset was considered for training and 25% for testing); two rules for classifying benign objects and two rules for classifying malignant objects as follows:

If MA is Low and SEC *is Low and BN is Low then object is benign with DC 0.58*

If CT is Low and UC *is Low then object is benign with DC 0.9*

If CT is High then object is malignant with DC 0.6

If BN is High and BC is High then object is malignant with DC 0.92

The classification rates and the degree of certainty of each rule is shown in Table 6.7. The average rule length is 2 in the antecedent part. This set of rules provides a classifier with an overall classification rate of 96.93%, with an error on the training set of 2.93% and an error on the testing set 3.51%. The average accuracy of this classifier is measured using the area under the ROC curve which is 98.6 % with 95% [97.54 99.66] as shown in Figure 6.13.

Table 6.7 Extracted rules of IT2FLS_3 for the WBCD dataset

No.	Antecedent	Class	Degree of Certainty	NCP[1]	NMP[2]
R1	[0 0 0 1 1 1 0 0 0]*	0	0.58	247	8
R2	[3 0 0 0 0 0 0 0 0]	1	0.6	120	11
R3	[0 0 0 0 0 3 3 0 0]	1	0.92	110	1
R4	[1 1 0 0 0 0 0 0 0]	0	0.9	185	1
			Sum =	662	21

*0, 1, 2, and 3 are equivalent to "don't care", Low, Middle and High linguistic terms

[1] Number of correctly classified patterns, [2] Number of misclassified patterns

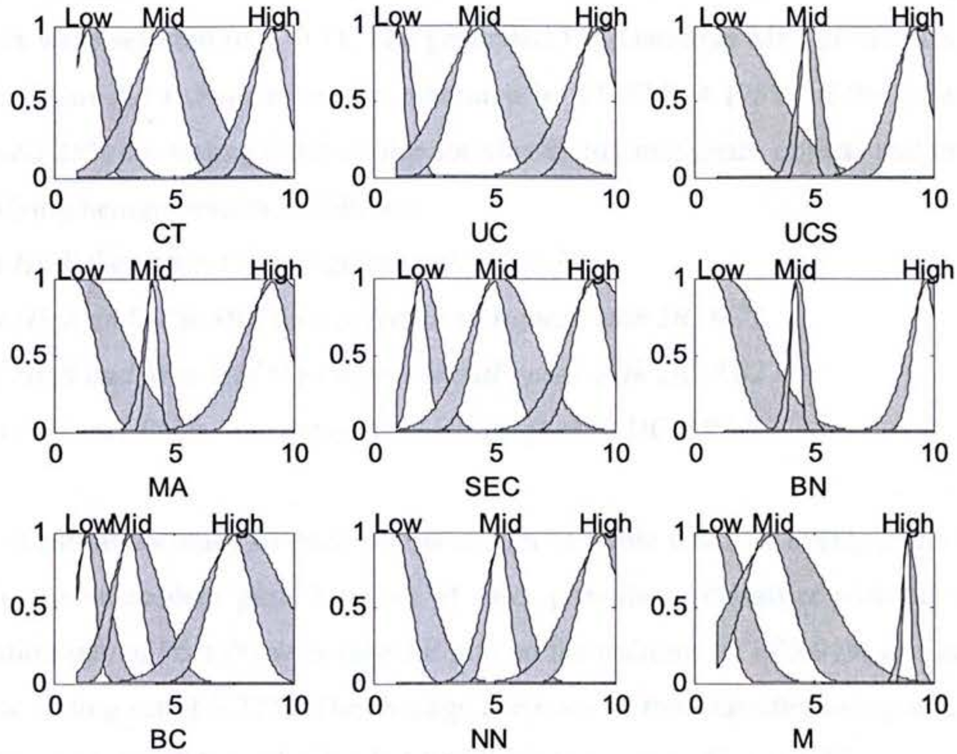


Figure 6.12: Membership function of IT2FLS_3, ($k_m = 0.2, k_v = 0.7$), for nine features of the WBCD dataset

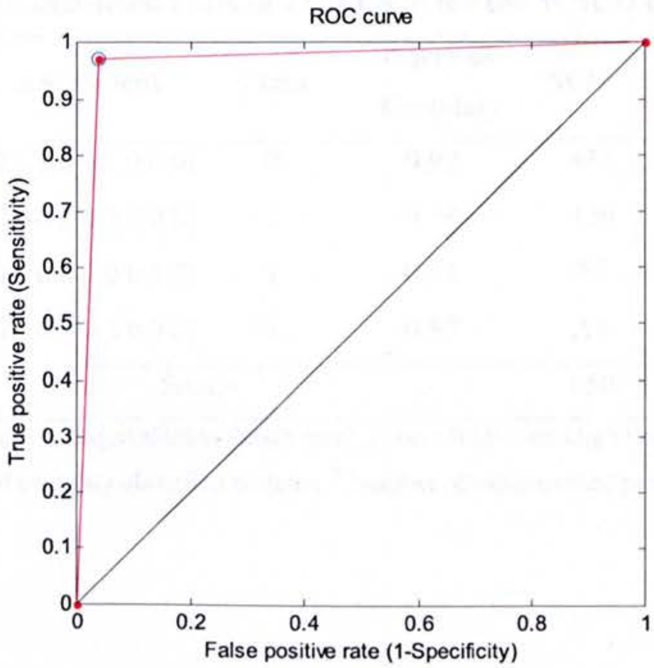


Figure 6.13: ROC curve of the IT2FLS_3 for the WBCD dataset

The fourth system, IT2FLS_4, generates rules with three term sets. The (k_m, k_v) parameters were selected (0.3, 0.7). The generated IT2 Gaussian MFs of IT2FLS_4 are shown in Figure 6.14. Four rules are generated by IT2FLS_4 (75% of the dataset for training and 25% for testing); three rules for classifying malignant objects and one rule for classifying benign objects as follows:

If BN is High then object is malignant with DC 0.79

If NN is High and M is Mid then object is malignant with DC 0.71

If BN is High and M is Mid then object is malignant with DC 0.82

If UC is Low and BN is Low then object is benign with DC 0.92

The classification accuracy of each rule is shown in Table 6.8. The average rule length is 1.75 in the antecedent part. This set of rules provides a classifier with an overall classification rate of 95.17% with an error rate on the training set of 3.91% and an error rate on the testing set of 8.77%. The average accuracy of this classifier using area under the ROC curve is 96.48 % with 95% [94.81 98.15] a shown in Figure 6.15.

Table 6.8 extracted rules of IT2FLS_4 for the WBCD dataset

No.	Antecedent	Class	Degree of Certainty	NCP[1]	NMP[2]
R1	[0 1 0 0 0 1 0 0 0]	0	0.92	413	2
R2	[0 0 0 0 0 3 0 0 0]	1	0.79	136	12
R3	[0 0 0 0 0 0 0 3 2]	1	0.71	82	18
R4	[0 0 0 0 0 3 0 0 2]	1	0.82	19	1
		Sum =		650	33

*0, 1, 2, and 3 are equivalent to "don't care", Low, Middle and High linguistic terms

[1] Number of correctly classified patterns, [2] Number of misclassified patterns

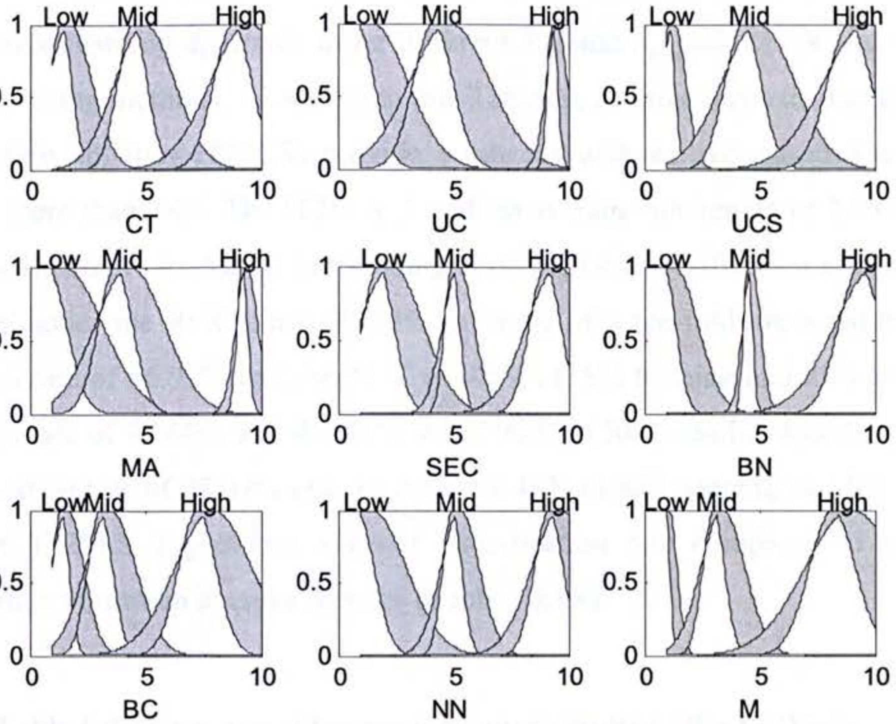


Figure 6.14: Membership function of IT2FLS_4, ($k_m = 0.3, k_v = 0.7$), for nine features of the WBCD

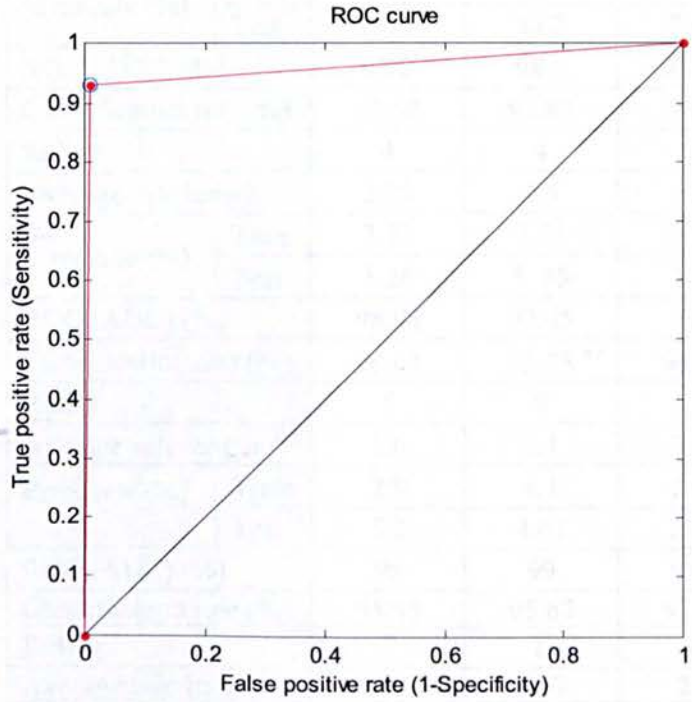


Figure 6.15: ROC curve of the IT2FLS_4 for the WBCD dataset

Table 6.9 summarizes the performance results of the obtained IT2FL rule sets of the GA IT2FL rule learning approach using different k_m and k_v parameters and different dataset partitioning methods. As shown in this Table, in average analysis using ten-fold cross-validation, all four IT2FLSs provide a rule set with an average area under the ROC curve more than 98%. The IT2FLS_3 with an average rule length of 2.88 provides a better classification rate for all partitioning systems; i.e., classification rate of 96.6% and the area under the ROC curve 98.73% (average of a ten-fold cross-validation), a classification rate of 96.93% and the AUC of 98.6% (75% training and 25% testing), a classification rate of 97.66% and the ROC AUC 98.8% (50% training and 50% testing) and classification rate of 97.07% and the AUC 98.49% (100% training and 0% testing). On average, IT2FLS_3 provides a better classification rate compared to the other IT2FLSs with a minimum average number of rules (2.88).

Table 6.9 Summary of the performance results of the IT2FLSs

			IT2FLS_1	IT2FLS_2	IT2FLS_3	IT2FLS_4
Average of a ten-fold cross-validation	Average rule #		3.28	3.25	2.88	3.3
	Average rule length		6.2	6.19	6.16	6.61
	Error rate (%)	Train	4.18	4.71	3.16	4.82
		Test	4.55	3.62	3.64	5.31
	ROC (AUC) (%)		98.2	98.8	98.73	98.1
	Classification rate (%)		95.64	95.83	96.6	94.93
75% training set and 25% testing set	Rule #		4	4	4	4
	Average rule length		2.25	1.5	2	1.75
	Error rate (%)	Train	2.73	3.71	2.93	3.91
		Test	5.26	5. 85	3.51	8.77
	ROC (AUC) (%)		98.04	97.45	98.6	96.48
	Classification rate (%)		96.63	95.75	96.93	95.17
50% training set and 50% testing set	Rule #		5	8	11	10
	Average rule length		1.6	2.1	2.64	3.2
	Error rate (%)	Train	2.9	4.1	2.64	3.81
		Test	5.2	4.67	2.05	5.26
	ROC (AUC) (%)		98	99	98.8	98
	Classification rate (%)		95.95	95.62	97.66	95.47
100% training set and 0% testing set	Rule #		7	6	8	8
	Average rule length		2.28	1.67	2.5	2.37
	Error rate (%)	Train	3.07	4.25	2.93	4.68
	ROC (AUC) (%)		97.95	98.10	98.49	96.79
	Classification rate (%)		96.93	95.75	97.07	95.31

6.7.4 Performance comparison with the other fuzzy rule extraction methods

In this section, the performance of the IT2FLSs rule sets generated by the GA IT2FL rule learning approach have been compared to rule sets generated using other fuzzy rule extraction approaches (reviewed in Section 6.6) and the results are summarized in Table 6.10.

Table 6.10 Comparison of the performance of the fuzzy rule extraction methods for the WBCD dataset

Method	Term Sets	Var per Rule	Rules	Classification Rate (%)	Initialization
VISIT (Chang and Lilly, 2004)	4	2	3	96.5	By expert
Fuzzy-GA (Pena-Reyes and Sipper, 1999)	2	2.8	5	97.8	Using the knowledge of existing rule sets in the literature
EMO (Ishibuchi and Yamamoto, 2003)	≤ 5	≤ 2	≥ 2	96.47	Heuristically by expert
SANFIS (Wang and Lee, 2002)	2	9	2	96.3	Automatically using the MCA clustering algorithm
DT based FC (Abonyi, Roubos and Szeifert, 2003)	3	1.5	2	96.5	Binary decision tree extracted from expert knowledge
IT2FL_1 (this work)	3	2.25	4	96.63	Automatically using the method explained in Sections 6.2 to 6.4
IT2FL_2 (this work)	3	1.5	4	95.75	Automatically using the method explained in Section s 6.2 to 6.4
IT2FL_3 (this work)	3	2	4	96.93	Automatically using the method explained in Sections 6.2 to 6.4
IT2FL_4 (this work)	3	1.75	4	95.17	Automatically using the method explained in Sections 6.2 to 6.4

IT2FLS_3 provides a better classification rate than the other fuzzy methods except the Fuzzy-GA (Pena-Reyes and Sipper, 1999). One of the main advantages of the GA

IT2FL rule learning approach is that it does not need a priori knowledge from experts or other proposed fuzzy systems for initializing the fuzzy system rules and membership functions. This approach automatically initializes the membership function parameters and rule set of an IT2FLS using the training dataset and the method explained in sections 6.2 to 6.4 of this chapter. Furthermore, it models uncertainty issues which cannot be managed using ordinary type-1 FLS. The main advantages of the proposed GA IT2FL rule learning approach in this study compared to other proposed approaches are as follows:

1. It models uncertainty in classification problems which lack expert knowledge and have an imperfect training dataset. The uncertainties in the training dataset are managed through the FOU of type-2 fuzzy sets in the antecedent part of the rules.
2. There is no need for expert knowledge and user interaction for initialization of the fuzzy logic parameters such as membership function parameters. The proposed type-2 fuzzy approach automatically generates membership functions and their FOU from training dataset.
3. The IT2FLS rule set with linguistic rules has higher interpretability for user interaction than the other non-fuzzy methods.
4. It defines the degree of certainty of extracted rules for the classifier decisions.
5. This approach maintains the trades-off between accuracy and interpretability by providing a comprehensive linguistic rule set with an average rule number about 3 and classification rate more than 95% in average.

6.8 Summary

This chapter has introduced the idea of uncertain rule-based pattern classification in vague environments with lack of expert knowledge. This approach has attempted to model uncertainty sources in the input of a classifier and the rule sets using an IT2FLS. The approach extracts IT2FLS rules from a training dataset and learns them for classification of future observations. The interval type-2 fuzzy GA evolutionary approach for rule extraction takes advantages of:

1. Type-2 fuzzy sets for modelling uncertainty sources in the training dataset and rule set
2. The genetic algorithm for searching in an optimized way in the solution space

To evaluate the performance of the proposed approach, it was applied to the anonymous WBCD dataset, which is one of the frequently applied dataset for performance analysis of the classifications. Furthermore, the noise and variation in this dataset, in addition to uncertainty sources in image processing techniques, makes it a suitable choice. For generating membership function, the FOU parameters (k_m and k_v) were selected using the cardinality measure. The IT2FLS with the minimum cardinality and maximum classification rate was considered for rule generation. Four different IT2FLS rule sets (IT2FLS_1 to 4) using different selected parameters (k_m and k_v) were investigated. Table 6.11 and 6.12 summarize the obtained rules for classifying malignant and benign objects for the WBCD dataset. As shown in these two tables, rules with higher (lower) degree of memberships are more involved for classifying malignant (benign) objects. The performance of the generated IT2FLSs was assessed using four different partitioning methods and was compared to the other fuzzy methods for rule extraction.

Table 6.11 Rules generated for classifying benign objects

IT2FLS_#	No	Antecedent	Class	Degree of certainty	NCP	NMP
1	R1	[0 0 1 0 0 1 0 0 0]	0	0.82	350	4
1	R3	[2 0 1 0 0 1 0 0 0]	0	1	77	2
2	R1	[0 0 1 0 0 0 0 0 0]	0	0.55	422	7
3	R1	[0 0 0 1 1 1 0 0 0]	0	0.58	247	8
3	R4	[1 1 0 0 0 0 0 0 0]	0	0.9	185	1
4	R1	[0 1 0 0 0 1 0 0 0]	0	0.92	413	2

*0, 1, 2, and 3 are equivalent to "don't care", Low, Middle and High linguistic terms

[1] Number of correctly classified patterns, [2] Number of misclassified patterns

Table 6.12 Rules generated for classifying malignant objects

IT2FLS_#	No	Antecedent	Class	Degree of certainty	NCP	NMP
1	R2	[0 0 0 3 0 3 0 0 0]	1	0.92	60	3
1	R4	[0 0 0 0 0 3 3 0 0]	1	0.8	173	14
2	R2	[0 0 0 3 0 3 0 0 0]	1	0.96	57	0
2	R3	[0 0 0 0 0 3 0 0 0]	1	0.65	150	22
2	R4	[3 0 0 0 0 0 3 0 0]	1	0.9	25	0
3	R2	[3 0 0 0 0 0 0 0 0]	1	0.6	120	11
3	R3	[0 0 0 0 0 3 3 0 0]	1	0.92	110	1
4	R2	[0 0 0 0 0 3 0 0 0]	1	0.79	136	12
4	R3	[0 0 0 0 0 0 0 3 2]	1	0.71	82	18
4	R4	[0 0 0 0 0 3 0 0 2]	1	0.82	19	1

*0, 1, 2, and 3 are equivalent to "don't care", Low, Middle and High linguistic terms

[1] Number of correctly classified patterns, [2] Number of misclassified patterns

The IT2FLS rule set competes with the best results of the other fuzzy approaches in terms of classification rates, number of rules, and number of variable per rules. Furthermore, the GA IT2FLS rule learning approach automatically initializes the membership function and rule sets of the FLS compared to most fuzzy approaches which need expert knowledge for the FLS initialization. The average accuracy of the best classifier (IT2FLS_3) with an average rule length of 2.88, after applying a ten-fold cross-validation was 96.6% with the area under the ROC curve 98.73%, which is comparable with the best of the previous fuzzy methods.

This approach is the first attempt toward modelling uncertainty in imperfect training dataset for pattern classification using membership functions and rules of an IT2FLS. The uncertainties are managed in membership functions (FOU) and rules of an IT2FLS. The presented method has the capability to be applied more widely to pattern classification problems that exhibit a lack of expert knowledge and with an imprecise and imperfect training dataset.

Chapter 7: Conclusion and Future Works

7.1 Conclusion

The purpose of the current study was to model uncertainty issues associated with a classification for the CAD applications. One of the main drawbacks of the current CAD applications is lack of consideration of uncertainty issues in design of the system components. Moreover, medical image analysis applications deal with the uncertainties inherent in intuitive knowledge of experts, inter- and intra- observer variability and ambiguity in the perception of the clinical words from the point of view of different experts. Chapter 2 has addressed and explained the uncertainty challenges using examples of real cases in CAD applications or medical diagnosis.

The study takes advantage of fuzzy logic as a rule-based classification method. However, membership functions of a type-1 fuzzy logic are more capable for managing uncertainty issues than deterministic methods, but they cannot manage all the uncertainties in the system. Remaining sources of uncertainty in a T1FLS have been explained in Chapter3. The study suggested that evolving a type-2 fuzzy logic for classification in a CAD application enables the system to manage high level of uncertainty. The uncertainty sources and capabilities of fuzzy logic rule-based classification and specifically type-2 fuzzy logic for managing the uncertainties have been demonstrated in detail in Part I of this study (Chapters 2 to 4).

Type-2 fuzzy logic has been introduced as a solution for modelling uncertainty in classification which cannot be modelled using a T1FLS. General architecture of a T2FLS, its components and fuzzy reasoning have been illustrated using a simplified nodule candidate classification example for a lung CAD system in Chapter 4. It should be noted that the main problem with type-2 fuzzy sets is the complexity of the type-2 fuzzy set mathematics because of three dimensional membership functions (refer to Section 4.3 Chapter 4) and difficulties in learning and understanding type-2 fuzzy logic concepts and theory compared with type-1 fuzzy logic. However, recent advances in simplifying type-2 fuzzy sets using interval type-2 fuzzy sets (applied in this study) significantly improves this problem and ameliorates the computational complexity of general type-2 fuzzy sets. In addition, because a decade has passed since the appearance and emergence of type-2 fuzzy set theory, the lack of software for the design and implementation of the T2FLS and performing operations on type-2 fuzzy sets still has an impact. Software developed in research groups only provides basic features of the theory. A T2FLS toolbox with the capability of learning from samples and adjustable to designer preferences would considerably speed up the design of the system.

The uncertainty issues associated with the input of a T2FLS classification (in the training dataset) have been managed using the FOU of IT2FMFs in an IT2FLS. The proposed IT2FLS approach for classification implicitly copes with accumulative effect of uncertainty issues (explained in Chapter 2) in the input of classification in the following ways:

- Uncertainty in image processing and analysis processes as well as uncertainty in mathematical models for measuring complex features of images are inherent in measurement of the features of the candidate objects in the input of classification. Uncertainty in the non-stationary features are represented in the training dataset as different input sets. Inter- and intra- uncertainty sources in the final decision of clinicians as well as vagueness in perception of the words are latent in different samples in the training dataset in the input of classifier.
- The accumulative effect of uncertainty sources in input dataset is managed in the FOU of the IT2FMFs through the proposed approach for modelling a Gaussian interval type-2 fuzzy membership function. Membership functions of the IT2FLSs are automatically designed using training dataset and the uncertainties associated with them are represented in overlap between the boundaries of the type-2 fuzzy membership functions as well as their FOUs. The FOU represents and models uncertainty in the interval type-2 fuzzy sets in antecedent and consequent part of the rules that are the union of several embedded T1FMFs. Furthermore, this method tunes the FOU of the IT2FMFs according to the uncertainties in the training dataset using the genetic algorithm with the proposed chromosome structure and the

improved initialization technique, which has fewer genes than other proposed GA methods and initializes the chromosome in a more efficient way.

- The ambiguity in the subjective knowledge of experts for initialization of the fuzzy logic parameters such as membership function parameters has been eliminated. The Gaussian IT2FMF generation approach automatically initializes membership function parameters and their FOU from training dataset.

On the other hand, comprehensibility is an essential factor in classification for a CAD application (as a second reader in the image analysis process). For this, the IT2FLS classifier with linguistic terms (maximum three) in the antecedent and consequent part of the rules has been developed which exhibits maximum interpretability between different fuzzy rule types.

7.2 novelty

This research, for the first time, introduces the concept of uncertainty in CAD application. It is also the first attempt toward modelling uncertainty issues in the design of classification component for a CAD application. The novelty of research is mainly represented by integrating a type-2 fuzzy logic for classification in a CAD application to manage uncertainties in input of classification.

Furthermore, the research introduces the idea of uncertain rule-based pattern classification for classification problems in vague environment which exhibit a lack of expert knowledge for initialization of the system parameters and rule set and only with an incomplete and uncertain training dataset as is the case for most of classification problems.

In addition, this is the first attempt at automatically generating membership function and rules for an interval type-2 fuzzy logic. This is significant from theoretical aspect of type-2 fuzzy logic and application of that for managing uncertainties in various applications.

7.3 Contribution

This study addresses and models accumulative effect of uncertainty sources in the design of classification component of the CAD application. The study suggested that integrating an IT2FL for in classification component of a CAD enables us to manage uncertainties latent in input of classifier. Uncertainty issues are modelled using FOU and rules of an IT2FL classifier. It also automatically extracts rules and membership functions for an IT2FLS from training datasets. These results are significant and contributes to medical diagnosis applications such as CAD technologies in at least two major aspects:

(1) Modelling uncertainties associated with the input of classification

(2) Extracting rules for classification problems with lack of expert knowledge and imperfect training dataset

The rest of this section explains these two aspects.

7.3.1 Modeling uncertainty associated with the input of classification

The approach for learning and tuning a Gaussian IT2FMF manages the uncertainties in the input of the classification (Chapter 5). The FOU of a Gaussian IT2FMF is estimated based on either an existing T1FLS or an imperfect training dataset. In the GA-based approach for learning the FOU of the IT2FMF, the structure of the chromosome has fewer genes than other GA-based methods and chromosome initialization is more precise. This property is significant for multi-dimensional classification problems with a large number of inputs where it is difficult to initialize IT2FMF parameters using the knowledge of experts. This method was applied to classify candidate nodules detected by a lung CAD application. Integrating the IT2FL classifier into the lung CAD application enables us to model the accumulative effect of uncertainties in input training sets such as inter- and intra uncertainties, word perception, and numerical measurements implicitly. In order to have an unbiased view of the classifier performance the ten-fold cross-validation technique was used. The maximum performance of the proposed classifier using the average performance of 10 runs of a ten-fold cross-validation is obtained for the IT2FLS based on the uncertain T1FLS defined by experts, with an average accuracy result of 95%. In general, the three IT2FLS methods described in Chapter 5, all outperform T1FLS counterpart (with an average accuracy of 65%). Analysis of the results reveals that IT2FLS is more capable of capturing the uncertainty in the model and achieving a better performance. For the nodule classification in a lung CAD application, the IT2FLS performance is 30% better than the T1FLS. Furthermore, the result of the proposed methods for automatic learning and tuning of the IT2FMFs based on the training dataset with an average accuracy 82%, is more promising for multidimensional classification problems that lack expert knowledge. These findings suggest that IT2FLS is capable of providing more efficient performance compared to T1FLS for classification of candidate nodules in a lung CAD application.

7.3.2 Extracting rules for uncertain classification problems

This study has presented the idea of uncertain rule-based pattern classification for classification problems with a lack of expert knowledge, or with time-consuming, complicated, or an expensive knowledge acquisition processes, and with an imperfect training dataset in Chapter 6. This approach is capable to model uncertainty associated with the input of a classifier through membership functions and rules of an IT2FLS. This is the first attempt toward automatically extracting rules and membership functions for an IT2FLS from training samples.

The FOU parameters (k_m and k_v) of the IT2FLS with the minimum cardinality and maximum classification rate were selected for rule generation. A genetic algorithm in combination with a punishment-reward scheme has been applied for learning rules and the degree of certainty. For this, whenever a rule classifies a pattern correctly the degree of certainty of the rule was increased otherwise it was decreased. In order to evaluate performance and investigate the generalization properties of the automatic generation of an IT2FLSs approach, it was applied to the anonymous WBCD dataset, a frequently used dataset for analysis of classification performance. The noise and variations in this dataset, in addition to uncertainties in image processing techniques for measuring features of an object, make it a suitable choice for this study. To evaluate the performance of the GA IT2FL rule learning approach, four different generated rule sets were assessed and compared to the rule set suggested by other fuzzy methods.

The rule set of the IT2FLS competes with the best results of the other fuzzy approaches in terms of classification rates and number of rules. Furthermore, the GA IT2FL rule learning approach automatically initializes the membership function and rule sets of the IT2FLS, while most fuzzy approaches use expert knowledge for initialization of the membership functions. The average accuracy of the best IT2FLS classifier with an average rule length 2.88, after applying a ten-fold cross-validation, was 96.6 % with the area under the ROC curve 98.73%, which is comparable with the best rule set of the previous fuzzy methods.

7.4 Benefits

Medical diagnosis is associated with human errors. For designing a rule-based IT2FLS for classification, knowledge about the membership function parameters and rules are required. Most previous methods based on IT2FLS assumed that this knowledge is available and can be extracted from a group of experts. Although prior studies employed the GA optimization properties for tuning IT2FMF parameters, but these methods require expert knowledge for initialization of the membership function parameters and their grammatical correctness in the chromosome structure. Moreover, the chromosome structure in these methods consists of all the parameters of the IT2FMFs. However, most medical diagnosis applications cope with a lack of expert knowledge, or high variations in their diagnosis associated with inter and intra observer variability and different perceptions about a clinical vocabulary using vague linguistic terms. These facts complicate the process of knowledge acquisition from a group of experts, especially when dealing with medical applications. Moreover, having access to a group of medical expert and fusing their knowledge is a difficult, time-consuming, and complicated process and sometimes impossible to achieve.

The approach presented in this study automatically designs membership functions and rules of an IT2FLS classifier from an imperfect training dataset and learns them for the prediction of new observations. The approach proposed for modelling uncertainties in the FOU of an IT2FMFs and rule set of an IT2FLS and its implementation details have been described in Part II (Chapters 5 and 6). The performance of the proposed approach has been evaluated by applying it to the nodule classification problem in a lung CAD application and to the anonymous breast cancer diagnosis classification problem using the WBCD dataset. The approach presented in this research for modelling uncertainty in classification component of a CAD application is significant from two major aspects:

(1) *Clinical view*: producing more accurate results for diagnosis problems can help to save more human lives. Modelling uncertainties in classification for a CAD provides a

powerful tool which improves the ambiguity of clinicians such as inter and intra observer variability for making final decision. Furthermore, the designed linguistic classifier using IT2FL has high interpretability, which can help radiologists as a second reader in the process of image analysis and interpretation. Furthermore, the automatic approach decreases the human errors in classification and provides a system with high true positive and low false positive rates.

(2) *Technical view*: modelling uncertainties in the design of a classifier using the automatic approach presented for IT2FLS membership and rules generation. This is critical for multi-dimensional classification problems with large number of inputs and lack of expert knowledge as is the case for most of medical diagnosis problems. This is significant for multi-dimensional classification problems with large number of inputs and lack of expert knowledge as is the case for most of medical diagnosis problems.

7.5 Future Works

Future investigations are suggested to improve the ability of the proposed approach for learning membership functions and rules of an IT2FLS from an imprecise and inadequate training dataset as follows:

- Design a multi-objective GA approach for maximizing the accuracy and degree of certainty and minimizing the number of rules and linguistic terms for an IT2FLS.
- Apply an adaptive genetic algorithm to define the probability of mutation and crossover automatically and based on the fitness of the population.

The GA selects the best candidate solution between various possible sets of solutions in each generation. The cross-over operation combines the best candidates and the mutation operation creates diversity in different generations to go through all possible combinations. While the advantage of the GA evolutionary approach to search for the global optimum in solution space has been considered for optimization problems using T2FLSs, one of the drawbacks of GA-based methods is its dependency on the initial population. More research is needed to improve initialization of the GA IT2FL rule-learning algorithm more intelligently. In addition, further investigations are suggested to employ other learning techniques like simulated annealing which promise to get to the global optimum and compared the results to the GA-based method.

Future studies on the current topic needs to be undertaken to improve the degree of certainty of the IT2FL classification rules by adjusting the FOU of the membership functions and decreasing the amount of the uncertainty in the system. An ideal IT2FLS classifier is a system which makes accurate decisions with the minimum uncertainty. Different uncertainty measures might be useful to demonstrate and quantify the amount of uncertainty in the system.

Apart from learning and certainty improvements, applying a feature extraction technique is suggested to improve the overall IT2FLS interpretability and ameliorate the computing complexity of the algorithm. Applying feature extraction techniques can dramatically reduce the dimensionality of a classification system with a large number of inputs. It is also recommended to apply the approach to the other pattern classification problems with lack of expert knowledge and imperfect dataset to evaluate the general performance of the approach and recognize other potential applications.

Future studies are suggested to model uncertainty sources in other components of the CAD application such as segmentation (delineating the boundaries of the objects), feature extraction (measuring complex feature of an object) taking advantages of type-2 fuzzy sets. Further studies are suggested to investigate robustness of the presented approach using different system variables such as datasets, uncertainty metrics and different features.

The research described in this thesis presents an approach for modelling uncertainties associated with classification. The uncertainties are modelled using membership functions and a rule set of a type-2 fuzzy logic system. The performance of the fuzzy classifier was evaluated on two different classification problems (nodule classification in a lung CAD and the Wisconsin breast cancer diagnosis problem) and the results demonstrate that the primary objectives of this study have been successfully achieved. The results reveal the superiority of the T2FLS classifier for managing high levels of uncertainty compared to the T1FLS counterpart.

Bibliography

A. De Luca, S. Termini. (1972) 'A definition of nonprobabilistic entropy in the setting of fuzzy sets theory', *Information and Computation 20,* pp. 301-312.

Abbass, H. A. (2002) 'An evolutionary artificial neural networks approach for breast cancer diagnosis', *Artificial Intelligence in Medicine,* 25 (3), pp. 265-281.

Abe, S. (2001) 'Pattern classification; Neuro-fuzzy methods and their comparison(Book)', *London, United Kingdom: Springer Verlag London, Ltd, 2001* .

Abonyi, J., Roubos, J. A. and Szeifert, F. (2003) 'Data-driven generation of compact, accurate, and linguistically sound fuzzy classifiers based on a decision-tree initialization', *International Journal of Approximate Reasoning,* 32 (1), pp. 1-21.

Akay, M. F. (2009) 'Support vector machines combined with feature selection for breast cancer diagnosis', *Expert Systems with Applications,* 36 (2), pp. 3240-3247.

Alcalá, R., Nojima, Y., Herrera, F. and Ishibuchi, H. (2009) *Generating single granularity-based fuzzy classification rules for multiobjective genetic fuzzy rule selection..* IEEE International Conference on Fuzzy Systems, pp. 1718-1723.

Arakawa, K., Kerre, E. and Nachtegael, M. (2000) 'Fuzzy rule-based image processing with optimization,"', *Fuzzy Techniques in Image Processing, vol.* 52, pp. 222-247.

Armato, S. G., Li, F., Giger, M. L., MacMahon, H., Sone, S. and Doi, K. (2002) 'Lung Cancer: Performance of Automated Lung Nodule Detection Applied to Cancers Missed in a CT Screening Program1', *Radiology,* 225 (3), pp. 685.

Baker, J. A., Rosen, E. L., Lo, J. Y., Gimenez, E. I., Walsh, R. and Soo, M. S. (2003) 'Computer-aided detection (CAD) in screening mammography: sensitivity of commercial CAD systems for detecting architectural distortion', *American Journal of Roentgenology,* 181 (4), pp. 1083.

Bezdek, J. C. and Pal, S. K. (1992) *Fuzzy models for pattern recognition: methods that search for structures in data,* Institute of Electrical and Electronics Engineers, IEEE Neural networks society.

Bishop, C. M. (1995) *Neural networks for pattern recognition.* Oxford university press.

Box, G. E. P. and Tiao, G. C. (1992) *Bayesian inference in statistical analysis.* Wisconsin University Madison, Department of Statistics;: Wiley Online Library.

Bradley, A. P. (1997) 'The use of the area under the ROC curve in the evaluation of machine learning algorithms', *Pattern Recognition,* 30 (7), pp. 1145-1159.

Cancer Research UK (2011) *UK breast cancer incidence statistics. Available at http://info.cancerresearchuk.org/cancerstats/types/breast/incidence.* (Accessed: 3 November 2011).

Casillas, J. (2003) *Accuracy improvements in linguistic fuzzy modeling.* Springer Verlag.

Casillas, J., Cordón, O., del Jesus, M. J. and Herrera, F. (2005) 'Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction', *IEEE Transactions on Fuzzy Systems,* Volume 13 (1), pp. 13-29.

Castillo, O., Martínez-Marroquín, R., Melin, P., Valdez, F. and Soria, J. (2010) 'Comparative study of bio-inspired algorithms applied to the optimization of type-1 and type-2 fuzzy controllers for an autonomous mobile robot', Information Sciences, Volume 192, June 2012, pp. 19-38.

Castillo, O. and Melin, P. (2008) *Type-2 fuzzy logic: theory and applications.* Springer Verlag.

Castro, J., Castillo, O., Melin, P., Martínez, L., Escobar, S. and Camacho, I. (2007) 'Building fuzzy inference systems with the interval type-2 fuzzy logic toolbox', *Analysis and Design of Intelligent Systems using Soft Computing Techniques,* pp. 53-62.

Castro, J., Castillo, O., Melin, P. and Rodríguez-Díaz, A. (2008) 'Building fuzzy inference systems with a new interval type-2 fuzzy logic toolbox', *Transactions on Computational Science I,* pp. 104-114.

Castro, J. R., Castillo, O. and Melin, P. (2007) *An interval type-2 fuzzy logic toolbox for control applications.*, in Proc. IEEE International Fuzzy Systems Conference, pp. 1–6.

Cazarez-Castro, N. R., Aguilar, L. T. and Castillo, O. (2010) 'Fuzzy logic control with genetic membership function parameters optimization for the output regulation of a servomechanism with nonlinear backlash', *Expert Systems with Applications,* 37 (6), pp. 4368-4378.

Chang, X. and Lilly, J. H. (2004) 'Evolutionary design of a fuzzy classifier from data', *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* Volume 34 (4), pp. 1894-1906.

Chi, Z., Yan, H. and Phạm, T. (1996) *Fuzzy algorithms: with applications to image processing and pattern recognition.* World Scientific Pub Co Inc.

Choi, B. I. and Chung-Hoon Rhee, F. (2009) 'Interval type-2 fuzzy membership function generation methods for pattern recognition', *Information Sciences,* 179 (13), pp. 2102-2122. Corcoran, A. L. and Sen, S. (1994), *Using real-valued genetic algorithms to evolve rule sets for classification.* In Proc. the First IEEE Conference on Computational Intelligence., Volume 1, pp. 120-124.

Cordón, O., del Jesus, M. J. and Herrera, F. (1999) 'A proposal on reasoning methods in fuzzy rule-based classification systems1', *International Journal of Approximate Reasoning,* 20 (1), pp. 21-45.

da Silva, E. C., Silva, A. C., de Paiva, A. C. and Nunes, R. A. (2008) 'Diagnosis of lung nodule using Moran's index and Geary's coefficient in computerized tomography images', *Pattern Analysis & Applications,* 11 (1), pp. 89-99.

Dehmeshki, J., Amin, H., Valdivieso, M. and Ye, X. (2008) 'Segmentation of pulmonary nodules in thoracic CT scans: a region growing approach', *IEEE Transactions on Medical Imaging,* 27 (4), pp. 467-480.

Dehmeshki, J., Valdivieso-Casique, M., Abaei, M., Kamangari N., Ebadian Dehkordi M., Roddie M.E. and Costello. J. (2003) *Computer Assisted Detection of Pulmonary Nodules on Thoracic CT Scans Using Image Processing and Classification Techniques, Lung cancer: new diagnostic approaches, Thorax, S122.*

Dehmeshki, J., Ye, X., Casique, M. and Lin, X. (2006) *A hybrid approach for automated detection of lung nodules in CT images.* IEEE Int. Symp. Biomedical Imaging: From Nano to Macro, pp. 506-509.

Dehmeshki, J., Ye, X., Lin, X. Y., Valdivieso, M. and Amin, H. (2007) 'Automated detection of lung nodules in CT images using shape-based genetic algorithm', *Computerized Medical Imaging and Graphics,* 31 (6), pp. 408-417.

Dhawan, A. P. (2011) *Medical image analysis*. Wiley-IEEE Press.

Di Lascio, L., Gisolfi, A. and Nappi, A. (2005) *Medical differential diagnosis through type-2 fuzzy sets*, IEEE International Conference on Fuzzy Systems, pp. 371-376.

Dougherty, E. R. and Lotufo, R. A. (2003) *Hands-on morphological image processing*. Society of Photo Optical, Volume 59.

Ensafi, P. and Tizhoosh, H. (2005) 'Type-2 fuzzy image enhancement', *Image Analysis and Recognition*, pp. 159-166.

Fawcett, T. (2006) 'An introduction to ROC analysis', *Pattern Recognition Letters*, 27 (8), pp. 861-874.

Ferlay, J., Shin, H., Bray, F., Forman, D., Mathers, C. and Parkin, D. (2011) *GLOBOCAN 2008 v1. 2, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10.*, Available at: http://globocan.iarc.fr (Accessed: 03/05/2011).

Fogel, D. B. and Wasson, E. C. (1995) 'Evolving neural networks for detecting breast cancer', *Cancer Letters*, 96 (1), pp. 49-53.

Frank, A. and Asuncion, A. (2010) 'UCI Machine Learning Repository [http://archive.ics.uci.edu/ml/], Irvine, University of California', *School of Information and Computer Science*.

Garibaldi, J. M. and Ozen, T. (2007) 'Uncertain fuzzy reasoning: a case study in modelling expert decision making', *IEEE Transactions on Fuzzy Systems*, Volume 15 (1), pp. 16-30.

Gomathi, M. and Thangaraj, P. (June 2010) 'Lung Nodule Detection using a Neural Classifier', *IACSIT International Journal of Engineering and Technology*, 2 pp. 291-295.

Gonzalez, R. C., Woods, R. E. and Eddins, S. L. (2004) *Digital image processing using MATLAB*. Pearson Education India.

Gould, M. K., Fletcher, J., Iannettoni, M. D., Lynch, W. R., Midthun, D. E., Naidich, D. P. and Ost, D. E. (2007) 'Evaluation of Patients With Pulmonary Nodules: When Is It Lung Cancer?', *Chest,* 132 pp. 108S-130S.

Hájek, P. (1998) *Metamathematics of fuzzy logic.* Springer.

Halligan, S., Altman, D. G., Mallett, S., Taylor, S. A., Burling, D., Roddie, M., Honeyfield, L., McQuillan, J., Amin, H. and Dehmeshki, J. (2006) 'Computed tomographic colonography: assessment of radiologist performance with and without computer-aided detection', *Gastroenterology,* 131 (6), pp. 1690-1699.

Hein, P. A., Krug, L. D., Romano, V. C., Kandel, S., Hamm, B. and Rogalla, P. (2010) 'Computer-aided detection in computed tomography colonography with full fecal tagging: comparison of standalone performance of 3 automated polyp detection systems', *Canadian Association of Radiologists Journal,* 61 (2), pp. 102-108.

Herrera, F. (2008) 'Genetic fuzzy systems: taxonomy, current research trends and prospects', *Evolutionary Intelligence,* 1 (1), pp. 27-46.

Hidalgo, D., Castillo, O. and Melin, P. (2008) 'Interval type-2 fuzzy inference systems as integration methods in modular neural networks for multimodal biometry and its optimisation with genetic algorithms', *International Journal of Biometrics,* 1 (1), pp. 114-128.

Hidalgo, D., Melin, P. and Castillo, O. (2010) *Optimal design of type-2 fuzzy membership functions using genetic algorithms in a partitioned search space.* , IEEE International Conference on Granular Computing, pp. 212-216.

Hosseini, R., Dehmeshki, J., Barman, S., Mazinani, M. and Qanadli, S. (2010a) *A Genetic type-2 fuzzy logic system for pattern recognition in computer aided detection systems.* IEEE International Conference on Fuzzy Systems, pp. 1-7.

Hosseini, R., Dehmeshki, J., Barman, S., Mazinani, M. and Qanadli, S. (2010b) *Modeling uncertainty in classification design of a computer-aided detection system,* SPIE, Computer Aided Detection, Volume 7624, pp. 76242V.

Hosseini, R., Qanadli, S., Barman, S., Mazinani, M., Ellis, T. and Dehmeshki, J. (2012) 'An automatic approach for learning and tuning Gaussian interval type-2 fuzzy membership functions applied to lung CAD classification system', *IEEE Transactions on Fuzzy Systems,* 20 pp. 224-234.

Innocent, P. R. and John, R. I. (2004) 'Computer aided fuzzy medical diagnosis', *Information Sciences,* 162 (2), pp. 81-104.

Innocent, P., John, R., Belton, I. and Finlay, D. (2001) *Type 2 fuzzy representations of lung scans to predict pulmonary emboli,* IFSA World Congress and 20th NAFIPS International Conference, Volume 4, pp. 1902-1907.

Ishibuchi, H., Kaisho, Y. and Nojima, Y. (2009) *Complexity, interpretability and explanation capability of fuzzy rule-based classifiers.* IEEE International Conference on Fuzzy Systems, pp. 1730-1735.

Ishibuchi, H., Nakashima, T. and Murata, T. (2001) 'Three-objective genetics-based machine learning for linguistic rule extraction', *Information Sciences,* 136 (1), pp. 109-133.

Ishibuchi, H., Nakashima, T. and Murata, T. (1996) 'A fuzzy classifier system that generates linguistic rules for pattern classification problems', *Fuzzy Logic, Neural Networks, and Evolutionary Computation,* pp. 35-54.

Ishibuchi, H. and Nojima, Y. (2008) 'Pattern Classification with Linguistic Rules', *Fuzzy Sets and their Extensions: Representation, Aggregation and Models,* pp. 377-395.

Ishibuchi, H. and Nojima, Y. (2007) 'Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning', *International Journal of Approximate Reasoning,* 44 (1), pp. 4-31.

Ishibuchi, H. and Yamamoto, T. (2005) 'Rule weight specification in fuzzy rule-based classification systems', *IEEE Transactions on Fuzzy Systems,* 13 (4), pp. 428-435.

Ishibuchi, H. and Yamamoto, T. (2003) *Effects of three-objective genetic rule selection on the generalization ability of fuzzy rule-based systems,*Evolutionary Multi-Criterion Optimization, Lecture Notes in Computer Science, Volume 2632, Springer, pp. 69.

Jarque, C. M.and Bera, A. K. (1987). 'A test for normality of observations and regression residuals', *International Statistical Review,* Volume **55** (2), pp.163–172

Jansen, M. (2001) *Noise reduction by wavelet thresholding.* Springer USA.

Jiang, L., Wang, D., Cai, Z. and Yan, X. (2007) 'Survey of improving naive Bayes for classification', *Advanced Data Mining and Applications,* pp. 134-145.

John, R. and Coupland, S. (2006) 'Extensions to type-1 fuzzy logic: Type-2 fuzzy logic and uncertainty', *Computational Intelligence: Principles and Practice,* pp. 89-101.

John, R. and Lake, S. (2001) *Type-2 fuzzy sets for modelling nursing intuition.* IFSA World Congress and 20th NAFIPS International Conference, Volume 4, pp. 1920-1925,

John, R. I. and Innocent, P. R. (2005) 'Modeling uncertainty in clinical diagnosis using fuzzy logic', *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* 35 (6), pp. 1340-1350.

John, R., Innocent, P. R. and Barnes, M. (2000) 'Neuro-fuzzy clustering of radiographic tibia image data using type 2 fuzzy sets', *Information Sciences,* 125 (1-4), pp. 65-82.

Kakeda, S., Moriya, J., Sato, H., Aoki, T., Watanabe, H., Nakata, H., Oda, N., Katsuragawa, S., Yamamoto, K. and Doi, K. (2004) 'Improved detection of lung nodules on chest radiographs using a commercial computer-aided diagnosis system', *American Journal of Roentgenology,* Volume 182 (2), pp. 505.

Kandel, A. and Friedman, A. M. (1999) *Introduction to pattern recognition: statistical, structural, neural, and fuzzy logic approaches.* World Scientific Pub Co Inc, Volume 32.

Karnik, N. N. and Mendel, J. M. (2001) 'Centroid of a type-2 fuzzy set', *Information Sciences,* 132 (1), pp. 195-220.

Karnik, N. N., Mendel, J. M. and Liang, Q. (1999) 'Type-2 fuzzy logic systems', *IEEE Transactions on Fuzzy Systems,* Volume 7 (6), pp. 643-658.

Karnik, N. and Mendel, J. (1998) *An introduction to type-2 fuzzy logic systems.*

Kerre, E. E. and Nachtegael, M. (2000) *Fuzzy techniques in image processing.* Springer Berlin Heidelberg.

Keserci, B. and Yoshida, H. (2002) 'Computerized detection of pulmonary nodules in chest radiographs based on morphological features and wavelet snake model', *Medical Image Analysis,* 6 (4), pp. 431-447.

Koza, J. and Poli, R. (2005) 'Genetic programming', *Search Methodologies,* pp. 127-164.

Lee, C., Hong, J. L., Lin, Y. C. and LAI, W. (2003) 'Type-2 fuzzy neural network systems and learning', *International Journal of Computational Cognition (Http://www.YangSky.com/yangijcc.Htm),* 1 (4), pp. 79-90.

Lee, C. S., Wang, M. H. and Hagras, H. (2010) 'A type-2 fuzzy ontology and its application to personal diabetic-diet recommendation', *IEEE Transactions on Fuzzy Systems,* 18 (2), pp. 374-395.

Lee, Y., Hara, T., Fujita, H., Itoh, S. and Ishigaki, T. (2001) 'Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique', *IEEE Transactions on Medical Imaging,* Volume 20 (7), pp. 595-604.

Leon, S., Brateman, L., Honeyman-Buck, J. and Marshall, J. (2009) 'Comparison of two commercial CAD systems for digital mammography', *Journal of Digital Imaging,* 22 (4), pp. 421-423.

MacMahon, H., Austin, J. H. M., Gamsu, G., Herold, C. J., Jett, J. R., Naidich, D. P., Patz, E. F. and Swensen, S. J. (2005) 'Guidelines for Management of Small Pulmonary Nodules Detected on CT Scans: A Statement from the Fleischner Society1', *Radiology,* 237 (2), pp. 395.

Mamdani, E. H. (1977) 'Application of fuzzy logic to approximate reasoning using linguistic synthesis', *IEEE Transactions on Computers,* 100 (12), pp. 1182-1191.

Martens, D., Baesens, B., Van Gestel, T. and Vanthienen, J. (2007) 'Comprehensible credit scoring models using rule extraction from support vector machines', *European Journal of Operational Research,* 183 (3), pp. 1466-1476.

Martinez, A., Castillo, O. and Garcia, M. (2008) 'Comparative Study of Type-1 and Type-2 Fuzzy Systems Optimized by Hierarchical Genetic Algorithms', *Soft Computing for Hybrid Intelligent Systems,* pp. 53-70.

Martínez, R., Castillo, O. and Aguilar, L. T. (2009) 'Optimization of interval type-2 fuzzy logic controllers for a perturbed autonomous wheeled mobile robot using genetic algorithms', *Information Sciences,* 179 (13), pp. 2158-2174.

Martínez-Marroquín, R., Castillo, O. and Soria, J. (2009) 'Particle Swarm Optimization Applied to the Design of Type-1 and Type-2 Fuzzy Controllers for an Autonomous Mobile Robot', *Bio-Inspired Hybrid Intelligent Systems for Image Analysis and Pattern Recognition,* pp. 247-262.

Mendel, J. M. (2007a) 'Advances in type-2 fuzzy sets and systems', *Information Sciences,* 177 (1), pp. 84-110.

Mendel, J. M. (2007b) 'Computing with words and its relationships with fuzzistics', *Information Sciences,* 177 (4), pp. 988-1006.

Mendel, J. M. (2004) 'Computing derivatives in interval type-2 fuzzy logic systems', *IEEE Transactions on Fuzzy Systems, Volume* 12 (1), pp. 84-98.

Mendel, J. M. (2001) *Uncertain rule-based fuzzy logic systems: introduction and new directions.* Prentice-Hall.

Mendel, J. M., John, R. I. and Liu, F. (2006) 'Interval type-2 fuzzy logic systems made simple', *IEEE Transactions on Fuzzy Systems,* Volume 14 (6), pp. 808-821.

Mendel, J. M. and John, R. I. B. (2002) 'Type-2 fuzzy sets made simple', *IEEE Transactions on Fuzzy Systems,* Volume 10 (2), pp. 117-127.

Mendel, J. M. and Liu, F. (2007) 'Super-exponential convergence of the Karnik–Mendel algorithms for computing the centroid of an interval type-2 fuzzy set', *IEEE Transactions on Fuzzy Systems,* Volume 15 (2), pp. 309-320.

Mendez, G. M., Cavazos, A., Soto, R. and Leduc, L. (2006) 'Entry temperature prediction of a hot strip mill by a hybrid learning type-2 FLS', *Journal of Intelligent and Fuzzy Systems,* 17 (6), pp. 583-596.

Meyer-Baese, A. (2004) *Pattern recognition for medical imaging.* Academic Press.

Michalewicz, Z. (1996) *Genetic algorithms data structures.* Springer.

Mitchell, H. B. (2005) 'Pattern recognition using type-II fuzzy sets', *Information Sciences,* 170 (2-4), pp. 409-418.

Montgomery, D. C., Runger, G. C. and Hubele, N. F. (2009) *Engineering statistics.* Fourth edn. Wiley.

N. Karnik, N. and M. Mendel, J. (2001) 'Operations on type-2 fuzzy sets', *Fuzzy Sets and Systems,* 122 (2), pp. 327-348.

Niewiadomski, A. (2008) 'A type-2 fuzzy approach to linguistic summarization of data', *IEEE Transactions on Fuzzy Systems,* Volume 16 (1), pp. 198-212.

Niewiadomski, A. (2007) 'Type-2 fuzzy summarization of data: An improved news generating', *Rough Sets and Intelligent Systems Paradigms,* pp. 241-250.

Niewiadomski, A. (2005) 'On two possible roles of type-2 fuzzy sets in linguistic summaries', *Advances in Web Intelligence,* pp. 945-946.

Niewiadomski, A. and Bartyzel, M. (2006) 'Elements of the Type-2 Semantics in Summarizing Databases', *Artificial Intelligence and Soft Computing–ICAISC 2006,* pp. 278-287.

Niewiadomski, A. and Szczepaniak, P. S. (2006) *News generating based on interval type-2 linguistic summaries of database,* Proceedings of IPMU Conference, pp. 2-7.

Oversight, T. (2011) 'Reduced lung-cancer mortality with low-dose computed tomographic screening', *N Engl J Med,* 2011 (365), pp. 395-409.

Ozen, T. and Garibaldi, J. M. (2003) *Investigating adaptation in type-2 fuzzy logic systems applied to umbilical acid-base assessment.* Citeseer, pp. 289–294.

Pal, S. K. and Mitra, S. (1999) *Neuro-fuzzy pattern recognition: methods in soft computing.* John Wiley & Sons, Inc.

Pena-Reyes, C. A. and Sipper, M. (1999) 'A fuzzy-genetic approach to breast cancer diagnosis', *Artificial Intelligence in Medicine,* 17 (2), pp. 131-155.

Peters, R. A. (1995) 'A new algorithm for image noise reduction using mathematical morphology', *IEEE Transactions on Image Processing,* Volume 4 (5), pp. 554-568.

Pham, D. L., Xu, C. and Prince, J. L. (2000) 'Current Methods in Medical Image Segmentation 1', *Annual Review of Biomedical Engineering,* 2 (1), pp. 315-337.

Pham, D. and Castellani, M. (2002) 'Action aggregation and defuzzification in Mamdani-type fuzzy systems', *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science,* 216 (7), pp. 747.

Picard, R. R. and Cook, R. D. (1984) 'Cross-validation of regression models', *Journal of the American Statistical Association,* pp. 575-583.

Priestley, M. B. (1988) 'Non-linear and non-stationary time series analysis', .

Roos, J. E., Paik, D., Olsen, D., Liu, E. G., Chow, L. C., Leung, A. N., Mindelzun, R., Choudhury, K. R., Naidich, D. P. and Napel, S. (2010) 'Computer-aided detection (CAD) of lung nodules in CT scans: radiologist performance and reading time with incremental CAD assistance', *European Radiology,* 20 (3), pp. 549-557.

Ross, T. J., Booker, J. M. and Parkinson, W. J. (2002) *Fuzzy logic and probability applications: bridging the gap.* Society for Industrial Mathematics.

Sampat, M. P., Markey, M. K. and Bovik, A. C. (2005) 'Computer-aided detection and diagnosis in mammography', *Handbook of Image and Video Processing,* 10 (4), pp. 1195-1217.

Sepulveda, R., Castillo, O., Melin, P., Montiel, O. and Aguilar, L. T. (2007) *Evolutionary optimization of interval type-2 membership functions using the human evolutionary model.* IEEE International Fuzzy System Conference, pp. 1-6.

Sethian, J. A. (2003) 'Level set methods and fast marching methods', *Journal of Computing and Information Technology,* 11 (1), pp. 1–2.

Setiono, R. (2000) 'Generating concise and accurate classification rules for breast cancer diagnosis', *Artificial Intelligence in Medicine,* 18 (3), pp. 205-219.

Setiono, R. (1996) 'Extracting rules from pruned neural networks for breast cancer diagnosis', *Artificial Intelligence in Medicine,* 8 (1), pp. 37-51.

Sezgin, M. and Sankur, B. (2004) 'Survey over image thresholding techniques and quantitative performance evaluation', *Journal of Electronic Imaging,* 13 pp. 146.

Sonka, M., Hlavac, V. and Boyle, R. (1999) 'Image processing, analysis, and machine vision second edition', *International Thomson*.

Spackman, K. A. (1989) *Signal detection theory: Valuable tools for evaluating inductive learning.* Morgan Kaufmann Publishers Inc.

Szmidt, E. and Kacprzyk, J. (2001) 'Entropy for intuitionistic fuzzy sets', *Fuzzy Sets and Systems,* 118 (3), pp. 467-477.

Taha, I. and Ghosh, J. (1996) *Characterization of the Wisconsin breast cancer database using a hybrid symbolic-connectionist system,* Proceedings of the Intelligent Engineering System.

Takagi, T. and Sugeno, M. (1985) 'Fuzzy identification of system and its applications to modelling and control', *IEEE Trans.Syst., Man, and Cyber,* Volume1, pp. 5.

Tizhoosh, H. (2008) 'Type II Fuzzy Image Segmentation', *Fuzzy Sets and their Extensions: Representation, Aggregation and Models,* pp. 607-619.

Tizhoosh, H. R. (2005) 'Image thresholding using type II fuzzy sets', *Pattern Recognition,* 38 (12), pp. 2363-2372.

van Ginneken, B., Armato III, S. G., de Hoop, B., van Amelsvoort-van de Vorst, S., Duindam, T., Niemeijer, M., Murphy, K., Schilham, A., Retico, A. and Fantacci, M. E. (2010) 'Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: The ANODE09 study', *Medical Image Analysis,* 14 (6), pp. 707-722.

Wagner, C. and Hagras, H. (2007) 'Evolving Type-2 Fuzzy Logic Controllers for Autonomous Mobile Robots', *Analysis and Design of Intelligent Systems using Soft Computing Techniques,* pp. 16-25.

Wang, C. H., Cheng, C. S. and Lee, T. T. (2004) 'Dynamical optimal training for interval type-2 fuzzy neural network (T2FNN)', *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* Volume 34 (3), pp. 1462-1477.

Wang, J. S. and Lee, C. S. G. (2002) 'Self-adaptive neuro-fuzzy inference systems for classification applications', *IEEE Transactions on Fuzzy Systems,* Volume 10 (6), pp. 790-802.

White, C., Pugatch, R., Teague, S., Dharaiya, E. and Read, K. (2007) 'Performance of Lung Nodule Computer Aided Detection Software: Effect of Slice Thickness on Chest CT', *Medicamundi,* 51 (2), pp. 3.

Wills, K., John, R., Innocent, P. and Lake, S. (2003) 'Modelling nursing intuition-a non-deterministic approach', *EUSFLAT2003.*

Witten, I. H. and Frank, E. (2005) *Data Mining: Practical machine learning tools and techniques.* Morgan Kaufmann.

Wolberg W. H., Mangasarian O. L, (1990) 'Multisurface method of pattern separation for edical diagnosis applied to breast cytology'. In Proceedings of the National Academy of sciences, U.S.A., vol. 87, pp. 9193–9196.

Wu, D. and Mendel, J. M. (2009) 'A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets', *Information Sciences,* 179 (8), pp. 1169-1192.

Wu, D. and Mendel, J. M. (2007) 'Uncertainty measures for interval type-2 fuzzy sets', *Information Sciences,* 177 (23), pp. 5378-5393.

Wu, D., Mendel, J. M. and Joo, J. (2010) *Linguistic summarization using IF-THEN rules.* IEEE International Conference on Fuzzy Systems, pp. 1-8.

Wu, D. and Wan Tan, W. (2006) 'Genetic learning and performance evaluation of interval type-2 fuzzy logic controllers', *Engineering Applications of Artificial Intelligence,* 19 (8), pp. 829-841.

Wu, H. and Mendel, J. M. (2002) 'Uncertainty bounds and their use in the design of interval type-2 fuzzy logic systems', *IEEE Transactions on Fuzzy Systems,* Volume 10 (5), pp. 622-639.

Wu, Z., Jiang, J. and Peng, Y. (2009) 'Computational Intelligence on Medical Imaging with Artificial Neural Networks', *Computational Intelligence in Medical Imaging: Techniques and Applications,* Chapter 1, pp. 1-22.

Ye, X., Lin, X., Dehmeshki, J., Slabaugh, G. and Beddoe, G. (2009) 'Shape-based computer-aided detection of lung nodules in thoracic CT images', *IEEE Transactions on Biomedical Engineering,* Volume 56 (7), pp. 1810-1820.

Yoshida, H. and Dachman, A. H. (2004) *CAD for CT colonography: current status and future.,* Volume 1268, Elsevier, pp. 973-977.

Zadeh, L. (2000) 'From computing with numbers to computing with words—from manipulation of measurements to manipulation of perceptions', *Intelligent Systems and Soft Computing,* pp. 3-40.

Zadeh, L. A. (2008) 'Is there a need for fuzzy logic?', *Information Sciences,* 178 (13), pp. 2751-2779.

Zadeh, L. A. (1994) 'Fuzzy logic, neural networks, and soft computing', *Communications of the ACM,* Volume 37 (3), pp. 77-84.

Zadeh, L. A. (1975) 'The concept of a linguistic variable and its application to approximate reasoning--I* 1', *Information Sciences,* Volume 8 (3), pp. 199-249.

Zadeh, L. A. (1965) 'Fuzzy sets*', *Information and Control,* Volume 8 (3), pp. 338-353.

Zeng, J. and Liu, Z. Q. (2007) 'Type-2 fuzzy sets for pattern recognition: The state-of-the-art', *Journal of Uncertain Systems,* 1 (3), pp. 163-177.

Zeng, J., Xie, L. and Liu, Z. Q. (2008) 'Type-2 fuzzy Gaussian mixture models', *Pattern Recognition,* 41 (12), pp. 3636-3643.

Appendix A

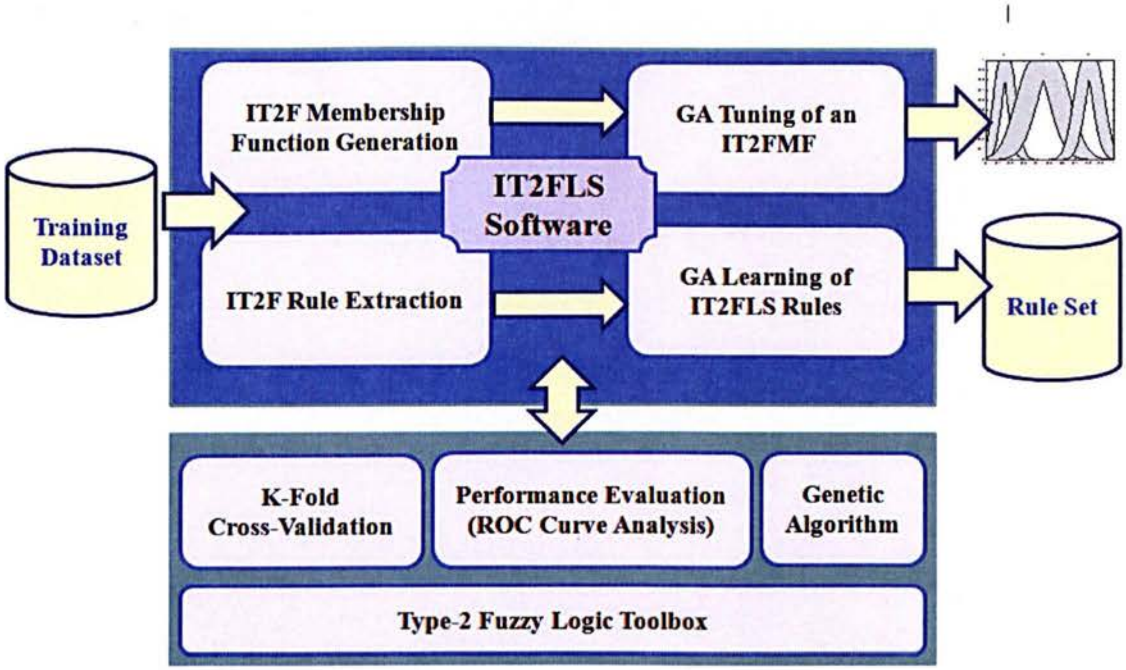


Figure A.1: IT2FLS software architecture

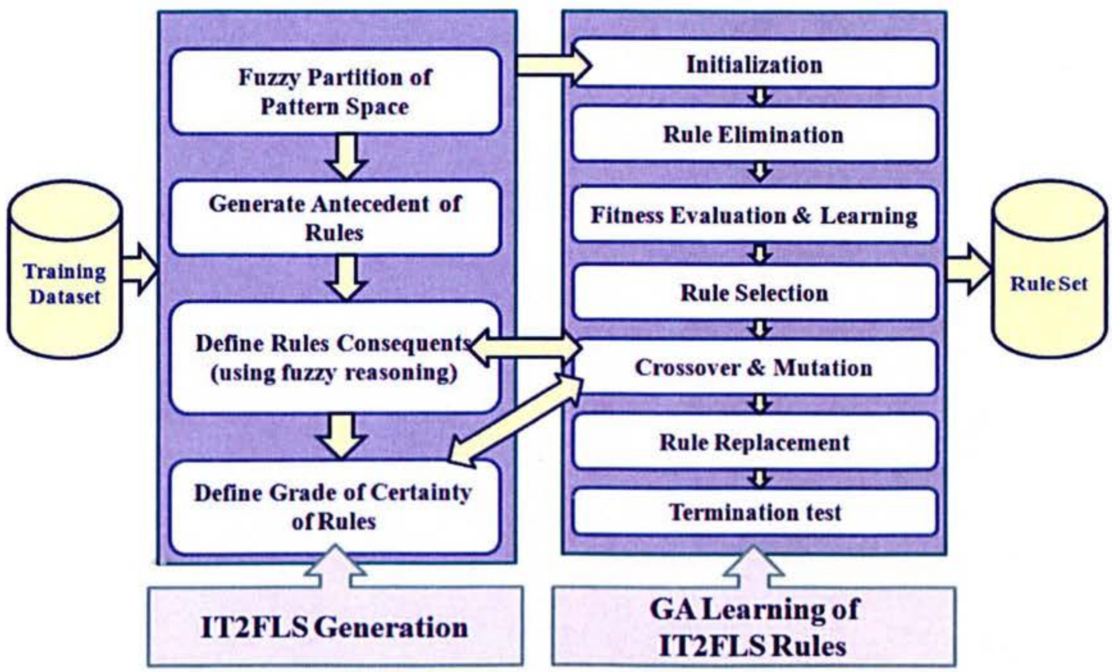


Figure A.2: Components of automatic IT2FLS generation software and relation between them