ADAPTIVE FUZZY LOGIC BASED FRAMEWORK FOR HANDLING IMPRECISION AND UNCERTAINTY IN PATTERN CLASSIFICATION OF BIOINFORMATICS DATASETS

BY

ZEEHASHAM RASHEED

A Thesis Presented to the
DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

Information & Computer Science

January 2009

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **ZEEHASHAM RASHEED** under the direction of his thesis advisor and approved by his thesis committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN COMPUTER SCIENCE**.

Thesis Committee

Dr. Tarek Helmy El-Bassuny (Thesis Advisor)

Dr. Mohamed Al-Mulhim (Thesis Co-Advisor)

Dr. Kanaan A. Faisal (Member)

Dr. Mostafa Elshafei (Member)

Dr. Alaa Eldin Ameen (Member)

Dr. Kanaan A. Faisal
Department Chairman

Dr. Salam A. Zummo
Dean of Graduate Studies

13/4/09

Date

Dedicated to my loving parents and my family

ACKNOWLEDGEMENT

In the name of Allah, the Most Beneficent Most Merciful

All praise is due to Allah, the source of all knowledge and strength. I acknowledge His infinite mercy and grace in making this work a success. And may His peace and blessings be upon his final messenger Muhammad, a guidance and inspiration to our lives. The successful completion of this work was made possible by a number of major contributions from different persons and organizations alike, to whom I wish to express my due gratitude.

I am grateful to King Fahd University of Petroleum & Minerals for giving me the opportunity to carry out this task under the guidance of a scholarly faculty and a proper research environment. My sincere gratitude and thanks goes to my thesis advisor Dr. Tarek Helmy El-Bassuny for his meticulous attention, his guidance, and his patience with me. I would also like to extend my appreciation to my thesis committee members Dr. Mohamed Al-Mulhim, Dr. Kanaan A. Faisal, Dr.Moustafa El-Shafei and Dr. Alaa Eldin Ameen for their encouragement and cooperation.

I am thankful to all colleagues and friends for their suggestions during the completion of this thesis. This work would not have been done so easily without their support, friendship, encouragement.

My heartfelt gratitude goes to my parents. I would never have been able to pursue this task without their cooperation and understanding. I am eternally indebted to my mother for all her prayers, concern, love and understanding. Whatever knowledge or abilities I have today have their roots in the love and teachings with which she has nurtured me. The value system ingrained by my father has driven me to always challenge myself and work hard towards my goals. I would like to especially thank my brothers and sister for their friendship and support which was both filial and maternal at the same time. Lastly, my deepest sense of gratitude to my fiance for her endless prayers, never ending concern, and much needed support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Full Name : Zeehasham Rasheed

Thesis Title : Adaptive Fuzzy Logic Based Framework for Handling Imprecision and Uncertainty in Pattern Classification of Bioinformatics Datasets

Major Field : Information and Computer Science

Date of Degree : January, 2009

Classification in the emerging field of Bioinformatics is a challenging task because the information about different diseases is either insufficient or lacking in authenticity as data is collected from different types of medical equipment. Also the limitation of human expertise in manual diagnoses leads to incorrect diagnoses. Moreover, the information gathered from various sources is subject to imprecision and uncertainty. Researchers utilized Artificial Neural Networks, Support Vector Machine and Bayesian Networks to achieve better classification, but the developed models are bedeviled by several limitations especially in uncertain situations. Recently, Type-1 and Type-2 Fuzzy Logic Systems (FLS) have been introduced as novel computational intelligence approaches for both prediction and classification. However Type-2 and other FLS have not been fully utilized in the bioinformatics and medical science. This thesis presents a Type-2 FLS-based classification framework for multivariate data to diagnose different types of diseases, which is capable of handling imprecision and uncertainty. As expected, this new computational intelligence approach overcomes the weaknesses of existing classifiers, particularly in the ability to handle data in uncertain situations such as uncertainty due to the existence of various types of noise, inconsistent expert opinions, ignorance and laziness. The classification accuracy and performance of the proposed framework are measured by using University of California, Irvine (UCI) well known medical datasets. The classification is performed on the basis of the nature of the inputs (e.g., singleton or non-singleton) and on whether uncertainty is present or absent. Empirical results have shown that the proposed FLS classification framework outperforms earlier implemented models with better classification accuracy among all existing classifiers. In addition, we conducted empirical studies on this classifier regarding the impact of various parameters of the proposed framework such as training algorithms and defuzzification methods.

خلاصــة الـرســالــة

الإسم	:	زي هاشم رشيد
عنوان الرسالة	:	إطار عمل باستخدام المنطق الغائم المكيف لمعالجة عدم الدقة وعدم التأكد في تصنيف النموذج لمجموعات بيانات البيوانفورماتيك
مجال التخصص	:	معلومات وعلوم الحاسب
تاريخ التخرج	:	يناير ، 2009

إن مشكلة التصنيف في مجال البيوانفورماتيك هي مهمة صعبة كون المعلومات حول الأمراض المختلفة هي إما غير كافية أو غير موثوقة بسبب تجميع البيانات باستخدام أدوات طبية مختلفة. إن محدودية الخبرة البشرية في تشخيص المرض يدوياً تؤدي الى تصنيف خاطىء في المجال الطبي. علاوة على ذلك المعلومات المجموعة من مصادر مختلفة قد تكون غير دقيقة أو غير مؤكدة. قدمت هذه الرسالة إطار عمل تصنيفي معتمد عى المنطق الغائم من النوع الثاني للبيانات المتغيرة لتشخيص الأنماط المختلفة من الأمراض وهو قادر على معالجة عدم الدقة وعدم التأكد. إن دقة التصنيف وأداء إطار العمل الجديدتم قياسها باستخدام UCI وهي مجموعات بيانات طبية معروفة وتمت مقارنة النتائج بمعظم المصنفات الشائعة في الأعمال المنشورة في علوم الحاسب والاحصائيات. لقد تم إنجاز التصنيف على أساس طبيعة الدخل كـ singleton أو non-singleton وفيما إذا كان عدم التأكد موجوداً في النظام أو لا. إن النتائج التجريبية أظهرت أن اطار العمل المقدم FLS يتفوق على النماذج المحققة السابقة وله دقة تصنيف أعلى من جميعها. بالإضافة لذلك، تم اجراء عدد من الدراسات التجريبية المتعلقة بتأثير البارامترات المختلفة لـ FLS مثل خوارزميات التدريب، و أساليب الـ defuzzification وخوارزميات التدريب.

CHAPTER 1

INTRODUCTION

The automatic diagnosis of cancer and other endemic diseases is an important, real-world medical problem. A major class of problems in medical science involves the diagnosis of disease, based upon various tests performed upon the patient. When several tests are involved, the ultimate diagnoses may be difficult to obtain, even for a medical expert. Over the past few decades, this has given rise to computerized diagnostic tools, intended to aid the physician in making sense of the confusing data in presence of imprecision and uncertainty [1]. These machine learning computerized tools and techniques have been employed by many disciplines to automate complex decision making and problem solving tasks. This field of research is immensely diverse, with applications in medicine, mathematics, computer science, chemistry, economics, business management and many other fields [2]. The majority of these tasks are concerned with pattern recognition problems. Pattern recognition is the act of taking in raw data and taking an action based on the “category” or the pattern [5]. In machine learning, the solution of pattern recognition problems lies within the field of supervised learning. The task is to learn (induce) the relationship between the dependent attributes (input) and the designated attribute (output) from a set of examples, i.e. generalize information collected from given data to unseen data. For example, given a set of training data consisting of the measurements of certain features (or variables) from examples of two categories (e.g. patients with malignant tumors and those with benign tumors), a learning system is required to determine the combination of features that is sufficient to distinguish one category (malignant) from the other (benign).

Data mining is defined as identifying valid, novel, potentially useful, and ultimately understandable patterns in data. In order to uncover these regularities, several techniques can be used, such as machine learning, statistical analysis, modeling techniques, database technology or human computer interaction [44]. These data mining methods originate in the field of Artificial Intelligence (AI) and machine learning [45].

Although data mining is quite a young discipline (about 25 years old), it is popular due to successful applications in telecommunications, marketing and tourism [46]. In recent years, its usefulness has also been proven in medicine [47]. Data mining aims to describe specific patterns (dependencies, interrelations, various regularities) which may be present in historical data. These patterns may be used to support future decisions in the diagnosis of new cases [46]. Such knowledge may also have an enormous value for decisions in treatment planning, risk analysis and other predictions. Prior to the mining process, it is essential to gain sufficient data [10]. This may require integrating data from multiple heterogeneous information sources and transforming it into a form specific to a target decision support application [48]. Afterwards the data has to be prepared for knowledge extraction by selecting the proper records and attributes.

1.1. Importance of Machine Learning in Medical Diagnosis

Human beings are always prone to make mistakes because of their limitations, and so correct diagnosis depends on the expertise of the doctor. Most physicians are confronted with the problem of deducing certain diseases or formulating a treatment based on more or less specified observations and knowledge [3]. Experience, which is the basis for a valid diagnosis, is obtained by the physician only after analyzing a sufficient number of cases. This expertise is reached only in the middle of a physician's career. This is especially true for rare or new diseases, where experienced physicians are in the same situation as newcomers. Principally, humans do not resemble statistical computers, but pattern recognition systems. Humans can recognize patterns or objects very easily but they fail when probabilities have to be assigned to observations [4]. A study was conducted to show that machine learning can help in making correct diagnoses [3]. The results of the study indicates that even the most experienced physician can diagnose correctly around 79% but diagnoses made with the help of machine learning are around 91% correct. From this result, one can conclude that computers make fewer errors than humans in making adhoc analyses of complex data.

A diagnosis is always based on symptoms identified in a patient's body and analyzed by a physician. There are four possible situations during this process [50]. First, True Positive (TP) means that the patient is correctly diagnosed as ill. True Negative (TN) means that the patient is correctly diagnosed as healthy. Those two situations are desired because they deliver correct predictions. On the other hand, there are situations when an ill patient is diagnosed as healthy (False Negative, FN) or when a healthy patient is diagnosed as ill (False Positive, FP). This concerns only two-class (binary) problems: sick-healthy, deceased-alive, etc. In the real world, however, a physician often has to decide which of many diseases a patient suffers from. The situation can get even more complicated if the patient suffers from several illnesses at once. The costs of FP and FN differ, especially in medicine. When an ill patient is classified as healthy, she would get no treatment and the unpredictable effects may include deterioration of the illness or even death. On the other hand, when a healthy patient is classified as ill, she would be treated in a wrong or inadequate way, which may cause health problems. Thus the real-world diagnosing process must be reflected in the data mining in the most appropriate way.

1.2. Imprecision and Uncertainty

The success of any classification framework that makes use of one or more sources of information is based on the availability of good historical data and experts' opinions. A framework that incorporates these things has two inherent problems: imprecision and uncertainty. Imprecision arises when an expert uses some quantitative criteria to differentiate between two or more classes. This is because an expert is a human being whose knowledge is imprecise due to the representation of knowledge in words. Moreover, the criteria defined are based on his past experience and may not be 100% precise. Therefore historical data is used by computers to make the boundaries between the classes precise.

Uncertainty comes into play when the device or instrument introduces noise while measuring the desired quantity. This can happen when different devices compute the same quantity (e.g., blood pressure or heart beat) but produces different outcomes due to different underlying understandings of the problem. The consequence of this is that the uncertainty in the internal attributes gives rise to uncertainty in the corresponding external attributes. Moreover, other factors that contribute to measure uncertainty are laziness/ignorance and to some extent the accuracy of the particular instrument. In general, uncertainty occurs mainly due to three reasons:

i. Laziness: There is too much work in listing all the antecedents and consequences in the problem domain.

ii. Theoretical Ignorance: We usually do not know enough about the domain to list every consideration.

iii. Practical Ignorance: Perhaps we do not have all the tests to run, or we do not want to run all the tests.

1.3. Problem Statement

A detailed literature review of the existing classifiers in bioinformatics reveals that there is no well defined classification framework which can handle imprecision and uncertainty in datasets. With respect to the various possible sources of uncertainties which we identified, our aim is to propose a new classification framework that can perform better than existing classifiers even in the presence of such uncertainties. In order to accomplish this, we need to design an adaptive Fuzzy Logic based framework that can deal with imprecision and uncertainties and will achieve promising classification accuracy. This work will also investigate the different parameters of the Fuzzy Logic System (FLS) and observe the impact on classification accuracy. These parameters include membership functions, nature of the training algorithms and defuzzification methods. The implementation of the proposed framework will be investigated with respect to different parameters. Classification accuracy will be evaluated on real time datasets and comparison will be carried out with the existing classifiers.

1.4. Thesis Organization

This thesis is organized as follows. The first chapter explains the motivation and significance of the proposed work, together with a general introduction and a problem statement. Chapter two reviews the relevant research literature. Chapter three gives a detailed overview of Fuzzy Logic Systems. Chapter four presents the proposed Fuzzy Logic based classification framework for handling imprecision and uncertainty. Chapter five discusses the impacts of algorithms and parameters on the performance of the proposed framework, and these parameters include membership functions, training algorithms, and defuzzification methods. The experimental setups, results and discussions, the conclusion and future work are presented in Chapter six.

CHAPTER 2

LITERATURE REVIEW

In the recent past, many classifiers were developed to explore various fields with the help of computer science. In fact, most of the research in the literature on disease classification used either statistical models or Artificial Neural Networks [6, 7, 8]. Statistical pattern recognition contributed greatly to the understanding of such classification problems. Widely used statistical methods included linear discriminant analysis, generalized linear regression, logistic regression, and nearest neighbor classification [5, 9].

In a health-care unit, a physician diagnosis a patient's condition based on the given symptoms. This information may be stored either in the medical unit's system or in the patient's files. This data may contain non-trivial dependencies [51], which may turn out to be valuable. Many methods and algorithms were used to mine data for hidden information. They included Artificial Neural Networks, Decision Trees, Fuzzy Logic systems, Naive Bayes, Support Vector Machines, Clusterization, Logistic Regression, and so on. The most frequently used algorithms for the medical support systems were the Decisions Trees (C4.5 algorithm), Artificial Neural Networks and the Naive Bayes [51, 52, 53]. These algorithms were able to reduce the time spent for processing symptoms and producing diagnoses, making them more precise at the same time. However, most of the research studies assessed the algorithms on a narrow set of medical databases (no more than three) [54, 55].

2.1. Earlier work on Artificial Neural Networks

Artificial Neural Networks (ANNs) are networks of units, called neurons, that exchange information in the form of numerical values via synaptic interconnections, inspired by the biological neural networks of the human brain. ANNs provided very powerful and flexible approaches to function approximation [10]. ANNs are mainly the Feed Forward Networks, such as multilayer perceptrons and radial basis function Neural Networks, which are widely used to develop diagnostic models [8].

Past studies sought ways of capitalizing the use of Neural Networks in medical diagnosis of breast cancer. To improve the results of breast cancer screenings, Gurcan et al. (2002), evaluated the performance of a back-propagation ANN to predict an outcome (cancer/not cancer) to be used as classifier [11]. ANN was trained on data from the family history of cancer, and socio-demographic, gyneco-obstetric and dietary variables. A complete method was proposed for fast detection of circumscribed mass in mammograms by employing Radial Basis Function Neural Networks (RBFNN) in which each neuron output is a nonlinear transformation of a distance measure of the neuron weights and its input vector [12]. The study in [13] indicated that oncologists can be helped when the result is obtained by different types of network, such as the desired Feed Forward Neural Network rule extraction algorithm or the Radial Basis Function, or the General Regression Neural Network, or the Probabilistic Neural Network.

The problem with Neural Networks is that they usually adopt gradient-based learning methods which are susceptible to local minima and long training times [6] especially when the number of classes/categories is high. For example, the Subsequent

Artificial Neural Network (SANN) in [7] had one ANN and up to 91 SANNs to be trained for each experiment. For each network, there are five modules, each consisting of 10 hidden nodes. This means that, for each experiment, up to 4,600 hidden nodes are needed for the training process. Conversely the Extreme Learning Machine technique required less than 50 hidden nodes [6, 7, 14, 18]. So ANNs usually produce less classification accuracy, and they need considerable training time while updating the input output weights.

The study in [59] introduces Artificial Neural Networks with back propagation for classification of heart disease cases. This solution was implemented in a medical system to support the classification of the Doppler signals in cardiology. The predictions yielded by [59] were more accurate than those presented in [60]. The authors of the article [61] claimed that the Multilayer Perceptron is one of the most frequently employed neural network algorithms in modern medical diagnosis systems. They discussed applications of this algorithm to classification of different cancers (hepatic, lung and breast cancers) and other diseases. The study in [62] presented two different Neural Network techniques are presented. NeuroRule and NeuroLinear were applied to diagnosis of hepatobiliary disorders. The Neural Networks' major disadvantage is complexity [62], which makes the classification process difficult to interpret. Nevertheless, the authors proved that they produced effective classifications in case of medical data. The medical application of Neural Networks was also presented in [63] and [64].

2.2. Earlier work on Bayesian Belief Networks

Some recent work was done with other artificial intelligence techniques, such as Bayesian Belief Networks (BBN) [15]. This statistical and graphical modeling approach is based upon direct application of Bayes Theorem, and it works on the assumption that the attributes are statistically independent from each other [5, 10]. Some optimization techniques were also used, such as the Hill Climbing Bayesian Network, Laplace smoothing, Decision Trees or C4.5L to improve the ranking of classifiers [16]. An application of Bayes' law in medical analyses was first proposed in 1959 [51] in an article about theoretical possibilities of applying this solution in physicians' everyday work. This idea was realized in 1972 by an implementation of a medical system to support the diagnosis of abdominal pain by using the Naive Bayes algorithm. This classifier assumes that all attributes are independent. For many years, scientists and medical staff tried to develop a suitable diagnosis system using the Bayesian theorem. Several studies on this problem were presented in [52] and [66]. Simplicity, learning speed and classification speed are the main advantages of the Bayesian classifier [67]. On the other hand, one of the most serious drawbacks is its ad-hoc restrictions placed on the graph, making the classifications hard to understand [51]. This method must be implemented with care, as diagnoses must be thoroughly understandable.

2.3. Earlier work on Decision Trees

Besides the Neural Network, Decision Trees are also utilized in medical knowledge extraction [47, 53]. Decision tree algorithms recursively partition the data, and they give rise to a tree-like structure [10]. The decisions are usually simple attribute tests, using one attribute at a time to discriminate the data. New data can be classified by following the conditions at the nodes down into a leaf. Decision trees have been used extensively in work from both machine learning and statistics. Their main advantage is simplicity and the easy-to-comprehend structure of the generated models [50]. Several algorithms generate trees. Vlahou et al. [53] and Duch et al. [47] applied Decision Trees classification for diagnosis of an ovarian cancer and a Melanoma skin cancer, respectively. Decision trees are applicable also in other fields of medicine. The authors of [65] compared the accuracy of Decision Tree with a Bayesian Network in diagnosis of female urinary incontinence. The obtained classifications were better in the Decision Tree, but the difference was small.

2.4. Earlier work on Support Vector Machines

Support Vector Machine (SVM) has been proposed as a very effective method for pattern recognition, machine learning and data mining [17]. The general idea is to map non-linearly D-dimensional input space into a high dimensional feature space. A linear classifier (separating hyper plane) is constructed in this high dimensional space to classify the data. The use of the kernel trick allows the classifier to be constructed without explicit knowledge of the feature space. SVM is considered to be a good method because of its high generalization performance. Intuitively given a set of points which belong to either one of the two classes, a SVM can find a hyper plane having the largest possible fraction of points of the same class on the same plane. This hyper plane is called the Optimal Separating Hyper plane (OSH) and it can minimize the risk of misclassifying examples of the test set. When the One-Versus-All (OVA) approach is used to make binary classifiers applicable to multi category problems, the number of classes increase as the complexity of the overall classifier also increases. So the system becomes more complex, and it requires extra computations [18].

2.5. Earlier work on Extreme Learning Machine

Huang et al. [19, 20, 21] proposed a new learning algorithm called the Extreme Learning Machine (ELM) for Single Hidden Layered Feed Forward Neural Networks (SLFNs). In ELM, one may randomly choose (according to any continuous sampling distribution) and fix all the hidden node parameters and then analytically determine the output weights of SLFNs [19]. After the hidden nodes parameters are chosen randomly, SLFN can be considered as a linear system, and the output weights can be analytically determined through a generalized inverse operation of the hidden layer output matrices. Studies have shown [19, 22] that ELM has good generalization performance for classification and can be implemented easily. Many nonlinear activation functions can be used in ELM such as sigmoid, algebraic sigmoid, sine, hard limit, and radial basis functions [20, 21], and complex activation functions [15].

2.6. Earlier work on Fuzzy Rule Based Systems

In the past fuzzy rule-based systems were applied mainly to control problems, but recently they are also applied to pattern classification problems. Various methods were proposed for the automatic generation of fuzzy if-then rules from numerical data for pattern classification [23, 24, 25] and they were shown to work well on a variety of problem domains. Several Artificial Intelligence (AI) techniques including Neural Networks and Fuzzy Logic [26, 27] were successfully applied to a wide variety of decisions in medical diagnosis [28, 29].

The authors of [56], [57] and [58] worked on medical rules induction. The article [56] presented study on unsupervised fuzzy clustering algorithms and rule based systems, which are useful in the labeling of tomography images. The presented methods turned out to be computationally efficient for one class of problems, as proven by the results of the studies. However in other applications their effectiveness was much lower. In some applications the generated rules are claimed to be easy to construct and modify. Furthermore, their independency allows for changing one rule without affecting the others.

In the paper [57] the rules extraction was achieved with the use of an Artificial Neural Network Multilayer Perceptron. The authors proposed an algorithm C-MLP2LN. It generates additional nodes, deletes the connections among them, and optimizes the rules. Such a solution leads to simpler and more accurate rules. The authors of [58] presented a study on the generation of rules which describe associations among attributes. The experiments were conducted on real medical data and their correctness is verified by statistical measures and physicians evaluations. This article analyzed of real data from St. Thomas' Hospital in London, and it described all the steps performed: from preprocessing, through data mining experiments, to verification of accuracy of the results.

In research on the analysis of stock prices, a Type-2 Fuzzy Rule Based Expert System was developed. The model applied the technical and fundamental indexes as the input variables, and it produced very encouraging results. Dongrui and Mendel [30] proposed a Vector Similarity Measure (VSM) for Interval Type-2 Fuzzy Sets, whose two elements measure the similarity in shape and proximity for a linguistic approximation problem, and they reported a bright performance. Li et al. [31] presented a model of the redundant structure in ecosystems by using a Type-2 Fuzzy Logic System to demonstrate quantitatively the communications between organisms and the environment. The result of the experiment showed the performance of the model in determining the community with higher reliability that has stronger fitness while the environment changes. Castillo et al. [32] used the Fuzzy Lyapunov Synthesis as proposed by Margaliot [33] to build a Lyapunov Stable Type-1 Fuzzy Logic Control System, and they made an extension from Type-1 to Type-2 to ensure the stability on the control system and to prove the robustness of the corresponding fuzzy controller. Research based on hybridization of Support Vector Machines (SVMs) and Type-2 FLS was performed by Chen et al. [34] to better handle the uncertainties existing in real classification data and in the Membership Functions (MFs) in the traditional Type-1 FLS. This was achieved by using type-2 fusion architecture to incorporate the classification results from individual SVM classifiers and to generate the combined classification decision as the output.

2.7. Uncertainty Handling Problems in Existing Classifiers

As a result of the above survey, we can see that there is a significant need for a classification framework that can handle the uncertainty in datasets. For this purpose, we investigated some deficiencies present in the earlier classifiers. For example, the Bayesian Network is based on probabilistic interactions of observations, and some of the observed values may not be correct. While some explanations highlight erroneous values as conflicting findings, no existing method can explain how errors may be compensated during inference. Similarly, current methods do not explain how the missing values along the influence paths to the target node are derived. Secondly, they require an exponential number of analyses for all the possible subsets of evidence variables. They are limited to analyzing findings individually, and are computationally intractable for reflecting variable interactions.

As far as Neural Networks are concerned, the influence of the noisy inputs on the output variable together with the transfer functions, implicit in the values of the weights. Hence an unattractive feature of such networks is that the number of weights and complexity increase greatly as the network grows. Also the weights may not always be easy to interpret if the data is imprecise and uncertain, which leads to the problem of under fitting or over fitting, and the problem becomes difficult to visualize from an examination of the weights.

In SVM classifiers, problems with corrupted inputs are more difficult than problems with no input uncertainty. Even if there is a large margin separator for the original uncorrupted inputs, the observed noisy data may become non-separable. For example, by using a kernel function in SVM, the input vector is mapped on to a usually high dimensional feature space, and the uncertainty in the input data introduces uncertainties in the feature space. To overcome this problem, researchers used total least square regression methods with SVM, but they could not achieve promising results.

CHAPTER 3

FUZZY LOGIC SYSTEMS

3.1. Fuzzy Logic

Fuzzy Logic was first proposed by Lotfi A. Zadeh, Professor of Systems Theory at the University of California, Berkeley, USA, in a publication in 1965 [35]. Fuzzy Logic was his term for a system of mathematics developed to model the human brain's curious way of processing and selecting words. The main motivation behind fuzzy logic was the imprecision in the measurement process, where by precise statements lose meaning and meaningful statements lose precision as the complexity rises [36].

3.2. Fuzzy Logic Systems

Fuzzy Logic System (FLS) is the name for the system which has a direct relationship between fuzzy logic and fuzzy concepts (fuzzy sets, linguistic variables, and so on). The most popular FLS in the literature may be classified into three types, namely Pure Fuzzy Logic Systems, Takagi and Sugeno's Fuzzy System, and Fuzzy Logic System with fuzzifier and defuzzifier. As engineering applications mostly produce crisp data as output and expect crisp data as input, the last type is the most widely used [37]. Figure 3-1 shows the basic configuration of a FLS with fuzzifier and defuzzifier.

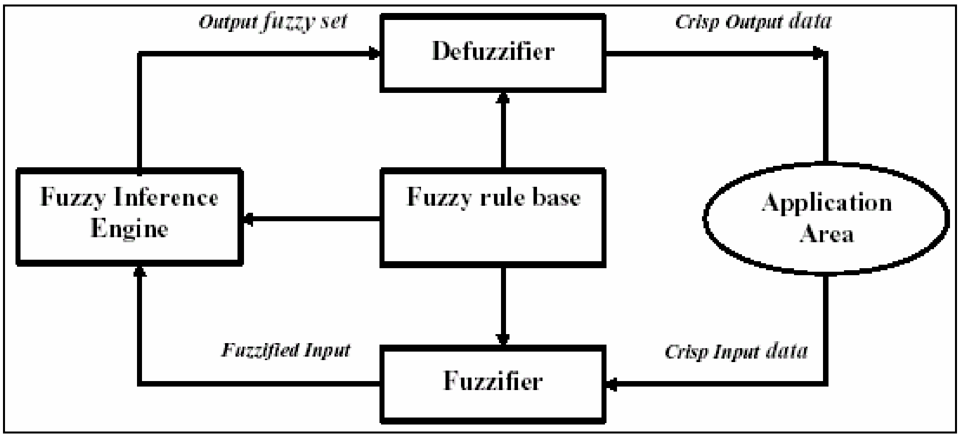

Figure 3-1: Fuzzy logic system with fuzzifier and defuzzifier

This type of FLS was first proposed by Mamdani [38]. It was successfully applied to a variety of industrial processes and consumer products. The main four components' functions are as follows.

Fuzzifier: Does a mapping from crisp input to a fuzzy set.

Fuzzy Rule Base: Fuzzy Logic Systems use fuzzy IF-THEN rules. A fuzzy IF THEN rule is of the form "IF X1 = A1 and X2 = A2 ... and Xn = An THEN Y = B" where Xi and Y are linguistic variables and Ai and B are linguistic terms. The IF part is the antecedent or premise, while the THEN part is the consequence or conclusion. In a FLS, the collection of fuzzy IF-THEN rules is stored in the fuzzy rule base, which is referred to by the inference engine when processing inputs.

Fuzzy Inference Engine: After all the crisp input values have been fuzzified into their respective linguistic values, the inference engine will access the fuzzy rule base of the fuzzy expert system to derive linguistic values for the intermediate as well as the output linguistic variables. The two main steps in the inference process are aggregation and composition. Aggregation is the process of computing for the values of the IF (antecedent) part of the rules. Composition is the process of computing for the values of the THEN (conclusion) part of the rules.

Defuzzifier: Does a mapping from the fuzzy output to the crisp output.

3.3. Adaptive Fuzzy Logic Systems

The definition of Adaptive Fuzzy System given by Wang in [37] is easy to understand – **"An Adaptive Fuzzy System is defined as a Fuzzy Logic System equipped with a training algorithm, where the fuzzy logic system is constructed from a set of fuzzy IF-THEN rules using fuzzy logic principles, and the training algorithms adjust the parameters of the Fuzzy Logic System based on numerical information**". Here parameters, e.g. position and mean or standard deviation, are the necessary values to construct the membership functions. Membership functions are adjusted by a set of input-output pairs. This is adaptive in the sense that the necessary changes are made only locally to the affecting membership functions, whereas trainable neural networks globally adjust all the weights. So, Adaptive Fuzzy Logic is a nice way of combining linguistic and numerical information, which can be done in two ways [37]:

i. Use linguistic information (experts' knowledge) to construct an initial FLS, and then adjust the parameters of the initial FLS based on numerical information.

ii. Use numerical information and linguistic information to construct two separate FLSs, and then average them to obtain the final FLS.

In the first case components of the FLS are based on the expert's opinion. These components include the number of rules, the shape and position of the membership functions, and the shape and position of the consequents. Historical data is then used to further tune the parameters of the FLS. The second way to make an adaptive FLS is straightforward in the sense that after the parameters of the individual systems are available, one can average these parameters to produce a final FLS.

The above discussion shows that in a FLS, the experts are the source of information for establishing rules. The rules are then treated by using the historical data. But the problem with the classical FLS (also known as Type-1 after the advent of Type-2 as discussed below) is that it handles only the imprecision and not the uncertainty. To handle both the uncertainty and imprecision, one needs to establish a Type-2 system, which is supposed to handle all sources of uncertainty [39].

3.4. Uncertainty in Fuzzy Logic Systems

Mendel [39] noted that uncertainty exists while building and using typical Fuzzy Logic Systems. He described four sources of uncertainty. Those are summarized here.

i. *Uncertainty about the meanings of the words that are used in a rule.* This uncertainty affects the membership functions, because these membership functions represent words in a FLS. It can be both antecedents and consequents.

ii. *Uncertainty about the consequent that is used in a rule.* This uncertainty affects the rule itself. A rule in FLS describes the impact of the antecedents on the consequent. Experts may vary in their opinions about the nature of this impact.

iii. *Uncertainty about the measurements that activate the FLS.* This uncertainty affects the crisp input values or measurements that activate the FLS systems. These measurements may be noisy or corrupted. This noise can again be in a certain range or totally uncertain meaning (stationary or non-stationary).

iv. *Uncertainty about the data that are used to tune the parameters of a FLS.* This uncertainty affects the measurements again. But these measurements are used to train the FLS whereas the measurements of (iii) are used to activate the FLS.

3.5. Uncertainty and Type-2 Fuzzy Logic Systems

Mendel has proposed Type-2 fuzzy sets and Type-2 FLS to deal with four types of uncertainties discussed in the previous section. Type-2 fuzzy sets were first proposed by Zadeh [40] in 1975, but they were first characterized in 1999 by Mendel and Liang [41]. Actually Type-2 fuzzy sets are the three-dimensional, whereas Type-1 is two-dimensional. This extra dimension allows uncertainty to be handled by Type- 2 fuzzy sets.

We will now see the definition of Type-2 fuzzy sets and how they can help to model uncertainty. We use the definition and figures from Mendel's book [39]. Type-2 fuzzy sets help us to deal with the first source of uncertainty, i.e. uncertainty about the meaning of the words. Type-1 fuzzy sets cannot deal with this type of uncertainty, because the degree of membership is considered as certain in Type-1 fuzzy sets. On the other hand, the blurred area, i.e. the second dimension in a Type-2 fuzzy set adapts the concept of uncertainty. Mendel calls this blurred area a footprint of uncertainty (FOU). Here the concept is to consider different degrees of membership for each of the values in the universe of discourse. Fuzzy sets are used to represent words or linguistic variables, and people differ in how to interpret a particular word. So, the concept of a second dimension in type-2 fuzzy set provides this flexibility to incorporate different persons' views in a fuzzy set. We will discuss this issue further in the later part of this thesis.

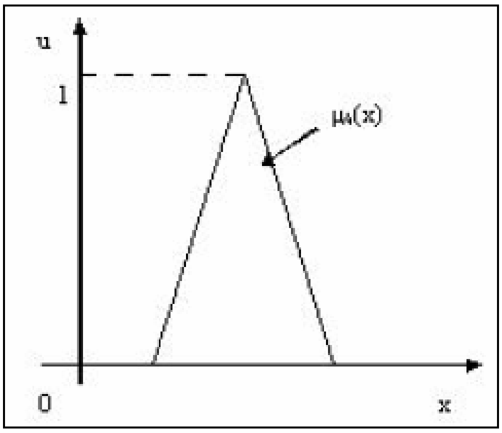

Figure 3-2: A Type-1 triangular membership function

Let us imagine blurring the Type-1 membership function depicted in Figure 3-2 by shifting the points on the triangle either leftwards or rightwards and not necessarily by the same amounts, as in Figure 3-3. Then, at a specific value of x, say x′, there is no longer a single value for the membership function. Instead, the membership function takes on values wherever the vertical line intersects the blur.

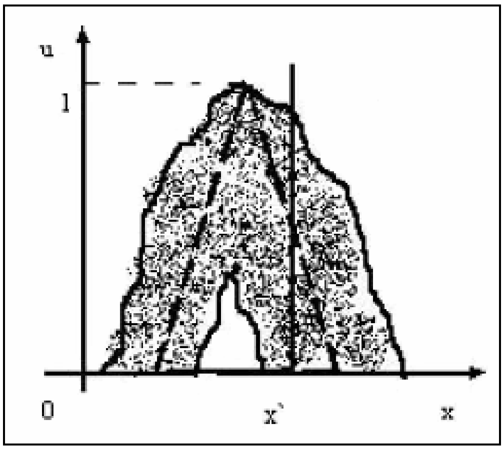

Figure 3-3: Blurred triangular membership function

Those values need not all be weighted the same. Hence, we can assign an amplitude distribution to all of those points. Doing this for all $x \in X$, we create a three-dimensional membership function (a Type-2 membership function) that characterizes a Type-2 fuzzy set. Type-2 membership functions have the same constraint as Type-1 membership functions. The degree of membership along the second dimension is always in the interval [0, 1]. The amplitude distribution i.e. the values along the 3rd dimension, also lie between the interval [0, 1]. So it is clear that, if the blur disappears, then a Type-2 membership function must reduce to a Type-1 membership function.

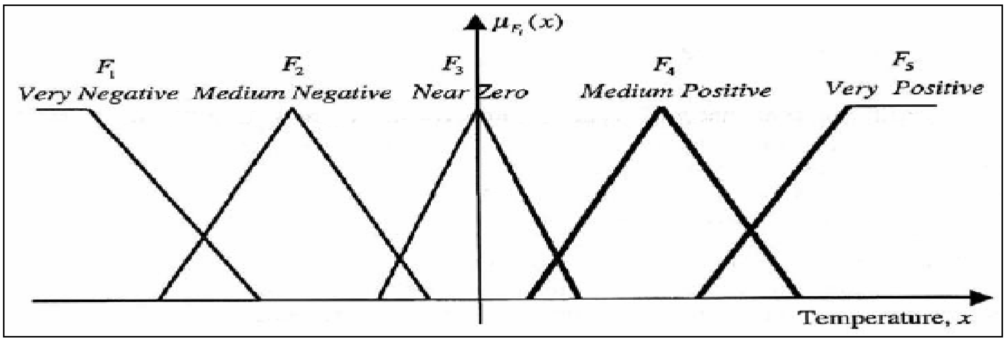

Figure 3-4: Type-1 Fuzzy sets

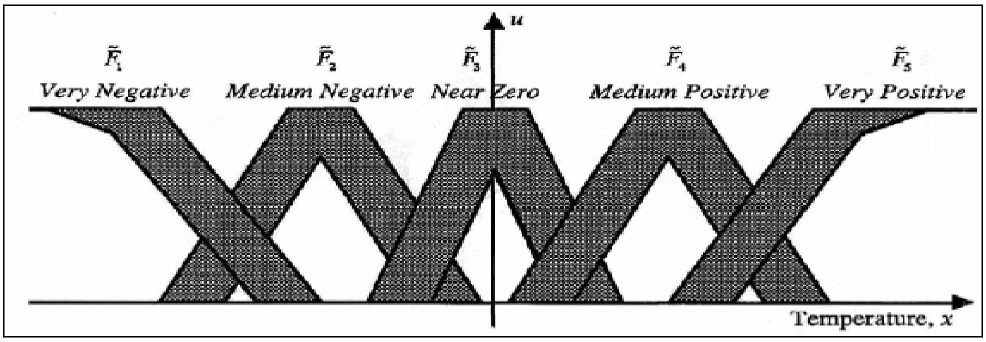

Figure 3-5: FOUs for membership functions of Figure 3-4

Figure 3-4 shows some triangular membership functions, and Figure 3-5 shows FOUs for those membership functions [39]. The shaded or blurred area is our FOU, i.e. the second dimension that helps to deal with uncertainty. We see in the figure that this FOU is uniformly shaded. It means that, at each point in the FOU, the membership degree is one. Membership functions of this type are called interval Type-2 membership functions. Imposing this constraint helps to build the Fuzzy Logic System, but it also poses some limitation. Now, let us see some examples of Type-2 Gaussian membership functions. Let us consider the case of a Gaussian membership function having a fixed standard deviation, σ, and an uncertain mean that takes on values in [m1, m2]. Figure 3-6 is an example.

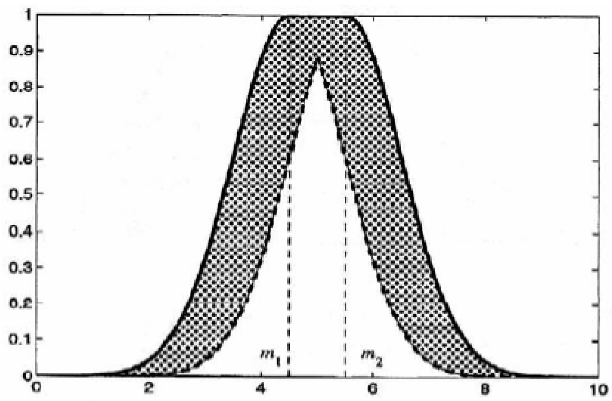

Figure 3-6: FOU for Gaussian membership function with uncertain mean

Similarly, let us consider a Gaussian membership function having a fixed mean, m, and an uncertain standard deviation that takes on values in [σ1, σ2]. Figure 3-7 is an example.

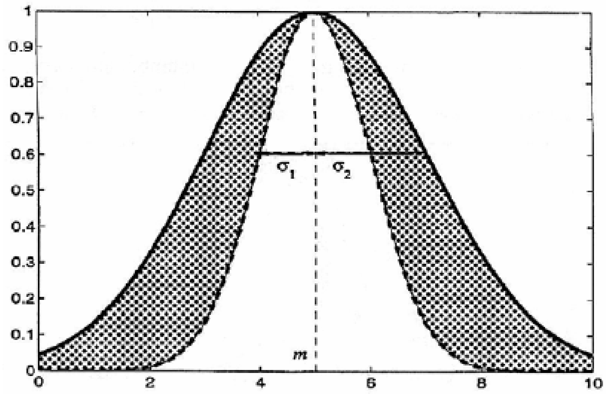

Figure 3-7: FOU for Gaussian membership function with uncertain standard deviation

It is easy to see here that both Gaussian membership functions are interval Type-2, as the shading is uniform.

3.6. Fuzzification in Type-2 Fuzzy Logic Systems

A Fuzzy Logic System is considered to be Type-2 as long as any one of its antecedent or consequent sets is Type-2. All the components of Figure 3-8 were discussed in detail by Mendel [39]. Fuzzifier is one of the most important components from the aspect of uncertainty. Here we shall discuss fuzzification, because it helps to deal with uncertainty.

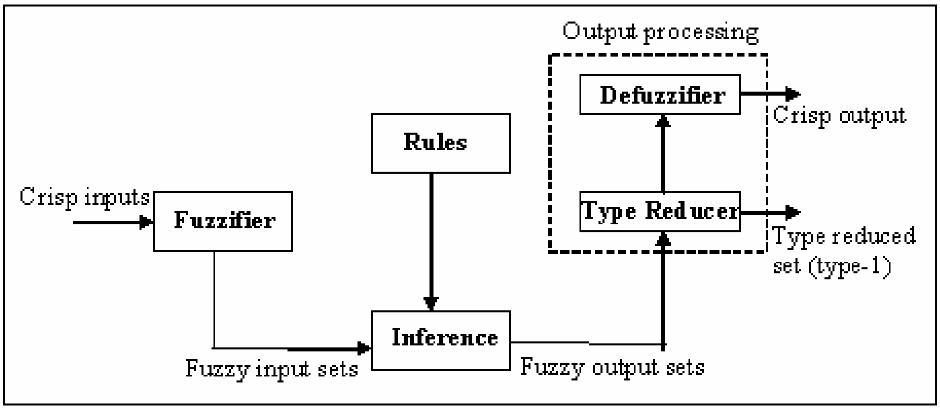

Figure 3-8: Type-2 FLS

Fuzzification can be done in two main ways (singleton and non-singleton). Singleton fuzzification considers the measurement that activates the FLS to be certain and noise free. Non-singleton considers the input crisp measurement to be uncertain. In singleton, the result of fuzzification is a fuzzy singleton, i.e. only at the input measurement the membership function has a value of 1. On the other hand, conceptually, a non-singleton fuzzifier implies that the given input value is the most likely to be correct from all the values in its immediate neighborhood. However, because the input is corrupted by noise, neighboring points are also likely to be the correct value, but to a lesser degree. So, fuzzy membership function is used for fuzzification where this function is centered at the measurement value. This non-singleton fuzzification can also be done in two ways (Type-1 and Type-2) based on the type of fuzzy sets used for fuzzification. When the noise is stationary, we can use the Type-1 non-singleton fuzzification. When the noise is non-stationary, we can use Type-2 non-singleton. Based on different types of fuzzification and different types of antecedent fuzzy sets, Mendel developed 5 different Fuzzy Logic Systems:

i. Singleton Type-1

ii. Non-singleton Type-1

iii. Singleton Type-2

iv. Non-singleton Type-2 with Type-1 inputs

v. Non-singleton Type-2 with Type-2 inputs

Figure 3-9 shows a pictorial description of these 5 different fuzzy logic systems [22].

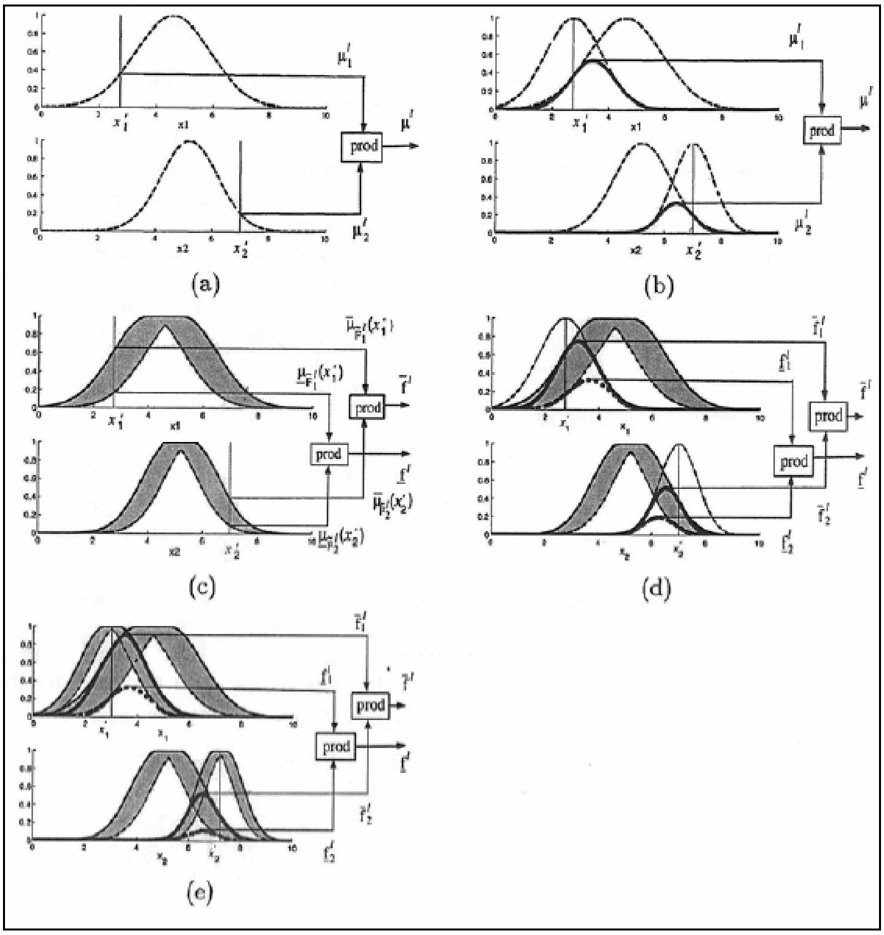

Figure 3-9: Different types of FLS – (a) singleton Type-1, (b) non-singleton Type-1, (c) singleton Type-2, (d) non-singleton Type-2 with type-1 inputs, (e) non-singleton Type-2 with Type-2 inputs

Mendel has shown which type of noise i.e. uncertainty can be handled by which FLS. See Table 3-1.

Table 3-1: Different Fuzzy Logic Systems to handle different types of noise

Type of FLS	**Measurement Noise**	**Training and Testing Data**	**Measurements that is used after building the FLS**
Singleton Type-1	None	Noise Free	Noise Free
Non-singleton Type-1	Stationary	Noisy	Noisy
Singleton Type-2	Stationary	Noisy	Noise Free
Type-1 non-singleton Type-2	Stationary	Noisy	Noisy
Type-2 non-singleton Type-2	Non-stationary	Noisy	Noisy

CHAPTER 4

HANDLING UNCERTAINTY IN THE PROPOSED FLS BASED CLASSIFICATION FRAMEWORK

As a result of our survey, we can see that there is significant need for a framework that can handle the uncertainty surrounding the classification process. This chapter provides solutions for handling various sources of uncertainty when performing classification. In order to counter the impact of uncertainty on the classification framework, we incorporate the concept of a Type-2 fuzzy logic system in our framework. The experimental results obtained by using Type-1 and Type-2 systems strengthened our intuition that Type-2 outperforms Type-1.

4.1. Major Components of Fuzzy System

The general architecture of a Type-2 FLS was already discussed in Chapter 3. In the following sections, we will look at the major components regarding fuzzification and rule base formulation by considering uncertainty issues.

4.1.1. Antecedent Fuzzy Sets

In building an FLS, we divide the whole range of all the internal and external attributes into several fuzzy sets. This fuzzy set classification can be obtained by the experts or by the analysis of numerical data sets. As we have already said, different experts may provide a different assessment, based on their past experience, regarding a particular fuzzy set range (e.g., LOW) of the internal attributes. Different experts can have different ranges. This causes an uncertainty as to which definition is reliable, when one wants to define antecedents while developing an FLS. This situation seems problematic, but at the same time it is interesting and advantageous, as Mendel himself says [39]:

> *"Uncertainty is good in that it lets people make decisions (albeit conservative ones)."*

This observation led Mendel to the idea of Type-2 fuzzy sets, which enable us to model uncertainty, caused by different experts' opinions as just discussed, in the FLS by blurring the antecedents' boundaries and defining the footprint of uncertainty (FOU). Similarly, obtaining fuzzy set classification through analysis of numerical data can cause uncertainty, if the data contains more than one metric value for a particular internal attribute. This uncertainty should be reflected in the antecedents by using Type-2 fuzzy sets, so that the impact of uncertainty can be propagated to the outputs through the FLS inference engine.

4.1.2. Consequent Fuzzy Sets

Uncertainty in consequent arises when two or more experts, based on their experience, relate the impact of the same antecedent fuzzy set on more than one consequent fuzzy set using fuzzy rules. In order to handle this situation Mendel proposed three possibilities:

i. Keep the response chosen by the largest number of experts.

ii. Find a weighted average of rule consequents for each rule.

iii. Preserve the distributions of the expert responses for each rule.

Mendel opted for the second solution, and he derived the defuzzification method which accomplishes this task.

4.1.3. Training FLS

After setting up the antecedents and the consequent fuzzy sets by incorporating the uncertainty through Type-2 representation, there is a need to train the parameters if one is employing an adaptive FLS, such as the one described in Section 3.3. In order to accomplish this, we need historical data having input attributes as well as output attribute as classes. Also, for the case where noisy input which gives rise to uncertainty, the situation can be handled if one uses non-singleton fuzzification by properly defining the parameters of the input membership functions [39].

4.1.4. Activating FLS

Finally, we need data to activate the FLS. This data may be testing data to validate the performance of the FLS, since the noise has already been taken care of during training and the parameters are already tuned.

4.2. Proposed Framework

The proposed framework is given in Figure 4-1. This framework is different from classical frameworks for two reasons. First, the training algorithms are adaptive, and they continuously adjust the parameters of the FLS based on numerical information. Second, the proposed framework has a Modified Height defuzzifier which takes care of all the deficiencies present in an ordinary Height defuzzifier. The basic idea and mathematical derivation of this Modified Height defizzifier are stated in section 5.2.2.

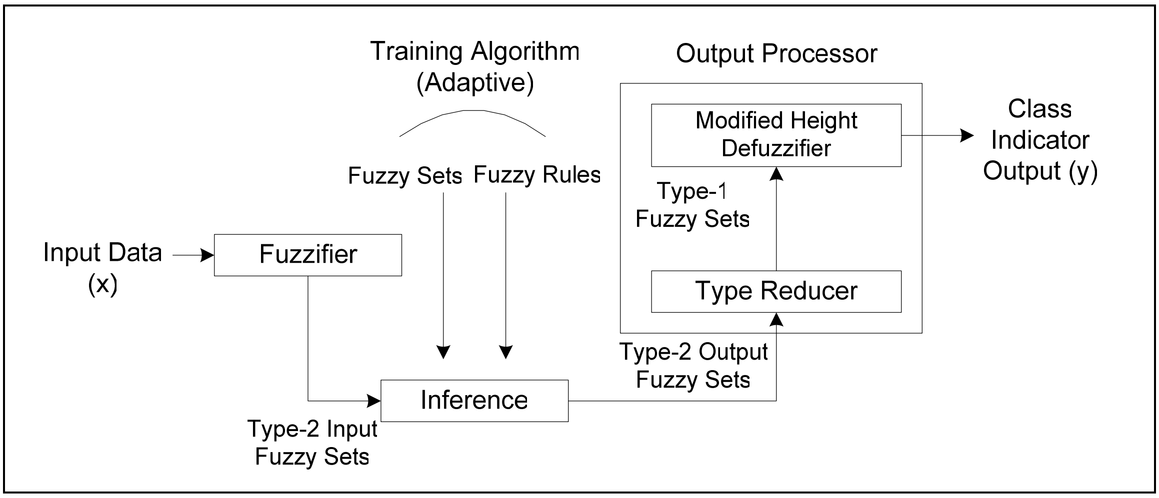

Figure 4-10:Proposed Adaptive FLS Classification Framework

For the initialization of the framework, we consider our antecedents and consequent membership functions with uncertain mean. After defining the type of membership functions, we need to classify antecedents and consequents into a suitable number of fuzzy sets. After that, we need to initialize the parameters of membership functions, i.e. means and standard deviations. Due to the adaptive nature of the training algorithm, the parameters of the antecedents can be initialized by running experiments and looking at the behavior of training, because here we assume imprecision and uncertainty due to laziness, and that is why we cannot compute means and standard deviations for only one time. Therefore, one can initialize the consequents by analyzing the training data or by running experiments and noticing the behavior of training. In developing the above framework, we consider (i) Initializing the Framework, (ii) Training the Framework, and (iii) Testing or Validating the Framework.

4.2.1. Initializing the Framework

For initializing the framework, we need to define some components. Initializing a FLS means initialization of its antecedents, consequents and the fuzzy rules. These components of an FLS can be initialized either from the numerical dataset or from the expert opinion. In this study, we initialized the FLS from the numerical dataset. In this model, we have: the antecedents and consequents; internal attributes or input fields as antecedents; and external attributes or class categories as consequents. Our framework will support one external attribute based on several internal attributes. If-Then rules will form the rule base by using these internal and external attributes.

To initialize from numerical data, we make use of a training data set taken from the available measurement data. For this, we need to define and initialize the various components of the Type-2 FLS developed by Mendel [39]. In this context we consider our antecedent and consequent to be Type-2 Gaussian with uncertain mean. The input membership functions will be Type-1 Gaussian. The Type-2 FLS expects that we shall have one or more measurements available in the dataset for each internal or external attribute. Our model will define the initial fuzzy sets for both antecedents and the consequent from the dataset. It should be noted at this point that we are using the FLS developed by Mendel [39], which tries to completely specify the FLS by using the training data, which can be interpreted as a collection of IF-THEN rules. Each rule is of the form shown below, where F_j^i are fuzzy sets described by the following Gaussian membership function, i.e.

$$\mu_A(x) = \exp\left\{-\frac{1}{2}\left(\frac{x-\sigma_{F_j^i}}{\sigma_{F_j^i}}\right)^2\right\}$$

where i=1….n, and j =1…p, with n and p representing the number of samples and number of attributes respectively.

Let us suppose that we need to initialize F fuzzy sets for the attribute A. Each attribute has m measurements. In the training dataset, we have attribute measurements for n samples. Table 4-1 shows the structure of the training dataset for one attribute.

Table 4-2: Structure of the training dataset

Sample #	**Measure 1**	**Measure 2**	**…**	**Measure m**	**Mean of Measure**	**Standard Deviation of Measure**
1	X_{11}	X_{12}	…	X_{1m}	μ_1	σ_1
2	X_{21}	X_{22}	…	X_{2m}	μ_2	σ_2
…	…	…	…	…	…	…
n	X_{n1}	X_{n2}	…	X_{nm}	μ_n	σ_n

Now, we have to calculate the following:

M_1 = Minimum ($\mu_1, \mu_2 \ldots\ldots \mu_n$)

M_2 = Maximum ($\mu_1, \mu_2 \ldots\ldots \mu_n$)

$M = \text{Mean}\ (\mu_1, \mu_2 \ldots \mu_n)$

$S = \text{Standard Deviation}\ (\mu_1, \mu_2 \ldots \mu_n)$

$R_1 = \text{Minimum}\ (\sigma_1, \sigma_2 \ldots \sigma_n)$

$R_2 = \text{Maximum}\ (\sigma_1, \sigma_2 \ldots \sigma_n)$

$R = \text{Mean}\ (\sigma_1, \sigma_2 \ldots \sigma_n)$

$T = (M_2 - M_1) / (F-1)$

Now, if U_{i1} and U_{i2} are the uncertain means for i^{th} fuzzy set, they are defined as follows:

$U_{i1} = M_i - \alpha R$

$U_{i2} = M_i + \alpha R$

$\alpha = \beta R$

where i = 1….F and F are the number of fuzzy sets.

In this framework, the consequents are computed by combining all the possibilities of fuzzy sets in the antecedents by using these formulas. Moreover the Type-2 FLS derived by Mendel [39] provides control over the consequents which represent the interval of confidence, and it can be initialized by using past experience or by observing the training behavior.

4.2.2. Training the Model with Type-2 Learning Procedures

After initializing the FLS, our model goes into the training step. For this purpose, the dataset contains the independent (input variables) and dependent (class variable) attributes. We have discussed in the previous section how this data should be organized and used to initialize the FLS. The same dataset will be used as training data. The objective of the training algorithm is to minimize the error function for E training Epochs.

$$\text{E(t)} = \left(\frac{f(x^{(t)}) - y^{(t)}}{y^{(t)}} \right)^2 \quad \text{t} = 1, \ldots, \text{N}$$

The steps are as follows:

i. Initialize all the parameters.

ii. Set the counter of the training epoch, starting from zero.

iii. Apply the means of the internal attribute measurements with their corresponding standard deviation to the Type-1 non-singleton Type-2 FLS.

iv. Use the modified height defuzzification in a defuzzifier.

v. Tune the uncertain means of the antecedent membership functions and the consequents by using the Steepest Descent or Heuristic based algorithm for the error function.

vi. Calculate error function. If e = Threshold then Stop. Otherwise start a new epoch

4.2.3. Testing or Validating the Model

Validation or testing is a very important requirement to show that any newly proposed framework really works. For validating the framework, we used 30% of the available data set, divided by using the stratifying sampling approach. The validation helps to determine whether our framework training works successfully. That is, with a numerical dataset the framework can train well to achieve a lower value of error function.

The validation also helps in assessing the performance of our framework in comparison to the other existing approaches.

4.3. Evaluation of Effectiveness and Accuracy

Nowadays scientists devote much time and effort to empirical studies which aim to determine the performance of data mining solutions. Some methods may yield better results for one type of problems, while others may be suitable for different ones. That is why it is important to find the pros and cons of each of them. This may help to avoid making a mistake resulting from an application of an unsuitable algorithm. The systems which implement data mining solutions may be usable in miscellaneous areas of life, such as banking, medicine or telecommunication. Such systems are expected to support decision making in a very reliable way, as a single mistake may cause irreversible consequences or even lead to someone's death (as it may be the case in medical systems). While estimating the performance of a method one can come across different problems such as a limited sample of data, or a difficulty in evaluating a hypothesis's performance for unseen instances, or a dilemma about and finally how to use an available dataset for training and testing.

To compare the performance and accuracy of our models with other models, we used two most common quality measures, viz. Classification Rate, Root Mean Squared Error (RMSE), Receiver Operating Characteristic curve (ROC), Precision, Recall and Area Under Curve (AUC). While comparing solutions, it is also crucial to consider costs of misclassifications. Making a correct decision is very important, and so the costs should be calculated. One way to show the errors of classification is to introduce a Confusion

Matrix [51]. Such a matrix, for a Boolean problem, consists of four fields (numbers): True Positive, True Negative, False Positive and False Negative, which are needed to generate the ROC curve. They all show the dependencies between the actual classes of instances and those delivered by a model. By using these values the overall success rate can be calculated.

The comparison of different machine learning solutions may also be done with the use of the ROC (Receiver Operating Characteristic) curves that are a graphical method for evaluating classifiers. ROC graphs are very useful in organizing classifiers and visualizing their performance. Based on the ROC curves and lift charts, it is possible to introduce two parameters: Recall and Precision. They are commonly used in information retrieval [51]. Recall is understood as the ratio of retrieved relevant documents to the total number of relevant documents. Precision is defined as the ratio between the total number of documents that are retrieved and the number of documents retrieved that are relevant. The author of [68] applies the ROC and AUC curves to evaluate the performance of a data mining model. The model was used to predict the cases of the corpus luteum deficiency in women with recurrent miscarriage. The ROC and AUC curves turned out to be valuable in comparing two or more data mining methods.

4.4. Dataset Description

In our experiments, we used eight types of datasets selected from the machine learning repository at the University of California Irvine (UCI) [43]. A brief description of these datasets is in Table 4-2

Table 4-3: Description of Datasets

Dataset	Size	# of Att	Missing	Classes
Wisconsin Breast Cancer	699	9	Yes	2
Hepatitis	155	19	Yes	2
Heart Statlog	270	13	Yes	2
Pima Diabetes	768	8	Yes	2
Mamographic Mass	961	6	Yes	2
Hypothyroid	3772	30	Yes	3
Lymphography	148	19	Yes	4
Heart Disease	303	14	Yes	5

For these datasets, Size represents the number of instances/entries in a dataset, # of Attributes represents the length of one instance representing how many values is contained in one instance, Missing shows whether there is any incomplete entry, and Class represents the number of categories to be classified in a dataset. For example, the Wisconsin Breast Cancer dataset has input attributes for each cell nucleus such as radius, texture, smoothness, compactness and symmetry. The patient's problem is then diagnosed as either benign or malignant. In the Hepatitis dataset, the input attributes are steroid, fatigue, liver big, liver firm, and so on. The patient is then classified as live or dies. The Heart Statlog dataset has the input attributes of chest pain type, resting blood pressure, blood sugar and heart rate. The data is then classified as absence or presence of heart disease. Likewise, the Pima Diabetes dataset has such input attributes as glucose concentration, blood pressure, serum insulin and body mass index. The problem is then classified as diabetes present or diabetes absent. In the Mamographic dataset, the input attributes consist of age, shape, margin, density and severity. The problem is then diagnosed as benign or malign. For the datasets having more than two classes, such as Hypothyroid datasets, the input attributes include thyroxine, thyroid surgery, pregnancy, sickness and tumor. The problem is then diagnosed as normal, hyper-thyroid or hypo-thyroid. In the Lymphograpy dataset, the input attributes consist of blockage, bypass, change in lymph and dislocation. The data is then classified as normal, meta states, malign lymph and fibrosis. Likewise, the Heart disease dataset has attributes such as age, sex, cholesterol value, fasting blood sugar and so on. The five output classes are categorized as absence, typical angina, atypical angina, non-anginal pain and asymptomatic.

We conducted our experiments on Matlab R2007a. The datasets are stored in MS Excel documents and can be read directly from Matlab. All the graphs are generated by using the same Matlab R2007a.

CHAPTER 5

IMPACT OF ALGORITHMS AND PARAMETERS ON THE PERFORMANCE OF THE PROPOSED FLS BASED CLASSIFICATION FRAMEWORK

The nature of the training algorithm and parameters play a vital role in an FLS based model [30]. Many different parameters can be combined in different ways in order to build an FLS. In addition, an FLS can be trained or tuned to optimize the developed fuzzy inference engine by using different algorithms e.g. Steepest Descent approach [37] and Heuristic based approach [42]. In our proposed study, we investigated the impact of various combinations and the nature of training algorithms to identify the algorithm that can provide better results in the context of classification models.

5.1. Training Algorithms

Our proposed study concerns the back propagation algorithm in the context of adapting an FLS. In the back propagation algorithm, none of the antecedents or consequents parameters are fixed ahead of time. They can be all tuned by using the following two most common approaches in the FLS community:

i. Steepest Descent Approach

ii. Heuristic Based Approach

5.1.1. Steepest Descent Approach

Consider an FLS with singleton fuzzification, max-product composition, product implication, height defuzzification and Gaussian membership functions. It is given by the equation:

$$y(x^{(i)}) = f_s(x^{(i)}) = \frac{\sum_{l=1}^{M} \bar{y}^{l} \prod_{k=1}^{p} \exp\left[-\frac{\left(x_k^{(i)} - m_{F_k^l}\right)^2}{\left(2\sigma_{F_k^l}^2\right)}\right]}{\sum_{l=1}^{M} \prod_{k=1}^{p} \exp\left[-\frac{\left(x_k^{(i)} - m_{F_k^l}\right)^2}{\left(2\sigma_{F_k^l}^2\right)}\right]} \qquad i = 1,\ldots,N \qquad (5\text{-}1)$$

where:

M is number of rules, p is number of antecedents and N is number of data points

Given an input-output training pair ($x^{(i)} : y^{(i)}$) also known as data point, we wish to design an FLS so that the error function is minimized. The Steepest Descent approach can be applied to obtain the following recursions to update all the design parameters of this FLS in order to minimize the error function.

$$m_{F_k^l}(i+1) = m_{F_k^l}(i) - \alpha_m [f_s(x^{(i)}) - y^{(i)}][\bar{y}^l(i) - f_s(x^{(i)})] \times \frac{[x_k^{(i)} - m_{F_k^l}(i)]}{\sigma_{F_k^l}^2(i)} \phi_l(x^{(i)}) \quad (5\text{-}2)$$

$$\bar{y}^l(i+1) = \bar{y}^l(i) - \alpha_{\bar{y}} [f_s(x^{(i)}) - y^{(i)}] \phi_l(x^{(i)}) \quad (5\text{-}3)$$

$$\sigma_{F_k^l}(i+1) = \sigma_{F_k^l}(i) - \alpha_\sigma [f_s(x^{(i)}) - y^{(i)}][\bar{y}^l(i) - f_s(x^{(i)})] \times \frac{[x_k^{(i)} - m_{F_k^l}(i)]^2}{\sigma_{F_k^l}^3(i)} \phi_l(x^{(i)}) \quad (5\text{-}4)$$

Now, the back propagation algorithm can be applied as follows:

Algorithm 5-1: Back propagation algorithm for FLS

1. Initialize the parameters of all the membership functions for all the rules, $m_{F_k^l}(0)$, $\bar{y}^l(0)$ and $\sigma_{F_k^l}(0)$.
2. Set an end criterion to achieve convergence.
3. **Repeat**
   i. **for all** data points ($x^{(i)} : y^{(i)}$) $i = 1,\dots, N$
      a) Propagate the next data point through the FLS.
      b) Compute error.
      c) Update the parameters of the membership functions using (5-2), (5-3) and (5-4).

   ii. **end for** *(*end for each input-output pair*)*

   iii. Compute the root mean square relative error (RMSRE) as (5-5).

   iv. Test the end criterion. If satisfied break.

Until *(*end for each epoch*)*

$$\text{RMSRE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\frac{f_s(x^{(i)}) - y^{(i)}}{y^{(i)}}\right]^2} \tag{5-5}$$

5.1.2. Heuristic Based Approach

Like the Steepest Descent approach, the Heuristic based algorithm propagates a data point in each loop, determines the output of the FLS, and computes the output error. The main information derived from the error value is whether the contribution of a fuzzy rule to the overall output values should be increased or decreased. The updated parameters of antecedents and consequent membership functions (MFs) are then determined by using the error value based on some heuristics. The consequent of the FLS is modified by using heuristics that take defuzzification procedure into consideration. The aim is to move the output of the FLS closer to the target value. This is achieved by shifting the support of the consequent fuzzy set such that the center of the fuzzy set moves closer to the target value.

The important point to note here is that in the Steepest Descent approach, all the rules are usually modified on the basis of computed error, whenever a data point is propagated through the FLS. Moreover, different rules do not share the same MFs, i.e. the instances of an MF of a particular fuzzy set across various rules are independent of each other. But in the Heuristic based approach, various rules share the same instance of a particular fuzzy set. Finally, the rule which contributes to the output value is modified.

5.2. System Parameters

We also investigated the impact of some systems' parameters that can affect the behavior of our proposed classification framework.

The parameters that can affect the behavior of the FLS are as follows. We wish to investigate the impact of the following on the proposed classification framework:

i. Height defuzzification versus Modified Height defuzzification.

ii. Triangular membership functions versus Gaussian membership functions.

5.2.1. Height Defuzzifier

The Height defuzzifier is also known as the center average defuzzifier. Here, each rule output fuzzy set is replaced by a singleton at the point having the maximum membership in that output set. The centroid of the Type-1 set, composed of these singletons, is then calculated. The output of a Height defuzzifier is given as:

$$y_h(x) = \frac{\sum_{l=1}^{M} \bar{y}^l \mu_{B^l}(\bar{y}^l)}{\sum_{l=1}^{M} \mu_{B^l}(\bar{y}^l)} \tag{5-6}$$

where $\bar{y}^l$ is the point having the maximum membership in the l^{th} output set, and its membership grade in the l^{th} output set is $\mu_{B^l}(\bar{y}^l)$. The problem with the Height defuzzifier is that it uses only the center of the support, $\bar{y}^l$, of the consequent membership function. It does not take into consideration the shape of the consequent membership function, whether or not the membership function is narrow. If it is known, it can be interpreted as an indication of a very strong belief in that rule. If it is very broad, it can be interpreted as an indication of much less belief in that rule. The height defuzzifier produces the same result [39].

5.2.2. Modified Height Defuzzification

The Modified Height defuzzifier takes care of the deficiency of the conventional Height defuzzifier. It is very similar, except that in the Modified Height defuzzifier each $\mu_{B^l}(\bar{y}^l)$ is scaled by the inverse of the spread of the l^{th} consequent set. This is our main contribution for this study work. After solving the problem of the simple Height defuzzifier, the output is given as:

$$y_h(x) = \frac{\sum_{l=1}^{M} \bar{y}^l \mu_{B^l}(\bar{y}^l) \big/ \delta^l}{\sum_{l=1}^{M} \mu_{B^l}(\bar{y}^l) \big/ \delta^l} \tag{5-7}$$

where δ^l is a measure of the spread of the l^{th} consequent set, $\bar{y}^l$ and $\mu_{B^l}(\bar{y}^l)$ have the same meaning as in (5-6). Now, consider an FLS with singleton fuzzification, max-product composition, product implication, Modified Height defuzzification and Gaussian membership functions. It is given by the equation:

$$y(x^{(i)}) = f_s(x^{(i)}) = \frac{\sum_{l=1}^{M} \bar{y}^l \prod_{k=1}^{p} \exp\left[-\frac{\left(x_k^{(i)} - m_{F_k^l}\right)^2}{\left(2\sigma_{F_k^l}^2\right)}\right] \Big/ \delta^l}{\sum_{l=1}^{M} \prod_{k=1}^{p} \exp\left[-\frac{\left(x_k^{(i)} - m_{F_k^l}\right)^2}{\left(2\sigma_{F_k^l}^2\right)}\right] \Big/ \delta^l} \quad i = 1,\dots,N \tag{5-8}$$

The training algorithm is then applied to obtain the following recursions, in order to update all the design parameters of this FLS, and thus to minimize the error function.

$$m_{F_k^l}(i+1) = m_{F_k^l}(i) - 2\alpha_m \left(\frac{f_s(x^{(i)}) - y^{(i)}}{\sum_{l=1}^{M} z^l \big/ \delta^l} \right) [\overline{y}^l(i) - f_s(x^{(i)})] \times \frac{[x_k^{(i)} - m_{F_k^l}(i)]}{\sigma_{F_k^l}^2(i)} \frac{z^l}{\delta^l} \quad (5\text{-}9)$$

$$\overline{y}^l(i+1) = \overline{y}^l(i) - \alpha_{\overline{y}}[f_s(x^{(i)}) - y^{(i)}] \frac{z^l / \delta^l}{\sum_{l=1}^{M} z^l / \delta^l} \quad (5\text{-}10)$$

and

$$\sigma_{F_k^l}(i+1) = \sigma_{F_k^l}(i) - 2\alpha_\sigma \left(\frac{f_s(x^{(i)}) - y^{(i)}}{\sum_{l=1}^{M} z^l \big/ \delta^l} \right) [\overline{y}^l(i) - f_s(x^{(i)})] \times \frac{[x_k^{(i)} - m_{F_k^l}(i)]^2}{\sigma_{F_k^l}^3(i)} \frac{z^l}{\delta^l} \quad (5\text{-}11)$$

$$\sigma^l(i+1) = \delta^l(i) + \alpha_\delta [f_s(x^{(i)}) - y^{(i)}][\overline{y}^l(i) - f_s(x^{(i)})] \frac{z^l}{\delta^{l2}(i)} \left(\frac{1}{\sum_{l=1}^{M} z^l \big/ \delta^l} \right) \quad (5\text{-}12)$$

where

$$z^l = \prod_{k=1}^{p} \exp\left[- \frac{\left(x_k^{(i)} - m_{F_k^l}\right)^2}{\left(2\sigma_{F_k^l}^2\right)} \right] \quad (5\text{-}13)$$

Now, the back propagation algorithm can be applied as explained above. Thus the Modified Height defuzzification in a defuzzifier first combines the output sets corresponding to the highest membership value, to obtain a single output set. Then it finds a crisp number that is representative of this combined output set.

5.2.3. Gaussian and Triangular Membership Functions

Besides other MFs, Gaussian and triangular MFs can be used to define antecedents or consequent fuzzy sets. In this research, we investigated the impact of these MFs on the adaptive FLS classification framework by using the Steepest Descent approach.

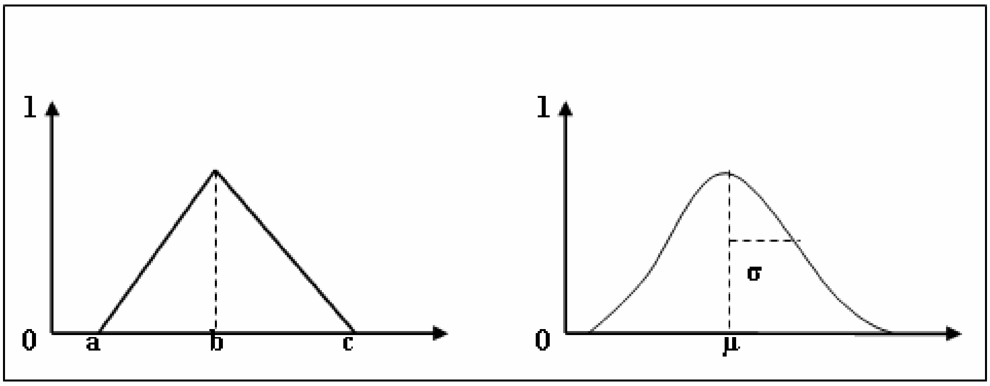

Figure 5-11: Gaussian and triangular Membership Functions

A triangular MF is a three point function, defined by minimum (a), maximum (c) and modal (b) values i.e., TMF (a,b,c), where ($a \le b \le c$). The membership value at any input x can be calculated as follows:

$$\mu_{a,b,c}(x) = \begin{cases} \dfrac{x-a}{b-a} & a \le x \le b \\ \dfrac{c-x}{c-b} & b \le x \le c \\ 0 & otherwise \end{cases} \tag{5-14}$$

Now, consider a FLS with singleton fuzzification, max-product composition, product implication, Height Defuzzification and Triangular MFs. It is given by the equation:

$$y(x^{(i)}) = f_s(x^{(i)}) = \frac{\sum_{l=1}^{M} \bar{y}^l \left(\prod_{k=1}^{p} \frac{\left(x_k^{(i)} - a_{F_k^l}\right)}{\left(b_{F_k^l} - a_{F_k^l}\right)} \right)}{\sum_{l=1}^{M} \left(\prod_{k=1}^{p} \frac{\left(x_k^{(i)} - a_{F_k^l}\right)}{\left(b_{F_k^l} - a_{F_k^l}\right)} \right)} \tag{5-15}$$

The training algorithm is then applied to obtain the following recursions to update all the design parameters of this FLS in order to minimize error function.

When $a \leq x \leq b$

$$a_{F_k^l}(i+1) = a_{F_k^l}(i) - \alpha \frac{[f_s(x^{(i)}) - y^{(i)}]}{b_{F_k^l}(i)} \frac{z^l}{b_{F_k^l}(i) - a_{F_k^l}(i)} \left(1 - \frac{b_{F_k^l}(i) - a_{F_k^l}(i)}{x_k^{(i)} - a_{F_k^l}(i)} \right) [\bar{y}^l(i) - f_s(x^{(i)})] \tag{5-16}$$

$$b_{F_k^l}(i+1) = b_{F_k^l}(i) + \alpha \frac{[f_s(x^{(i)}) - y^{(i)}]}{b_{F_k^l}(i)} z^l \left(\frac{1}{b_{F_k^l}(i) - a_{F_k^l}(i)} \right) [\bar{y}^l(i) - f_s(x^{(i)})] \tag{5-17}$$

when $b \leq x \leq c$

$$c_{F_k^l}(i+1) = c_{F_k^l}(i) - \alpha \frac{[f_s(x^{(i)}) - y^{(i)}]}{b_{F_k^l}(i)} \frac{z^l}{c_{F_k^l}(i) - b_{F_k^l}(i)} \left(\frac{c_{F_k^l}(i) - b_{F_k^l}(i)}{c_{F_k^l}(i) - x_k^{(i)}} - 1 \right) [\bar{y}^l(i) - f_s(x^{(i)})] \tag{5-18}$$

$$b_{F_k^l}(i+1) = b_{F_k^l}(i) - \alpha \frac{[f_s(x^{(i)}) - y^{(i)}]}{b_{F_k^l}(i)} \frac{z^l}{c_{F_k^l} - b_{F_k^l}(i)} [\bar{y}^l(i) - f_s(x^{(i)})] \tag{5-19}$$

Where

$$z^l = \prod_{k=1}^{p} \frac{\left(x_k^{(i)} - a_{F_k^l}\right)}{\left(b_{F_k^l} - a_{F_k^l}\right)} \tag{5-20}$$

Now, the back propagation algorithm can be applied as explained above. The only difference is that, while modifying the parameters, we must make sure that the condition ($a \le b \le c$) holds.

CHAPTER 6

EXPERIMENTS AND RESULTS

6.1 Experiment 1: Uncertainty Handling in the Fuzzy Logic Based Classification Framework

In this experiment, we used eight different datasets. These datasets cover all the variations, such as small number of instances, large number of instances, number of attributes and so on, which are required for good experiments. As this is our first experiment, we excluded all the missing values in the datasets and we assumed that all the datasets are in original form and noise free.

6.1.1. Training and Testing

We conducted this experiment with eight datasets. In each experiment, we employed the Type-2 FLS algorithm as described above, and we compared the results with the corresponding Type-1 FLS. It is important to note here that the performance of Type-1 and Type-2 training algorithms cannot be compared on the same step size, because the Type-1 training algorithm requires a smaller value for the step size than Type-2. Thus, in all the experiments, we provided the suitable step size values to both the FLS types, so that the parameters are converged on the same number of epochs. Upon the completion of each training epoch, the RMSRE values on both training and testing datasets were computed.

6.1.2. Results and Discussion

It is evident from Table 6-1 and 6-2 that none of the Type-1 or Type-2 fuzzy logic systems has preference over one another. For some datasets Type-1 has shown better performance, whereas for others Type-2 has got some edge over the Type-1 FLS based classification framework. The reason is that, for small and simple datasets, Type-1 has better generalization property as it keeps the overall process simple. However if we look at the RMSRE and ROC graphs, we can say that Type-2 FLS outclassed the Type-1 FLS. Therefore our intuition for handling the uncertainty in a classification framework by using Type-2 FLS is justified.

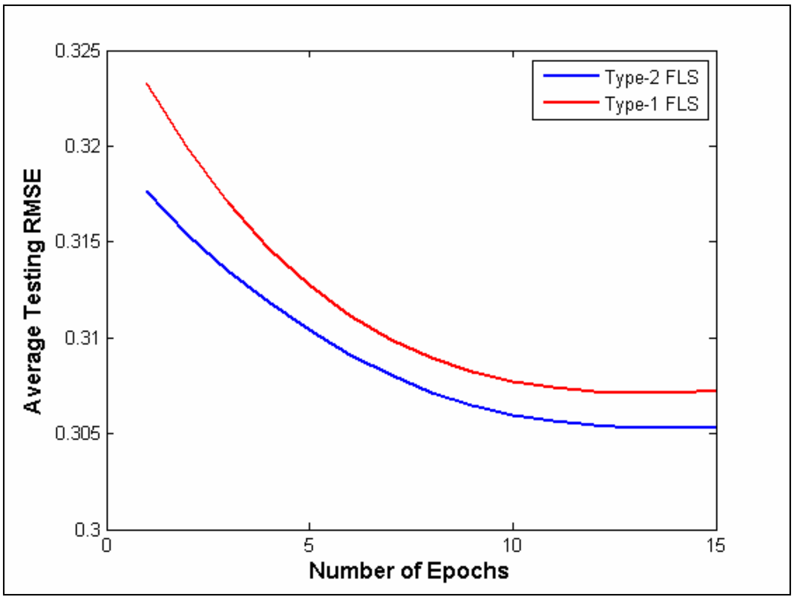

Figure 6-12: Average RMSRE graph during testing of Type-1 and Type-2 FLS on Wisconsin Breast Cancer Dataset

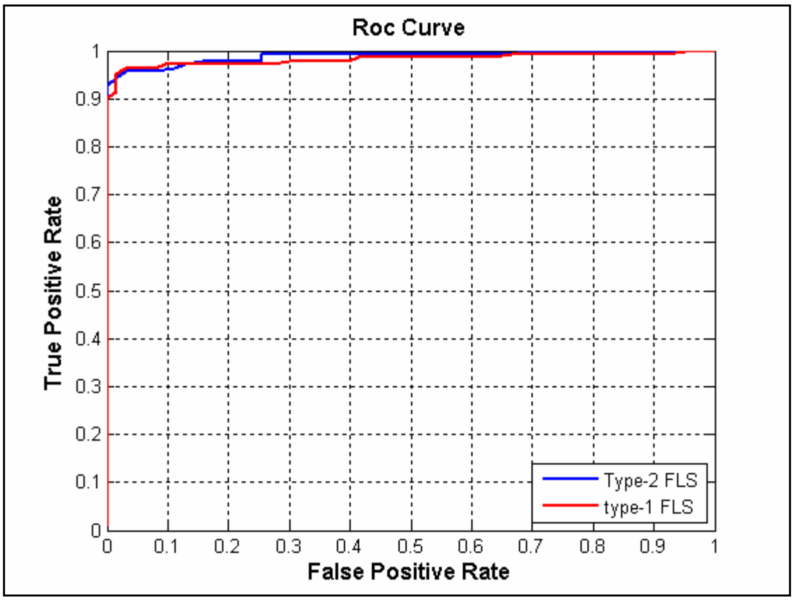

Figure 6-13: ROC Curve during testing of Type-1 and Type-2 FLS on Wisconsin Breast Cancer Dataset

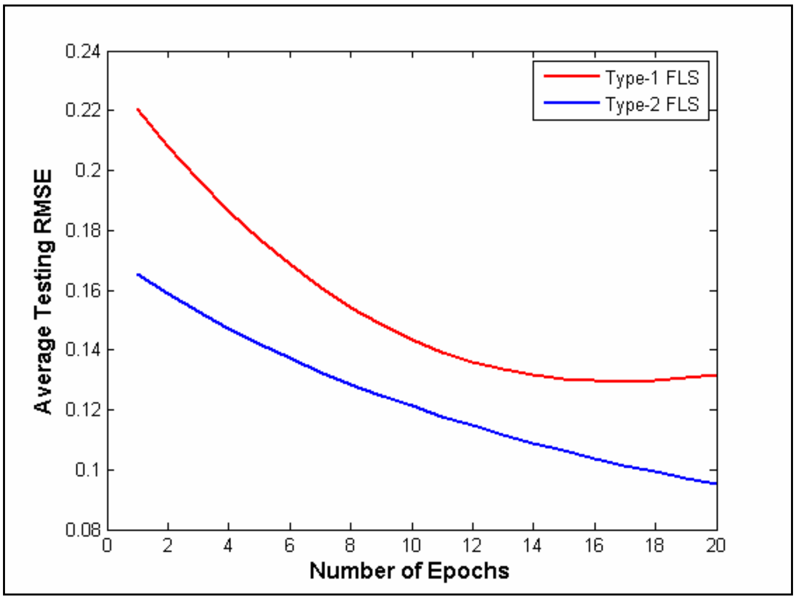

Figure 6-14: Average RMSRE graph during testing of Type-1 and Type-2 FLS on Pima Diabetes Dataset

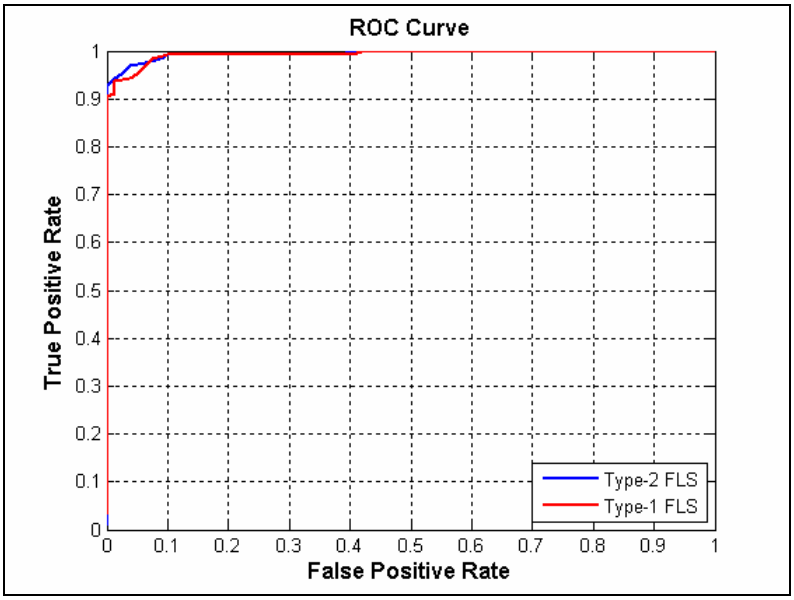

Figure 6-15: ROC Curve during testing of Type-1 and Type-2 FLS on Pima Diabetes Dataset

Table 6-4: Summary of Classification Accuracy of eight different datasets on training and testing data

Dataset	Type-1 Fuzzy Logic System		Type-2 Fuzzy Logic System	
	Classification Accuracy on Train Data	**Classification Accuracy on Test Data**	**Classification Accuracy on Train Data**	**Classification Accuracy on Test Data**
Wisconsin Breast Cancer	98.66 %	97.90 %	99.68 %	99.28 %
Hepatitis	91.30 %	89.95 %	91.80 %	90.15 %
Heart Statlog	90.83 %	90.45 %	90.10 %	89. 67 %
Pima Diabetes	81.55 %	79.89 %	83.65 %	83.19 %
Mamographic Mass	88.42 %	87.25 %	91.72 %	90.56 %
Hypothyroid	96.50 %	94.25 %	97.50 %	95.30 %
Lymphography	89.35 %	87.60 %	90.35 %	88.67 %
Heart Disease	87.13 %	86.48 %	89.13 %	88.24 %

Table 6-5: Summary of Precision, Recall and Area under Curve (AUC) on training and testing data

Dataset	Type-1 Fuzzy Logic System			Type-2 Fuzzy Logic System		
	Precision (%)	Recall (%)	AUC	Precision (%)	Recall (%)	AUC
Wisconsin Breast Cancer	97.6 / 98.4	98.2 / 98.6	0.93	98.3 / 98.5	99.3 / 98.9	0.94
Hepatitis	88.5 / 89.1	90.1 / 90.7	0.81	89.4 / 91.1	90.3 / 90.5	0.83
Heart Statlog	89.3 / 90.4	89.8 / 90.6	0.83	89.3 / 89.7	89.8 / 90.6	0.83
Pima Diabetes	79.6 / 80.8	80.2 / 80.3	0.78	83.6 / 82.8	84.2 / 83.3	0.80
Mamographic Mass	87.9 / 86.6	88.3 / 87.8	0.85	92.2 / 91.6	91.3 / 92.8	0.88
Hypothyroid	93.2 / 94.8	94.5 / 94.2	0.92	95.3 / 94.8	95.2 / 96.3	0.94
Lymphography	85.4 / 86.7	86.9 / 87.1	0.84	87.4 / 88.6	86.7 / 87.4	0.86
Heart Disease	86.6 / 85.1	85.8 / 86.3	0.85	88.2 / 87.9	88.9 / 88.3	0.87

6.2 Experiment 2: Uncertainty Handling in Fuzzy Logic Based Classification Framework when Dataset is taken as a Non-Singleton Input

For the dataset to be taken as non-singleton, we need to generate Gaussian membership functions that represent size as non-singleton input.

Algorithm 6-1: Generate datasets that represent uncertainty by using Gaussian membership function.

1. Take a mean value of all the values in each of the input attributes.
2. For each mean size, compute the random standard deviation so that the Gaussian membership function spans R to Q percent of the mean value on both sides of the mean.
3. For each mean value, generate noise up to at most R percent, and add that noise to the corresponding input values.
4. Partition the generated dataset into training and testing datasets. The training dataset consists of 70 percent of the entire dataset, whereas the remaining 30 percent is left for testing.

6.2.1. Training and Testing

We conducted this experiment in two parts with eight different datasets. In the first part these datasets were assumed to have at most 15 percent uncertainty in their respective values, and in the second part the uncertainty was scaled up to 25 percent. In all the experiments, we have generated Gaussian input noise such that they spanned around 15 and 25 percent of the mean value on both sides of the mean. The rest of the details are the same as we have discussed in experiment 1.

6.2.2. Results and Discussion

It is evident from Tables 6-3 and 6-4 that Type-2 FLS is better than Type-1 FLS on the basis of Classification accuracy, Precision, Recall and Area Under Curve. This is because Type-2 is more immune to noise and can perform better in presence of uncertainties. Here in all the experiments Type-2 FLS was better than Type-1. The RMSRE and ROC graphs show that Type-2 FLS outclassed the Type-1 FLS. Therefore our intuition for handling uncertainty in classification frameworks by using Type-2 FLS is justified.

While investigating the results of Tables 6-5 and 6-6 when the uncertainty scale is increased to 25 percent, we can see that Type-2 is still performing better than Type-1. Also the results of Type-1 at 25 percent uncertainty show a drastic decrease in prediction accuracy as compared to 15 percent uncertainty, while Type-2 fuzzy is more immune to noise and uncertainties.

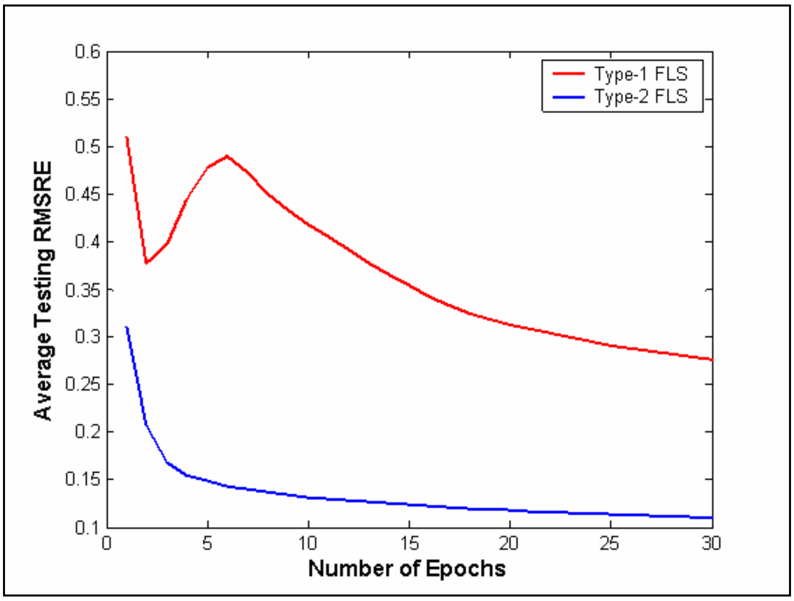

Figure 6-16: Average RMSRE graph during testing of Type-1 and Type-2 FLS on Wisconsin Breast Cancer Dataset as a non-singleton input

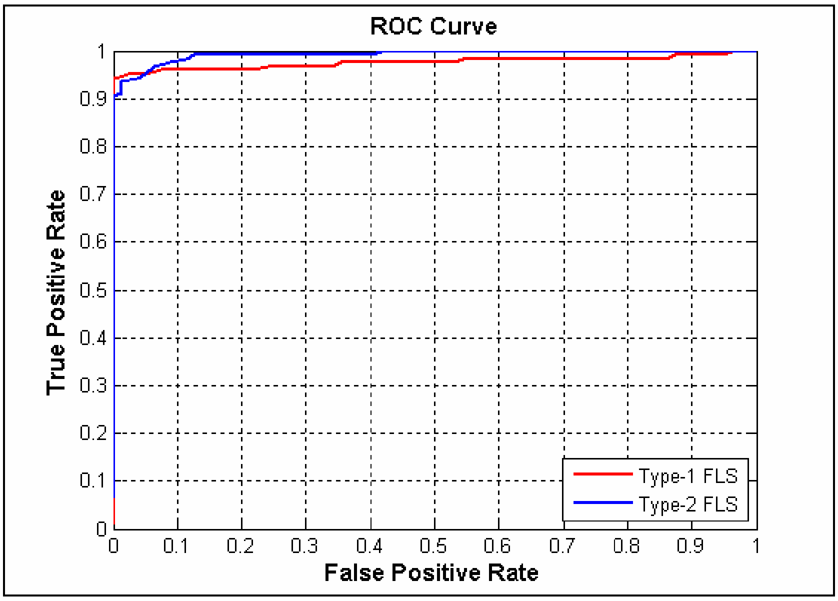

Figure 6-17: ROC Curve during testing of Type-1 and Type-2 FLS on Wisconsin Breast Cancer Dataset

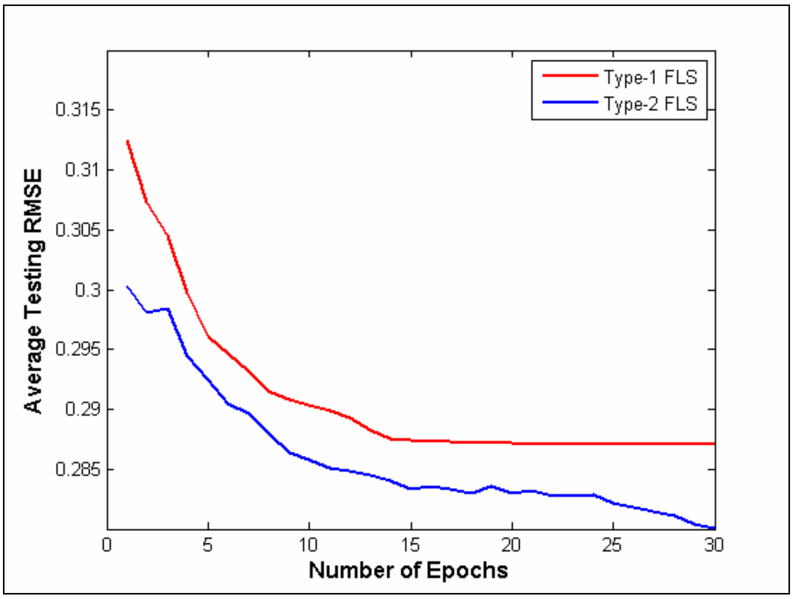

Figure 6-18: Average RMSRE graph during testing of Type-1 and Type-2 FLS on Pima Diabetes Dataset

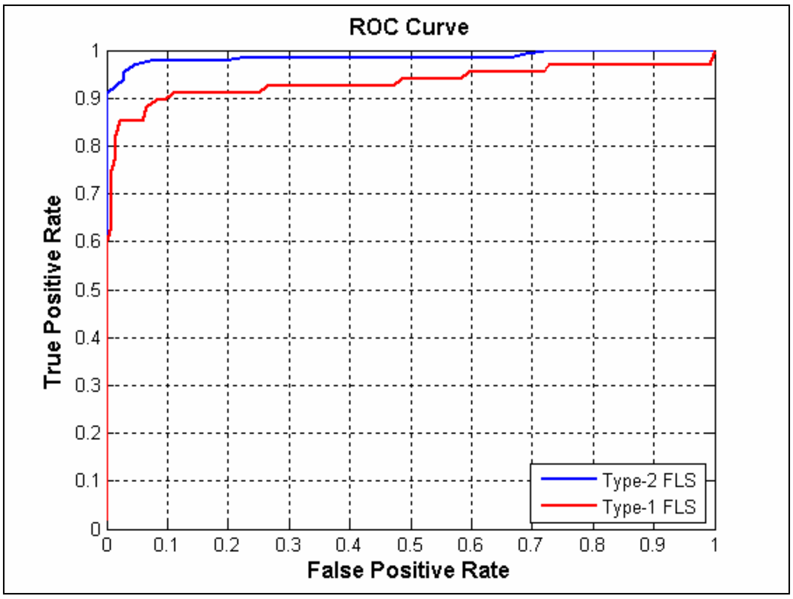

Figure 6-19: ROC Curve during testing of Type-1 and Type-2 FLS on Pima Diabetes Dataset

Table 6-6: Summary of Classification Accuracy for eight different datasets on training and testing data when uncertainty is scaled to 15 percent

Dataset	Type-1 Fuzzy Logic System		Type-2 Fuzzy Logic System	
	Classification Accuracy on Train Data	**Classification Accuracy on Test Data**	**Classification Accuracy on Train Data**	**Classification Accuracy on Test Data**
Wisconsin Breast Cancer	92.06 %	91.50 %	96.63 %	96.26 %
Hepatitis	87.65 %	86.15 %	89.84 %	88.45 %
Heart Statlog	88.13 %	87.55 %	90.92 %	90. 05 %
Pima Diabetes	76.58 %	76.15 %	80.45 %	79.85 %
Mamographic Mass	84.18 %	83.54 %	89.76 %	89.62 %
Hypothyroid	94.30 %	93.28 %	96.25 %	95.30 %
Lymphography	87.56 %	85.63 %	89.35 %	86.24 %
Heart Disease	85.33 %	83.18 %	87.18 %	85.89 %

Table 6-7: Summary of Precision, Recall and Area under Curve (AUC) on training and testing data when uncertainty is scaled to 15 percent

Dataset	Type-1 Fuzzy Logic System			Type-2 Fuzzy Logic System		
	Precision (%)	Recall (%)	AUC	Precision (%)	Recall (%)	AUC
Wisconsin Breast Cancer	92.5 / 91.2	91.8 / 92.2	0.89	96.4 / 96.3	96.6 / 97.1	0.93
Hepatitis	85.7 / 86.9	85.3 / 85.1	0.81	89.4 / 89.1	88.3 / 89.5	0.84
Heart Statlog	87.3 / 86.8	87.4 / 87.5	0.83	90.3 / 89.7	90.8 / 90.4	0.88
Pima Diabetes	77.4 / 76.9	76.2 / 77.3	0.72	80.6 / 79.8	80.6 / 79.3	0.74
Mamographic Mass	83.1 / 83.4	84.3 / 83.5	0.79	89.2 / 88.6	89.3 / 89.7	0.87
Hypothyroid	92.1 / 93.2	94.5 / 94.2	0.90	94.3 / 93.5	94.2 / 95.3	0.93
Lymphography	84.2 / 85.3	85.9 / 86.2	0.82	86.4 / 87.6	85.7 / 86.4	0.85
Heart Disease	85.6 / 86.6	84.4 / 85.3	0.83	87.2 / 86.7	87.9 / 86.9	0.86

Table 6-8: Summary of Classification Accuracy for eight different datasets on training and testing data when uncertainty is scaled to 25 percent

Dataset	Type-1 Fuzzy Logic System		Type-2 Fuzzy Logic System	
	Classification Accuracy on Train Data	**Classification Accuracy on Test Data**	**Classification Accuracy on Train Data**	**Classification Accuracy on Test Data**
Wisconsin Breast Cancer	89.23 %	88.90 %	95.30 %	95.38 %
Hepatitis	84.45 %	83.68 %	87.12 %	86.45 %
Heart Statlog	84.89 %	84.25 %	89.56 %	88. 90 %
Pima Diabetes	72.18 %	71.50 %	78.35 %	78.36 %
Mamographic Mass	80.20 %	79.85 %	88.46 %	87.55 %
Hypothyroid	91.35 %	90.29 %	94.75 %	93.10 %
Lymphography	84.62 %	82.53 %	88.35 %	86.24 %
Heart Disease	82.38 %	80.18 %	86.48 %	85.59 %

Table 6-9: Summary of Precision, Recall and Area under Curve (AUC) on training and testing data when uncertainty is scaled to 25 percent

Dataset	Type-1 Fuzzy Logic System			Type-2 Fuzzy Logic System		
	Precision (%)	Recall (%)	AUC	Precision (%)	Recall (%)	AUC
Wisconsin Breast Cancer	88.7 / 87.4	88.2 / 86.9	0.83	95.3 / 94.7	95.2 / 96.5	0.91
Hepatitis	83.8 / 84.5	82.3 / 84.1	0.75	86.9 / 87.8	86.3 / 87.2	0.82
Heart Statlog	83.8 / 84.1	83.6 / 83.9	0.76	89.4 / 88.6	89.2 / 89.1	0.86
Pima Diabetes	74.6 / 73.6	74.2 / 73.2	0.67	78.6 / 79.5	77.9 / 78.6	0.73
Mamographic Mass	80.1 / 80.2	79.5 / 79.1	0.72	88.4 / 87.2	87.5 / 86.3	0.86
Hypothyroid	90.1 / 89.7	89.5 / 90.3	0.87	93.4 / 92.9	93.2 / 94.3	0.91
Lymphography	82.2 / 81.3	82.9 / 82.2	0.80	85.4 / 85.6	86.7 / 85.4	0.84
Heart Disease	83.6 / 84.2	82.8 / 83.5	0.81	86.2 / 85.5	86.9 / 85.9	0.85

6.3 Experiment 3: Uncertainty Handling in Fuzzy Logic System Based Classification Framework due to Laziness / Ignorance Uncertainty

The logic of imposing laziness and ignorance in the dataset is very simple. We introduced some missing values, and we slightly changed some input values, so as to make the dataset uncertain. The rest of the details are the same as in experiment 1.

6.3.1. Training and Testing

We conducted this experiment with eight datasets. Each dataset was divided into 70% of training and 30% of testing data points.

6.3.2. Results and Discussion

It is evident from Table 6-7 and Table 6-8 that Type-2 FLS is preferable to Type-1 when dealing with laziness/ignorance in the input data. The same conclusion can be drawn from RMSRE ad ROC graphs, where Type-2 FLS has outclassed the Type-1 FLS. Therefore our intuition for handling laziness/ignorance in classification frameworks by using Type-2 FLS is justified.

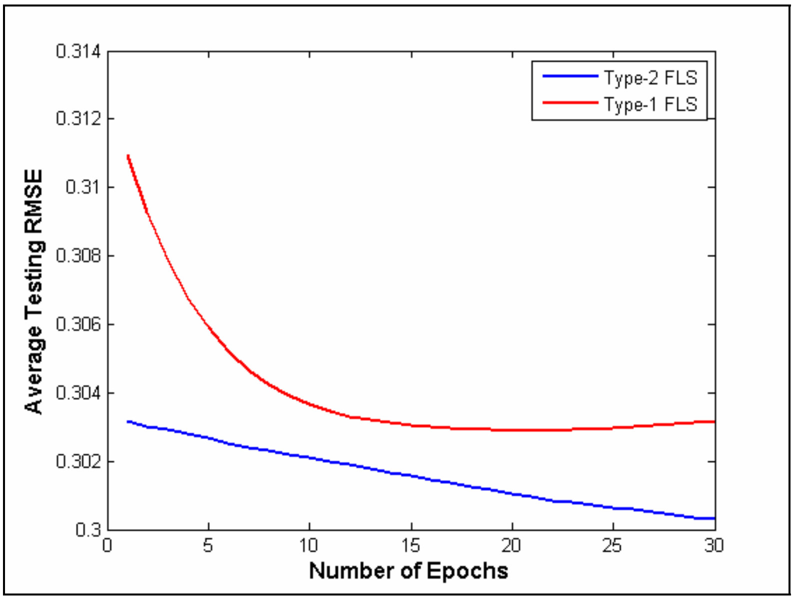

Figure 6-20: Average RMSRE graph during testing of Type-1 and Type-2 FLS on Wisconsin Breast Cancer Dataset

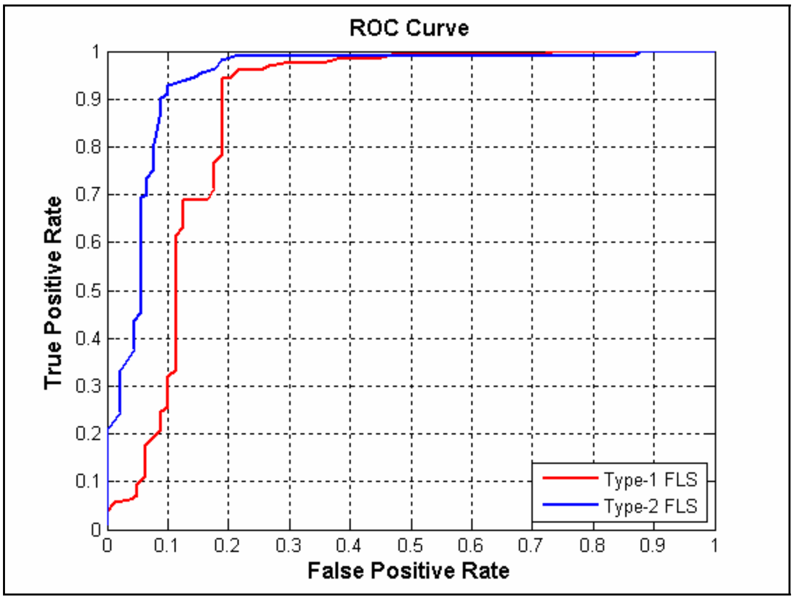

Figure 6-21: ROC Curve during testing of Type-1 and Type-2 FLS on Wisconsin Breast Cancer Dataset

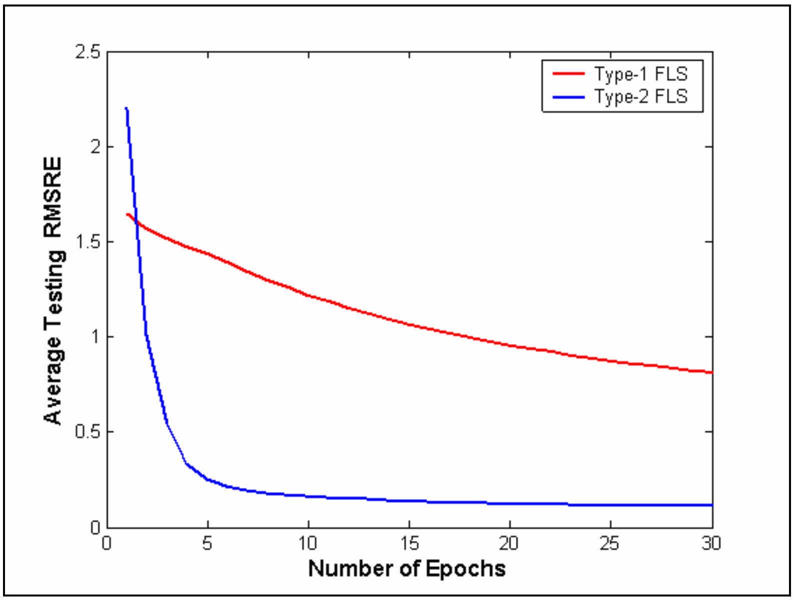

Figure 6-22: Average RMSRE graph during testing of Type-1 and Type-2 FLS on Pima Diabetes Dataset

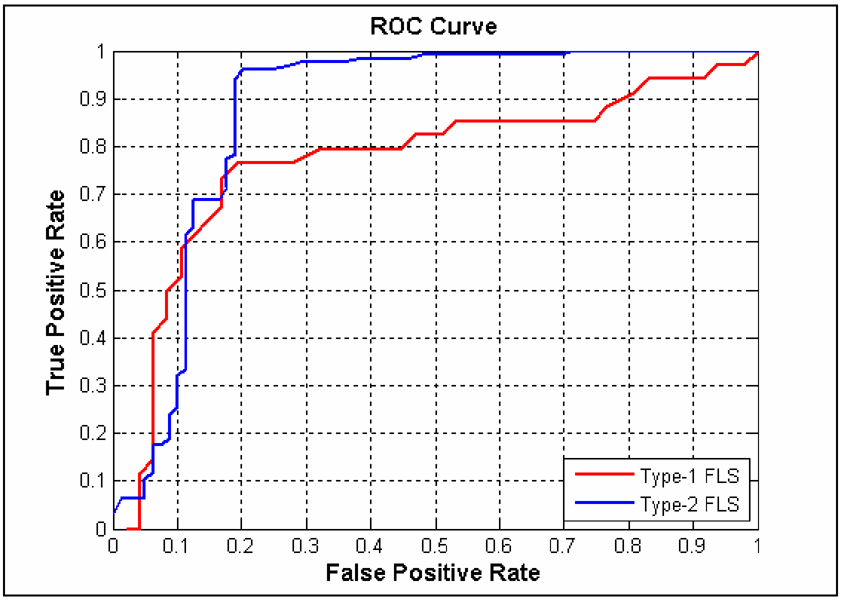

Figure 6-23: ROC Curve during testing of Type-1 and Type-2 FLS on Pima Diabetes Dataset

Table 6-10: Summary of Classification Accuracy for eight different datasets on training and testing data

Dataset	Type-1 Fuzzy Logic System		Type-2 Fuzzy Logic System	
	Classification Accuracy on Train Data	Classification Accuracy on Test Data	Classification Accuracy on Train Data	Classification Accuracy on Test Data
Wisconsin Breast Cancer	85.34 %	85.20 %	91.25 %	91.16 %
Hepatitis	80.67 %	79.55 %	87.36 %	86.40 %
Heart Statlog	82.14 %	81.35 %	88.72 %	88. 25 %
Pima Diabetes	70.82 %	70.15 %	77.90 %	76.80 %
Mamographic Mass	79.48 %	79.10 %	84.46 %	84.72 %
Hypothyroid	89.30 %	88.18 %	93.25 %	92.80 %
Lymphography	82.67 %	80.63 %	86.35 %	85.28 %
Heart Disease	81.43 %	80.18 %	85.45 %	85.95 %

Table 6-11: Summary of Precision, Recall and Area under Curve (AUC) on training and testing data

Dataset	Type-1 Fuzzy Logic System			Type-2 Fuzzy Logic System		
	Precision (%)	Recall (%)	AUC	Precision (%)	Recall (%)	AUC
Wisconsin Breast Cancer	86.2 / 85.4	85.4 / 87.2	0.81	91.2 / 91.6	92.3 / 91.5	0.88
Hepatitis	80.6 / 79.4	81.2 / 80.7	0.77	86.4 / 87.4	87.3 / 86.7	0.83
Heart Statlog	81.8 / 81.9	82.2 / 82.6	0.79	87.9 / 88.4	88.6 / 87.5	0.84
Pima Diabetes	70.8 / 70.9	71.2 / 70.3	0.69	77.2 / 77.8	76.4 / 76.3	0.73
Mamographic Mass	78.2 / 79.6	79.1 / 80.2	0.75	84.8 / 84.2	84.3 / 83.9	0.79
Hypothyroid	88.1 / 87.2	88.5 / 87.2	0.86	92.3 / 93.5	92.2 / 93.3	0.90
Lymphography	80.7 / 81.3	81.9 / 80.2	0.78	86.4 / 85.6	86.7 / 85.4	0.84
Heart Disease	80.6 / 80.2	80.6 / 81.3	0.78	85.2 / 84.9	85.9 / 84.2	0.84

6.4 Experiment 4: Investing the Effects of Parameters on Proposed FLS Classification Framework

We conducted this experiment with eight different datasets for each study. Each dataset consisted of 70% for training and 30% for testing data points.

6.4.1. Comparing Height and Modified Height Defuzzification

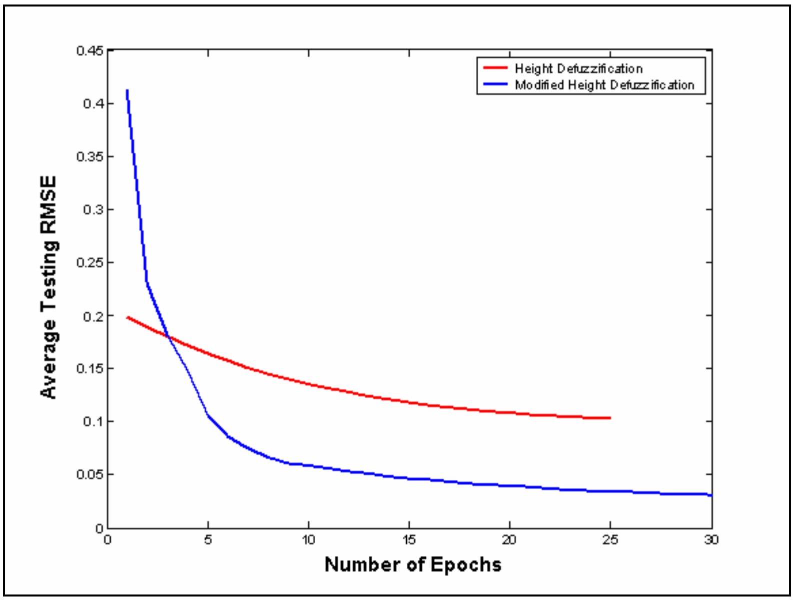

Figure 6-24: Average RMSRE graph during testing of FLS on Wisconsin Breast Cancer Dataset with Height and Modified Height Defuzzification

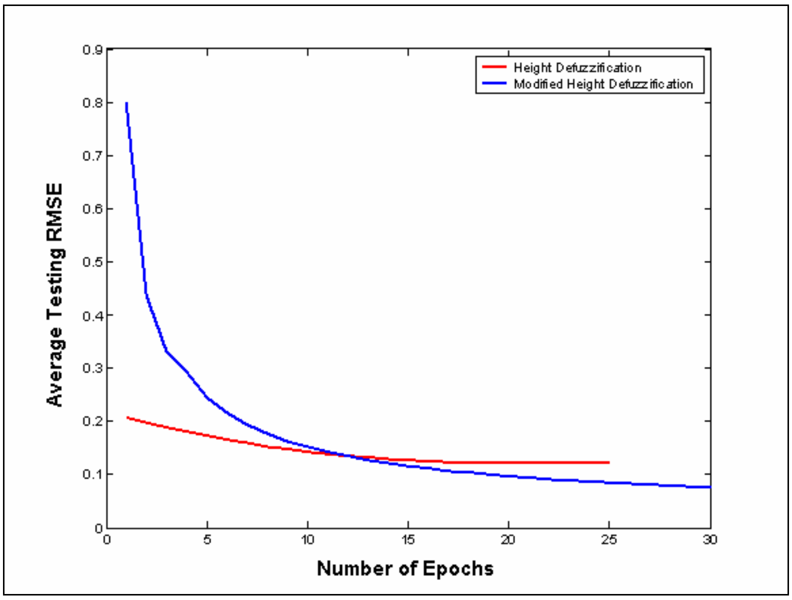

Figure 6-25: Average RMSRE graph during testing of FLS on Pima Diabetes Dataset with Height and Modified Height Defuzzification

Table 6-12: Summary of Classification Accuracy using Height and Modified Height Defuzzification methods on eight different datasets

Dataset	Height Defuzzification		Modified Height Defuzzification	
	Classification Accuracy on Train Data	Classification Accuracy on Test Data	Classification Accuracy on Train Data	Classification Accuracy on Test Data
Wisconsin Breast Cancer	98.66 %	97.90 %	98.80 %	98.45 %
Hepatitis	91.30 %	89.95 %	93.34 %	92.48 %
Heart Statlog	90.83 %	90.45 %	92.12 %	91. 95 %
Pima Diabetes	81.55 %	79.89 %	81.90 %	81.20 %
Mamographic Mass	88.42 %	87.25 %	89.36 %	88.74 %
Hypothyroid	96.50 %	94.25 %	97.50 %	95.30 %
Lymphography	89.35 %	87.60 %	90.35 %	88.67 %
Heart Disease	87.13 %	86.48 %	89.13 %	88.24 %

6.4.2. Comparing Gaussian and Triangular Membership Functions

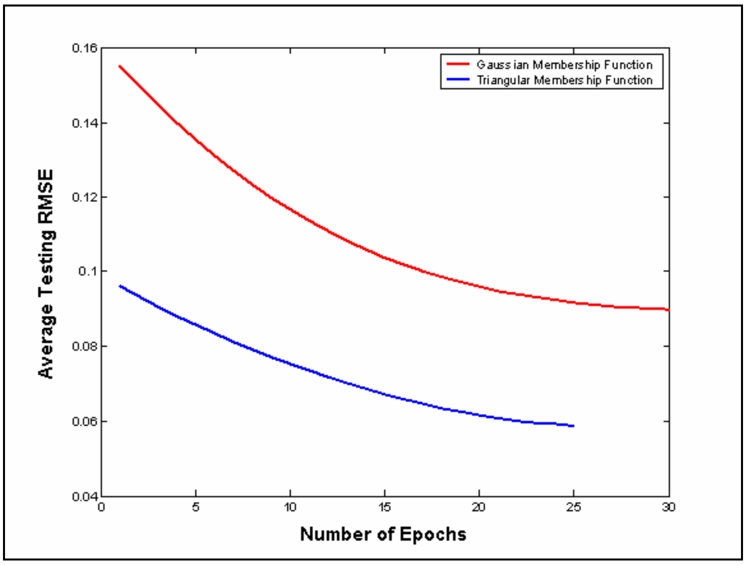

Figure 6-26: Average RMSRE graph during testing of FLS on Wisconsin Breast Cancer Dataset with Gaussian and Triangular Membership Functions

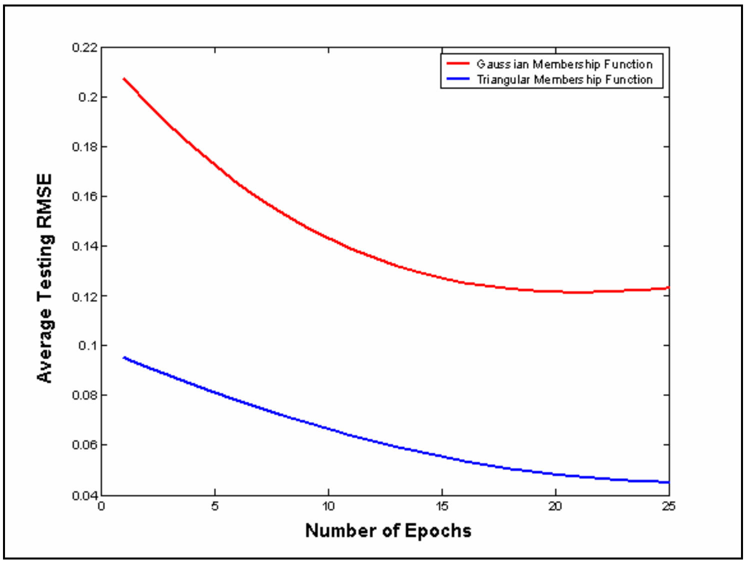

Figure 6-27: Average RMSRE graph during testing of FLS on Pima Diabetes Dataset with Gaussian and Triangular Membership Functions

Table 6-13: Summary of Classification Accuracy using Gaussian and Triangular Membership Functions

Dataset	Gaussian Membership Function		Triangular Membership Function	
	Classification Accuracy on Train Data	Classification Accuracy on Test Data	Classification Accuracy on Train Data	Classification Accuracy on Test Data
Wisconsin Breast Cancer	98.26 %	96.90 %	98.90 %	97.45 %
Hepatitis	91.30 %	89.95 %	91.90 %	90.18 %
Heart Statlog	90.83 %	90.45 %	91. 28 %	91.05 %
Pima Diabetes	81.55 %	79.89 %	83.45 %	82.70 %
Mamographic Mass	88.42 %	87.25 %	89.60 %	89.15 %
Hypothyroid	95.50 %	94.25 %	96.50 %	96.20 %
Lymphography	88.35 %	87.60 %	89.35 %	88.67 %
Heart Disease	87.13 %	86.48 %	89.13 %	88.24 %

6.4.3. Results and Discussion

After analyzing Tables 6-9, 6-10 and the corresponding RMSRE graphs, we have come to the following conclusion:

i. Modified Height defuzzification has superiority over Height defuzzification. The reason is that, unlike Height defuzzification, Modified Height defuzzification method considers the spread of consequent membership function. This spread portrays a better picture of the contribution of a particular fuzzified input to the corresponding consequent.

ii. Triangular membership function has better classification accuracy than Gaussian membership function in the FLS based classification framework. This is because triangular MF provides a better local control over the shape, i.e. spread, of the membership function. Whether an input is fuzzified to the left or to the right of the mean for Gaussian MF, a modification to the spread of MF is applied equally to both the sides of the mean. On the other hand in the case of triangular MF, a modification to the spread of MF is applied only to the side which was fuzzified with the input. Although this modification disturbs the other side to some extent, it is still far less than in Gaussian MF.

6.5 Experiment 5: Effect of Training Algorithm on Proposed Classification Framework

We conducted this experiment with eight datasets for each study. Each dataset consist of 70% for training and 30% for testing data points.

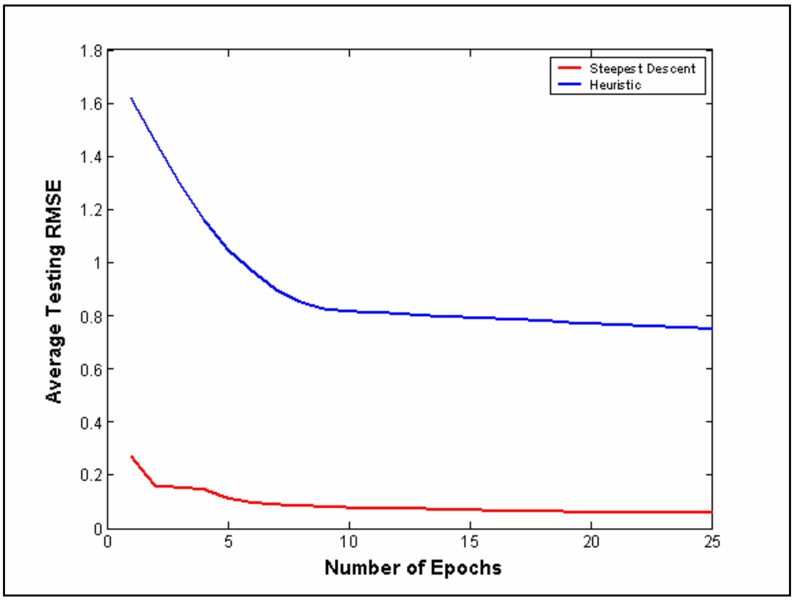

Figure 6-28: Average RMSRE graph during testing of FLS on Wisconsin Breast Cancer Dataset with Steepest Descent and Heuristics approaches

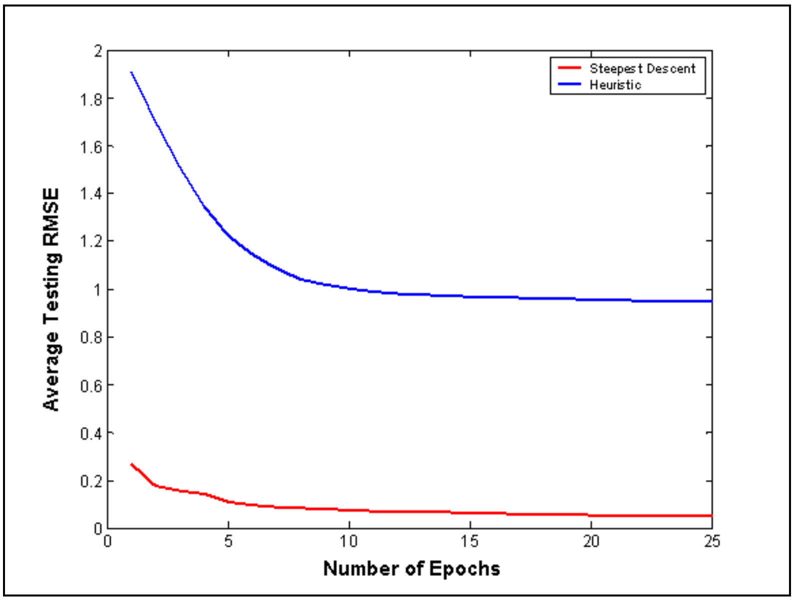

Figure 6-29: Average RMSRE graph during testing of FLS on Pima Diabetes Dataset with Steepest Descent and Heuristics approaches

Table 6-14: Summary of Classification Accuracy using Steepest Descent and Heuristic based approaches

Dataset	Steepest Descent		Heuristic	
	Classification Accuracy on Train Data	Classification Accuracy on Test Data	Classification Accuracy on Train Data	Classification Accuracy on Test Data
Wisconsin Breast Cancer	98.66 %	97.90 %	95.95 %	95.20 %
Hepatitis	91.30 %	89.95 %	89.74 %	89.15 %
Heart Statlog	90.83 %	90.45 %	89. 48 %	88.94 %
Pima Diabetes	81.55 %	79.89 %	79.76 %	78.50 %
Mamographic Mass	88.42 %	87.25 %	86.60 %	85.24 %
Hypothyroid	96.50 %	94.25 %	95.20 %	94.30 %
Lymphography	89.35 %	87.60 %	88.35 %	86.87 %
Heart Disease	87.13 %	86.48 %	85.13 %	84.94 %

6.5.1. Results and Discussion

It is evident from Table 6-11 and RMSRE graphs that the performance of the training algorithm is better with the Steepest Descent approach than with the Heuristic based approach in the FLS classification framework. We believe the reason is that the Steepest Descent Approach allows the separate training of each rule. Each rule gets its own copy of membership functions, and the modification to the membership functions in a rule is applied locally. In the heuristic based approach, by contrast, the membership functions are modified rather than the rules. If a modification to any rule is needed, it is applied to the same copy of the membership functions. Thus Steepest Descent training converges better than the heuristic based approach, and it provides better results during activation or testing.

6.6 Experiment 6: Comparing Proposed FLS Classification Framework with other Classification Frameworks

In this experiment, we compared our proposed FLS based classification framework with other frameworks in the literature. We employed the algorithms for generating datasets as discussed earlier. We have conducted our experiments with eight different datasets to cover all the variations, such as small number of instances, large number of instances, number of attributes and number of classes, which are required for good experiments. Each dataset consisted of 70% for training and 30% for testing data points. The comparison was made on the basis of classification accuracy.

Table 6-15: Summary of Classification Accuracy Refer to Experiment 1 for eight different datasets

Dataset	Proposed FLS	ELM	F.Nets	SVM	FBNC	C.4.5 L
Wisconsin Breast Cancer	99.68 %	98.95 %	98.78 %	95.95 %	97.25 %	91.86 %
Hepatitis	91.80 %	91.10 %	90.65 %	80.86 %	86.90 %	81.50 %
Heart Statlog	90.10 %	88.74 %	88.24 %	82.48 %	83.81 %	79.85 %
Pima Diabetes	83.65 %	80.28 %	81.05 %	74.43 %	74.85 %	73.88 %
Mamographic Mass	91.72 %	89.36 %	87.90 %	86.05 %	87.30 %	85.68 %
Hypothyroid	97.50 %	94.39%	95.80 %	93.48%	93.16%	93.24%
Lymphography	90.35 %	87.29%	86.15 %	83.64%	85.20%	78.21%
Heart Disease	89.18 %	84.24%	84.45 %	82.57%	83.34%	79.61%

Table 6-16: Summary of Classification Accuracy Refer to Experiment 2 for eight different datasets

Dataset	Proposed FLS	ELM	F.Nets	SVM	FBNC	C.4.5 L
Wisconsin Breast Cancer	94.06 %	91.15 %	90.25 %	90.05 %	89.75 %	89.06 %
Hepatitis	88.65 %	87.34 %	87.50 %	86.16 %	83.90 %	82.36 %
Heart Statlog	89.13 %	87.67 %	86.90 %	84.80 %	82.81 %	80.42 %
Pima Diabetes	79.58 %	75.22 %	74.45 %	74.13 %	72.85 %	71.38 %
Mamographic Mass	87.18 %	83.10 %	83.24 %	81.25 %	79.45 %	78.18 %
Hypothyroid	95.85 %	91.39%	92.20 %	88.45%	86.10%	85.75 %
Lymphography	88.40 %	84.50%	83.34 %	81.64%	82.20%	76.21%
Heart Disease	88.26 %	85.14%	83.45 %	82.57%	83.34%	79.61%

Table 6-17: Summary of Classification Accuracy Refer to Experiment 3 for eight different datasets

Dataset	Proposed FLS	ELM	F.Nets	SVM	FBNC	C.4.5 L
Wisconsin Breast Cancer	91.34 %	86.10 %	86.45 %	84.35 %	83.25 %	82.86 %
Hepatitis	86.67 %	80.42 %	81.10 %	79.84 %	78.90 %	76.50 %
Heart Statlog	87.14 %	82.36 %	82.90 %	80.48 %	79.90 %	78.42 %
Pima Diabetes	77.82 %	72.61 %	73.55 %	71.43 %	70.05 %	70.88 %
Mamographic Mass	85.48 %	80.95 %	81.14 %	79.62 %	78.87 %	76.98 %
Hypothyroid	94.68 %	85.39 %	86.38 %	86.45 %	84.10 %	82.75 %
Lymphography	87.40 %	81.50 %	80.42 %	79.64 %	78.52 %	77.25 %
Heart Disease	86.96 %	80.75 %	80.15 %	78.38 %	77.20 %	75.58 %

It is evident from Tables 6-12, 6-13 and 6-14 that our proposed FLS has better preference than other machine learning techniques in all the experiments carried out in this research. Also, in the results of hypothyroid data set which has the largest data points, all the existing techniques showed a drastic decrease in prediction accuracy whereas our proposed FLS showed a very small change. Therefore we can conclude that our proposed FLS is capable of manipulating large datasets with promising results. Hence our intuition for handling laziness/ignorance in classification frameworks by using the proposed FLS is justified.

7. Conclusion and Future Work

The work resulted in the following contributions:

i. An extensive critical survey of some of the existing machine learning classification techniques was presented. Sources of uncertainty were discovered and well examined.

ii. Several types of FLS were classified and examined. This classification was performed on the basis of the nature of the inputs e.g., singleton or non-singleton; and on whether uncertainty is present in the system or not.

iii. A novel framework for handling imprecision and uncertainty by using Type-2 FLS was presented. The uncertainty due to various types of noise, inconsistent experts' opinion, ignorance and laziness etc. is encountered. We introduced about 25% uncertainty level in the datasets, and we proved that our proposed framework is much more robust than other techniques. We found that Type-2 FLS handles uncertainty better than Type-1 FLS.

iv. Impact of various FLS parameters on FLS based classification framework was studied empirically. The parameters studied were the defuzzification method and the shape of membership functions.

v. The impact of the nature of the training algorithm on the FLS based classification framework was investigated empirically. The approaches studied were Steepest Descent and Heuristic based.

vi. RMSE and ROC graphs showed that the proposed framework has better classification performance than Type-1 FLS. The experimental results showed that our proposed adaptive FLS based classification framework outperforms other existing classifiers in terms of classification accuracy.

vii. Some future work can be directed towards this framework to make it more robust, as other membership functions are yet to be introduced. Also more research can be done with the aim of validating this framework by using expert data which is usually obtained through surveys. It is hoped that more work can be carried out on using parallel computing to speed up the operation of Type-2 FLS, so that a Type-2 FLS can be run in about the same time as it takes to run a Type-l FLS.

REFERENCES

[1] Kiyana, T., Yildirim, T., “Breast cancer diagnosis using statistical neural networks”, In International XII, Turkish symposium on artificial intelligence and neural networks, pp. 754-760, 2003.

[2] Ramalingam, V., Palaniappan, B., Panchanatham, N., & Palanivel, S, “Measuring advertisement effectiveness - a neural network approach”, Expert Systems with Applications, 31(1), pp.159–163, 2006.

[3] Brause, R., "Medical analysis and diagnosis by neural networks, In Computer Science Department, Frankfurt a.M., Germany, 2001.

[4] Salim, N., Medical diagnosis using neural network, http://www.Generation.org/content /2004/.

[5] Duda, R.O., Hart, P.E. and Stork, D.G., Pattern Classification, John Wiley & Sons, Inc, New York, second edition, 2001.

[6] Statnikov, A., Aliferis, C.F., Tsamardinos, I., Hardin, D. and Levy, S., “A Comprehensive Evaluation of Multi-category Classification Methods for Micro array Gene Expression Cancer Diagnosis,” Bioinformatics, vol. 21, no. 5, pp. 631-643, 2005.

[7] Linder, R., Dew, D., Sudhoff, H., Theegarten, D., Remberger, K., Poppl, S.J. and Wagner, M., "The Subsequent Artificial Neural Network Approach Might Bring More Classificatory Power to ANN-Based DNA microArray analyses", Bioinformatics 20 (2004), pp. 3544–3552

[8] Khan, J., Wei, J.S., Ringner, M., Saal, L.H., Ladanyi, M., Westermann, F., Berthold, F., Schwab, M., Antonescu, C.R., Peterson, C., and Meltzer, S., "Classification and Diagnostic Prediction of Cancers Using Gene Expression Profiling and Artificial Neural Networks," Nature Medicine, vol. 7, no. 6, pp. 673-679, 2001.

[9] Hastie, T., Tibshirani, R. and Friedman, J., The elements of statistical learning-data mining, inference, and prediction. Springer, New York, ISBN 0-387-95284-5 2001.

[10] Michell, T.M., Machine Learning. McGraw-Hill, 1997.

[11] Gurcan, M. N., Chan, H. P., Sahiner, B., Hadjiiski, L., Petrick, N., Helvie, M.A "Optimal neural network architecture selection: Improvement in computerized detection of micro calcifications", Academic Radiology 9, 420–429. (2002).

[12] Christoyiani, I., Dermatas, E., & Kokkinakis, G., „Fast detection of masses in computer aided mammography", IEEE Signal Processing Magazine, pp. 54–64, 2000.

[13] Kiyaná, T., Yildirim, T., "Breast cancer diagnosis using statistical neural networks. In International XII, Turkish symposium on artificial intelligence and neural networks, 2003.

[14] Lee, J.W., Lee, J.B., Park, M., and Song, S.H., “An Extensive Comparison of Recent Classification Tools Applied to Microarray Data,” Computational Statistics and Data Analysis, vol. 48, pp. 869-885, 2005.

[15] Jiang, J.S., Zhang, H., "Full Bayesian Network Classifiers, "Proceedings of the 23 rd International Conference on Machine Learning, Pittsburgh, PA, 2006.

[16] Provost, F. J., & Domingos, P., "Tree induction for probability-based ranking" Machine Learning 52(3), pp.199-215, 2003.

[17] Cortes, C., and Vapnik, V., “Support-vector networks. Machine Learning, 20(2), pp. 273–297, 1995.

[18] Ringner, M., Peterson, C. and Khan, J., “Analyzing Array Data Using Supervised Methods,” Pharmacogenomics, vol. 3, no. 3, pp. 403-415, 2002.

[19] Huang, G.B., Zhu, Q.Y., and Siew, C.K., “Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks,” Proc. Int’l Joint Conf. Neural Networks (IJCNN ’04), July 2004.

[20] Huang, G.B., and Siew, C.K., “Extreme Learning Machine: RBF Network Case,” Proceeding of 8th Int’l Conf. Control, Automation, Robotics, and Vision (ICARCV ’04), Dec. 2004.

[21] Huang, G.B., and Siew, C.K., “Extreme Learning Machine with Randomly Assigned RBF Kernels,” Int’l J. Information Technology, vol. 11, no. 1, 2005.

[22] Zhang, R., Huang, G.B., Sundararajan, N., Saratchandran, P., “Multicategory Classification Using an Extreme Learning Machine for Microarray Gene Expression Cancer Diagnosis”, IEEE/ACM Transaction on Computational Biology and Bioinformatics, pp. 485-495, 2007

[23] Grabisch, M. and Dispot, F., "A comparison of some methods of fuzzy classification on real data," in 2nd Int. Conference on Fuzzy Logic and Neural Networks, pp. 659-662, 1992.

[24] Ishibuchi, H. and Nakashima, T., "Performance evaluation of fuzzy classifier systems for multi-dimensional pattern classification problems", IEEE Trans. Systems, Man and Cybernetics - Part B: Cybernetics, vol. 29, pp. 601-618, 1999.

[25] Ishibuchi, H. and Nakashima, T., "Improving the performance of fuzzy classifier systems for pattern classification problems with continuous attributes," IEEE Trans. on Industrial Electronics, vol. 46, no. 6, pp. 1057-1068, 1999.

[26] Hayashi, Y., "Neural Expert System Using Fuzzy Teaching Input and Its Application to Medical Diagnosis", Proceedings of the Second International Conference on Fuzzy Logic and Neural Networks, Iizuka, Japan, pp. 989-993, 1992.

[27] Watanabe, H. et.al. “The Application of a Fuzzy Discrimination Analysis for diagnosis of Valvular Heart Disease” IEEE trans. on Fuzzy Systems, pp. 267-276, 1994.

[28] Jain R. and Abraham A., A Comparative Study of Fuzzy Classifiers on Breast Cancer Data, 7th International Work Conference on Artificial and Natural Neural Networks, Lecture Notes in Computer Science- Volume 2687, Jose Mira and Jose R. Alverez (Eds.), Springer Verlag, Germany, pp. 512-519, 2003.

[29] Ishibuchi, H., and Nakashima, T., “A Study on Generating Classification Rules Using Histogram” Edited by Jain L.C., and Jain, R.K., KES’98, pp. 132-140, (1998)

[30] Dongrui W, JM. Mendel, “A vector similarity measure for linguistic approximation: Interval type-2 and type-1 fuzzy sets”, Information Sciences 178, pp. 381–402, 2008.

[31] Li Y, Sun X, Hua J, Gong C, “Modeling redundant structure in ecosystem by type-2 fuzzy logic system”, ecological modeling article in press, Elsevier, 2007.

[32] Castillo, O., Aguilar L., Cazarez, N., Cardenas S., “Systematic Design of a Stable Type-2 Fuzzy Logic Controller”. Applied Soft Computing Journal, 2007.

[33] Margaliot, M. and Langholz, G., “New Approaches to Fuzzy Modeling and Control: Design and Analysis”, World Scientific, Singapore, 2000.

[34] Chen X., Li Y., Harrison R., Zhang Y.Q., “Type-2 fuzzy logic based classifier fusion for support vector machines”. Applied Soft Computing Journal, 2007.

[35] Zadeh, L.A, “Fuzzy Sets”, Information and Control 8, 1965, pp. 338-353.

[36] The MathWorks – MATLAB and Simulink for Technical Computing
http://www.mathworks.com/access/helpdesk/help/toolbox/fuzzy/fp72.html

[37] Wang, L.: “Adaptive Fuzzy System and Control: Desing and Stability Analysis”, Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632, 1994.

[38] Mamdani, E. H.: “Applications of Fuzzy Algorithms for Simple Dynamic Plant”, Proc. IEEE 121, 12 (1974).

[39] Mendel, J. M.: “Uncertain Rule-Based Fuzzy Logic Systems”, Prentice-Hall, Upper Saddle River, NJ 07458, 2001.

[40] Zadeh, L.A.: “The Concept of a Linguistic Variable and Its Application to Approximate Reasoning-1”, Information Sciences, Volume. 8, pp. 199-249, 1975.

[41] Mendel, J.M. and Liang, Q.: “Pictorial Comparison of Type-1 and Type-2 Fuzzy Logic Systems”, In Proceedings of IASTED International Conference on Intelligent Systems & Control, Santa Barbara, CA, Oct., 1999.

[42] Nauck, D.: “Data Analysis with Neuro-Fuzzy Methods”, Habilitation Thesis, University of Magdeburg, 2000.

[43] http://mlearn.ics.uci.edu/MLRepository.html

[44] Fayyad U. M., Data mining and knowledge discovery: Making sense out of data. IEEE Expert, 1996, vol. 11, no. 5, 20-25.

[45] Aftarczuk K., Kozierkiewicz A, Nguyen N. T., Using Representation Choice Methods for a Medical Diagnosis Problem. Knowledge-Based Intelligent Information & Engineering Systems 2006, 805-812.

[46] Tsymbal A., Bolshakova N., Guest Editorial Introduction to the Special Section on Mining Biomedical Data, IEEE Transactions On Information Technology In Biomedicine, 2006, vol. 10, no. 3, 425-428.

[47] Duch W, Grabczewski K, Adamczak R, Grudzinski K, Hippe Z.S, Rules for melanoma skin cancer diagnosis, 2001 http://www.phys.uni.torun.pl/publications/kmk/ retrieved on 4.04.2007

[48] Vassiliadis P., Quix C., Vassiliou Y., Jarke M., DataWarehouse Process Management. Information Systems, 2001, vol. 26, no. 3, 205–236.

[49] Cios K., Moore G., Uniqueness of Medical Data Mining. Artificial Intelligence in Medicine, 2002, vol. 26, 1-24.

[50] Witten I. H., Frank E., Data Mining, Practical Machine Learning Tools and Techniques, 2nd Elsevier, 2005

[51] Van Gerven M.A.J., Jurgelenaite R., Taal B.G., Heskes T., Lucas P.J.F., Predicting carcinoid heart disease with the noisy-threshold classifier. Artificial Intelligence in Medicine, 2007, vol. 40, 45-55.

[52] Spiegelhalter D. and Knill-Jones R., Statistical and knowledgebased approaches to clinical decisionsupport systems, with an application in gastroenterology. Journal of the Royal Statistical Society, 1984, vol. 147, 35-77.

[53] Vlahou A., Schorge J.O., Gregory B.W., Coleman R.L., Diagnosis of ovarian cancer using decision tree classification of mass spectral data. Journal of Biomedicine and Biotechnology, 2003 vol. 5 308-314.

[54] Herron P., Machine Learning for Medical Decision Support: Evaluating Diagnostic Performance of Machine Learning Classification Algorithms, INLS 110, Data Mining, 2004.

[55] Li L.et al., Data mining techniques for cancer detection using serum proteomic profiling, Artificial Intelligence in Medicine, 2004, vol. 32, 71-83.

[56] Cosic D., Loncaric S., Rule-based labeling of CT head image. Lecture Notes in Artificial Intelligence, Berlin, Germany, Springer-Verlag, 1999, vol. 1211, 453–456.

[57] Duch W., Adamczak R., Grabczewski K., Zal G., Hayashi Y., Fuzzy and crisp logical rule extraction methods in application to medical data. Computational Intelligence and Applications, Berlin, Germany, Springer-Verlag, 2000, vol. 23, 593–616.

[58] Richards G., Rayward-Smith V. J., Sönksen P. H., Carey S., Weng C., Data mining for indicators of early mortality in a database of clinical records. Artificial Intelligence in Medicine, 2000, vol. 22, no. 3, 215–231.

[59] Comak E., Arslan A., Turkoglu I., A decision support system based on support vector machines for diagnosis of the heart valve diseases. Elsevier, 2007, vol. 37, 21-27.

[60] Turkoglu I., Arslan A., Ilkay E., An expert system for diagnosis of the heart valve diseases. Expert Systems with Applications, 2002, vol. 23, no.3, 229–236.

[61] West D., West V., Model selection for a medical diagnostic decision support system: a breast cancer detection case. Artificial Intelligence in Medicine, 2000, vol. 20, 183-204.

[62] Hayashi Y., Setiono R., Yoshida K., A comparison between two neural network rule extraction techniques for the diagnosis of hepatobiliary disorders. Artificial Intelligence in Medicine, 2000, vol.20, no. 3, 205–216.

[63] Cunningham P., Carney J., Jacob S., Stability problems with artificial neural networks and the ensemble solution. Artificial Intelligence in Medicine, 2000, vol. 20, no. 3, 217–225.

[64] Zhou Z.-H., Jiang Y., Yang Y.-B., Chen S.-F., Lung cancer cell identification based on artificial neural network ensembles, Artificial Intelligence in Medicine, 2002, vol.24, no.1, 25-36.

[65] Hunt M., von Konsky B., Venkatesh S., Petros P., Bayesian networks and decision trees in the diagnosis of female urinary incontinence, Engineering in Medicine and Biology Society, Proceedings of the 22nd Annual International Conference of the IEEE, 2000, vol. 1, 551-554.

[66] Teach R. and Shortliffe E., An analysis of physician attitudes regarding computer-based clinical consultation systems. Computers and Biomedical Research, 1981, vol. 14, 542-558.

[67] Domingos P. and Pazzani M., On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. Machine Learning, 1997, vol. 29, no. 2-3, 103-130.

[68] Daya S., Diagnostic test - receiver operating characteristic (ROC) curve. Evidence-based Obstetrics and Gynaecology, 2006, vol. 8, no. 1-2, 3-4.