Data Mining and Analytics 2008

Proceedings of the
Seventh Australasian Data Mining Conference (AusDM'08),
Glenelg, South Australia, 27-28 November, 2008

John F. Roddick, Jiuyong Li, Peter Christen and
Paul Kennedy, Eds.

Volume 87 in the Conferences in Research and Practice in Information Technology Series.
Published by the Australian Computer Society Inc.

Published in association with the ACM Digital Library.

Data Mining and Analytics 2008. Proceedings of the Seventh Australasian Data Mining Conference (AusDM'08), Glenelg, South Australia, 27-28 November, 2008

Conferences in Research and Practice in Information Technology, Volume 87.

Editors: **John F. Roddick**
School of Computer Science, Engineering and Mathematics
Flinders University
GPO Box 2100, Adelaide, SA, 5001, Australia
Email: john.roddick@flinders.edu.au

Jiuyong Li
School of Computer and Information Science
University of South Australia, Mawson Lakes
GPO Box 2471, Adelaide, SA, 5001, Australia
Email: jiuyong.li@unisa.edu.au

Peter Christen
Department of Computer Science
Faculty of Engineering and Information Technology
The Australian National University
Canberra ACT 0200 Australia
Email: peter.christen@anu.edu.au

Paul J. Kennedy
Faculty of Engineering and Information Technology
University of Technology, Sydney
Broadway, NSW, 2007, Australia
Email: paulk@it.uts.edu.au

Series Editors:
Vladimir Estivill-Castro, Griffith University, Queensland
John F. Roddick, Flinders University, South Australia
Simeon Simoff, University of Western Sydney, NSW
`crpit@csem.flinders.edu.au`

Publisher: Australian Computer Society Inc.
PO Box Q534, QVB Post Office
Sydney 1230
New South Wales
Australia.

Conferences in Research and Practice in Information Technology, Volume 87
ISSN 1445-1336
ISBN 978-1-920682-68-2

Printed November 2008 by Flinders Press, PO Box 2100, Bedford Park, SA 5042, South Australia.
Cover Design by Modern Planet Design, (08) 8340 1361.

The *Conferences in Research and Practice in Information Technology* series aims to disseminate the results of peer-reviewed research in all areas of Information Technology. Further details can be found at http://crpit.com/.

wFDT - Weighted Fuzzy Decision Trees for Prognosis of Breast Cancer Survivability

Umer Khan[1], Hyunjung Shin[2], Jong Pill Choi[3], Minkoo Kim[1]
[1]Graduate School of Information and Communication Engineering, AJOU University, South Korea
[2]Department of Industrial and Information Systems Engineering, AJOU University, South Korea
[3]Centre for Genome Sciences, Division of Biomedical Informatics, National Institute of Health, South Korea
umer@ajou.ac.kr, shin@ajou.ac.kr, cjp@ajou.ac.kr, minkoo@ajou.ac.kr

Abstract

Accurate and less invasive personalized predictive medicine can spare many breast cancer patients from receiving complex surgical biopsies, unnecessary adjuvant treatments and its expensive medical cost. Cancer prognosis estimates recurrence of disease and predict survival of patient; hence resulting in improved patient management. To develop such knowledge based prognostic system, this paper examines potential hybridization of accuracy and interpretability in the form of Fuzzy Logic and Decision Trees, respectively. Effect of rule weights on fuzzy decision trees is investigated to be an alternative to membership function modifications for performance optimization.

Experiments were performed using different combinations of: number of decision tree rules, types of fuzzy membership functions and inference techniques for breast cancer survival analysis. SEER breast cancer data set (1973-2003), the most comprehensible source of information on cancer incidence in United States, is considered. Performance comparisons suggest that predictions of weighted fuzzy decision trees (wFDT) are more accurate and balanced, than independently applied crisp decision tree classifiers; moreover it has a potential to adapt for significant performance enhancement.

Keywords: Prognosis, knowledge based, hybridization, accuracy, interpretability, membership functions, inference, crisp and fuzzy

1 Introduction

According to National Cancer Institute of United States, estimated number of new breast cancer cases in 2008 is 182,460 (female) and 1,990 (male), while the estimation of deaths is 40,480 (female); 450 (male) (National Cancer Institute 2008). Based on current rates, 12.7 percent of women born in US today will be diagnosed with breast cancer at some time in their lives. Surgical biopsies confirm malignancy with high level of sensitivity, but are considered costly and can affect patient's psychology as well (Iliaas, Elias and Ioannis 2007).

After confirmation of malignancy, oncologists get indulged into prognostic decision making. Surgery, radiation, chemotherapy, hormone therapy or any combination of them are considered to be the successful treatment methods. But again, selection of treatment method without considering the resulting tumour behaviour can lead to severe consequences. Therefore, being able to predict disease outcomes more accurately would help physicians make informed decisions regarding the potential necessity of adjuvant treatment. This may also lead to the development of individually tailored treatments to maximize the efficacy of treatment.

Ultimately, breast cancer mortality would also be decreased. This idea is the basic motivation behind the growing trend of focusing on accurate and less invasive personalized predictive medicine using machine learning techniques. This approach can spare many breast cancer patients from receiving complex surgical biopsies, unnecessary adjuvant treatments and its expensive medical cost (Yijun et al. 2007). Moreover, in situations where experienced oncologists are not available, predictive models created with data mining techniques can be used to support physicians in decision making with acceptable accuracy (Amir et al. 2007).

Prognosis helps in establishing a treatment plan by predicting the outcome of a disease. There are three predictive foci of cancer prognosis: 1) prediction of cancer susceptibility (risk assessment), 2) prediction of cancer recurrence and 3) prediction of cancer survivability. Focus of this paper is prediction of survivability of a particular patient suffering from breast cancer. ''Survival'' is generally defined as a patient remaining alive for a specified period of time after the diagnosis of disease. For this research effort, survival is considered as any incidence of breast cancer where the person is still living after 1825 days (5 years) from the date of diagnosis, as recommended by Dursun et al. 2004, Brenner and Gefellor 2002 and Cox DR. 1984.

In this research project, we surveyed various research efforts (Joseph et al. 2006, Ilias and Elias 2007, Dursun et al. 2004, Crokett et al. 2006, Andres and a-reyes 1999 and others will be mentioned later in this paper) in the application of different machine learning techniques to breast cancer prognosis. Some of the obvious trends which account for the motivation behind proposals and experiments presented in this manuscript are:

1. About 70% of all reported studies [Dursun et al. 2004] use Neural Networks which yield "Black Box" models for physicians to interpret.
2. Majority of reported studies in surveys like [Dursun et al. 2004] used machine learning

techniques independently without considering potential in those techniques to cooperate with each other in a hybrid model.
3. Fuzzy logic has been rarely used in cancer prognosis. Being non-crisp, it can act as a natural ally of a physician in prognostic decision making process.
4. Lack of attention paid to data size. Data sets considered are not sufficiently large that can be reasonably partitioned into disjoint training and test sets.

Before going into the details of these observations, let us first analyze their conceptual importance to intelligent cancer prognosis at the grass root level.

The design of any decision support system always faces a critical trade-off known as accuracy-interpretability trade-off. This trade-off becomes very sensitive and important in case of prognostic decision making for cancer treatment. Such data analysis systems, intended to assist a physician, are highly desirable to be accurate, human interpretable and balanced, with a degree of confidence associated with final decision. Accuracy and interpretability are highly conflicting requirements; since complexity of system usually increases as a result of accuracy maximization, resulting in reduced comprehensibility of system's overall behavior. "Improving accuracy while preserving interpretability" is a challenging research issue being actively pursued by designers of decision support systems (Gonzalez et al. 2007, Rafael et al. 2006, Ralf et al. 2004, Cristina and Louis 2004). This trade-off is one of the basic motivational factors behind the model presented in this paper. As mentioned earlier, majority of research efforts in breast cancer diagnosis and prognosis used neural networks (Joseph et al. 2006). This is because relative ease in their use, abilities to provide gradual responses and good classification performance. But in prognostic decision making systems where physicians want to understand and justify the decisions, they act totally as "Black Boxes" with poor interpretability because it is difficult for humans to interpret the symbolic meaning behind the learned weights. Moreover, neural network learning with too many attributes, as in case of breast cancer data (SEER 1973-2003), can result in over-fitting (Joseph et al. 2006).

Unlike neural network, decision trees have always been praised for comprehensibility of their knowledge representation and inference procedures. They have been shown to be problem independent and able to treat large scale industrial applications (Cristina and Louis 2003). Pruned decision tree effectively overcomes over-fitting problem when dealing with large number of attributes (Joseph et al. 2006). The fundamental weakness of decision trees is that the decision boundaries are sharp at each node (for continuous valued attributes), due to which even small changes in attribute values may result in misclassifications (Crockett et al 2006, Cristina and Louis 2003). That is why; they are recognized to be unstable, with high variance. Therefore, decision boundaries need to be softened and there should be a gradual transition between attribute values. Here comes the role of fuzzy logic, as explained next.

Based upon above mentioned observations, we propose to investigate a hybrid scheme based on weighted fuzzy decision trees (wFDT), as an efficient alternative to crisp classifiers that are applied independently. Fuzzy sets, along with fuzzy logic and approximate reasoning methods, provide the ability to model fine knowledge details (Lotfi A. Zadeh 1983). Accordingly, fuzzy representation is becoming increasingly popular in dealing with problems of uncertainty, noise, and inexact data (Cezary Z. Janikow 1998). That is why we believe, it can act as natural ally of physicians. To help decision trees, the role of fuzzy logic becomes very crucial in softening the sharp decision boundaries because of the elasticity of fuzzy sets formalism. An important aspect of this model is an interesting simultaneous cooperation between Fuzzy Logic and Decision Trees. This bidirectional cooperation tries to soften the accuracy/interpretability trade-off, and can be realized as follows:

1. Fuzzy representation, with its approximate reasoning handles uncertainty and gradual processing to help soften the crisp decision tree boundaries. This results in reduced misclassifications and increased accuracy.
2. There are two approaches of fuzzy modelling (FM) depending on problem domain (Rafael et al 2006):
   - Linguistic FM: based on Linguistic (Mamdani 1974) fuzzy rule based systems. These systems have high interpretability but strive to achieve improved accuracy.
   - Precise FM: based on Takagi-Sugeno 1985 fuzzy rule based systems. Such systems focus on accuracy but lack in interpretability.

   Figure-1 describes how linguistic and precise fuzzy modelling tend to achieve required optimal. For breast cancer prognosis problem, precise fuzzy modelling is used i.e. rules are in the form of Takagi-Sugeno 1985. Therefore, such systems lack in interpretability and need to achieve an optimal level of comprehensibility. In this case, decision trees will help such fuzzy representation.

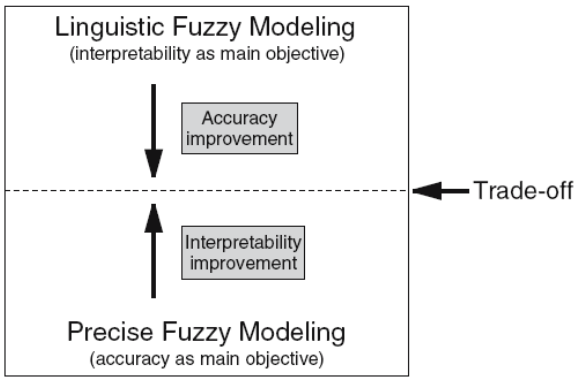

Figure 1: Precise fuzzy modelling tends to achieve optimal interpretability (Rafael et al. 2006)

Fuzzy decision tree (FDT) IF-THEN rules for an m-class pattern classification problem with 'n' attributes

can be written as:

Rule R_i: If $\boldsymbol{x_1}$ **is** $\boldsymbol{A_{i1}}$ AND $\boldsymbol{x_2}$ **is** $\boldsymbol{A_{i2}}$ AND……AND $\boldsymbol{x_n}$ **is** $\boldsymbol{A_{in}}$ THEN ***Class*** $\boldsymbol{C_i}$ $i = 1,2,…..N$

where $x = (x_1, x_2,…,x_n)$ n-dimensional pattern vector,
A_{ij} is antecedent Precise fuzzy set (like <30, >=30),
C_i Consequent class (one of the given m-classes in labelled date),
N is the number rules used in a particular model.

Effect of rule weights on fuzzy decision trees is analysed using certainty grade concept. It determines the degree of confidence in the decision of a particular rule. A FDT rule with an associated certainty grade can be written as:

Rule R_i: If $\boldsymbol{x_1}$ **is** $\boldsymbol{A_{i1}}$ AND $\boldsymbol{x_2}$ **is** $\boldsymbol{A_{i2}}$ AND……AND $\boldsymbol{x_n}$ **is** $\boldsymbol{A_{in}}$ THEN ***Class*** $\boldsymbol{C_i}$ with $\boldsymbol{CF_i}$ $i = 1,2,…..N$

Usually, CFi is a real number in unit interval ($0 \le CFi \le 1$). A special rule weighting technique is used which learns certainty grades from training data. Using certainty grade, compatibility of each rule is calculated for each incoming input record to be classified. The most compatible rule for a particular input record decides its final class. Effect of this weighting is investigated to be an alternative to membership function optimization. A significant performance enhancement is achieved by weighting rules, which also helps oncologists to have certain degree of confidence in the final decision.

The overall aim of this research is to determine the potential of wFDTs for prediction of breast cancer survivability in particular, and breast cancer prognosis in general. wFDT is studied in detail and compared with FDT and crisp decision tree. Experiments were performed rigorously using different combinations of: number of rules in a model, types of fuzzy membership functions and inference techniques. Results show that wFDT achieved much improved prediction accuracy and much reduced variance, as compared with crisp decision tree.

Rest of the paper is organized as follows: section 2 presents the related work, section 3 describes materials and methods used in this research, section 4 presents the experimental evaluations and finally section 5 concludes this manuscript.

2 Related Work

In (Joseph et al. 2006), authors conducted a broad survey of the different types of machine learning methods being used, the types of data being integrated and the performance of these methods in breast cancer prediction and prognosis. To get possible research directions in application of machine learning techniques for cancer prognosis, this survey is the only detailed manuscript (by date) especially for researchers new to this application area. In (Dursun et al. 2004), a comparison between two data mining techniques namely decision trees and neural networks and a statistical method namely logistic regression, is presented. These techniques were applied independently on SEER breast cancer data (SEER 1973-2003) to predict survivability. This research effort concluded that decision trees proved to be the best classifier in that experimental environment. We propose that this performance can be extensively increased using weighted and fuzzified decision tree i.e. wFDT. This was another reason (besides others mentioned earlier) to select decision trees for constructing crisp rule base.

In (Carlos and Moshe 1999), fuzzy rules for cancer diagnosis are generated by randomly selecting data instances from training data, and performing rigorous genetic search evolving different models and then selecting the best ones. According to our approach, an efficient and well tested classifier can be used to build initial rule base, avoiding complexities and optimization errors due to random selection of training records. Moreover, rigorous and repetitive genetic search through a realistically huge cancer patient data (like one used in this research) would result in tedious time and memory complexities.

In recent research efforts for cancer prediction (Ilias et al. 2007 and Leonardo et al 2007), support vector machine (SVM) and neural network (ANN) modelling were performed. In both the cases, main focus was accuracy and no doubt, they would have achieved "high peaks" of accuracy. But a clinician, involved in sensitive decision making about a patient's treatment, demand more than that. Factors including interpretability, system's ability to adopt human reasoning behaviour to deal with uncertainties and performance consistency were ignored.

Let us review research efforts specifically focused on hybridization of fuzzy logic and decision trees, other than cancer prediction domain. To cope with sharp decision boundaries problem, a number of approaches (Cristina and Louis 2003, C.Z Janikow 1999 and 2004, C.W. Olaru 2003 and Yuan et al. 1995) have made use of fuzzy theory to create fuzzy trees. In (Webb and Ting 2005, Umano et al. 1994) fuzzy tree is induced directly from pre-fuzzified data. The difference between these approaches and the one used in this manuscript is that, they focus on modification of decision tree pruning algorithm and require fuzzy parameters to be set by domain experts. Here fuzzy sets produced can be the outcome of subjective perception. This way an additional aspect of uncertainty is introduced in the system, when there are conflicting opinions between domain experts. In (Crockett et al. 2006), a similar architecture is proposed in which pre-constructed tree is fuzzified without modifying ID3 algorithm. But the rule weighing technique in their inference procedure is very trivial and they did not focus on analysing the strength of certainty grades in system's performance and comprehensibility.

Effect of rule weights in fuzzy rule-based classification systems is examined in (Hisao and Tomoharu 2001). They presented an effective analysis of applying rule weights as an efficient alternative to membership function learning and optimization. Effect of certainty grades on the decision areas of fuzzy rules is illustrated. The larger the certainty grade of a rule, the larger will be its decision area. Their experiments are focused on linguistic values of fixed membership functions. In this work, we have shown the same effect on precise fuzzy modelling, in section 4.

3 Materials and Methods

3.1 Prognostic and Predictive Factors in Breast Cancer

Survival of patients with breast cancer depends on two different types of prognostic factors: 1) Chronological [indicators of how long the cancer has been present (e.g. tumor size)], 2) Biological [indicators of metastatic aggressive behaviour of a tumour (e.g. tumor grade)] as described in (Bundred N.J. 2001). They determine, either or not a particular tumor might respond to a specific therapy. Definitions and effects of some of the most important prognostic factors in breast cancer are given below:

Lymph node status: Lymph nodes, where cancer cells get accumulated (usually under the arm pits). Both number of nodes and level of involvement worsen the prognosis. If the lymph nodes accumulate cancerous cells, they are called positive nodes, otherwise negative.

Stage: Defined by the size of tumor and its spread. Survival is inversely proportion to size of tumor. The probability of long-term survival is better with smaller tumors than with larger tumors (Bundred N.J. 2001). Let us see some examples of breast cancer stage from (National Cancer Institute 2008).

1. Stage-1 is an early stage of invasive cancer. Tumor is no more than 2 cm. Cancer cells remained inside breast.
2. In Stage-2, tumor size can be from 2-5 cm. Cancer cells may or may not spread out of breast.

Similarly, there in total four types of stages (further divided into many sub-types), in which tumor size keeps on increasing along with spreading of cancer cells. Higher the stage, difficult is survival.

Grade: How does the tumor look like and its resemblance to more or less aggressive tumors. Histological grade is a combination of mitotic rate, nuclear grade and architectural morphological appearance (Rampaul 2001). Here also, patients with grade-1 tumor have higher chances of survival than patients of grade-3.

Figure-1 shows ranking of survivability attributes in terms of their decisive strength, calculated using information gain (IG) applied to breast cancer data, as described in subsection 3.3.

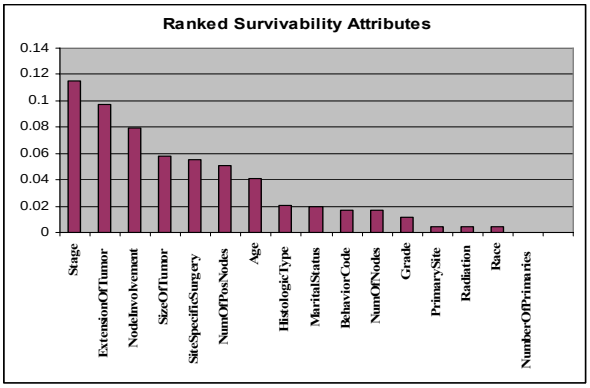

Figure 2: Ranked Survivability Attributes

3.2 Data

In this research work, SEER (Surveillance, Epidemiology, and End Results) data (1973-2003) is used for breast cancer prognosis. Files were requested through website (www.seer.cancer.gov) of SEER program which is a part of Surveillance Research Program at National Cancer Institute. The data set is considered to be the most comprehensive source of information on cancer incidence in USA and SEER program claims quality and completeness of data. A search for term 'SEER' on PUBMED (National Library of Medicine's database), gives a list of more than 1500 publications for the time period of 1978-2008.

Initially, there were 505,367 records each with 86 variables. These variables describe socio-demographic and cancer-specific information of an incidence of cancer. We used Clementine data analysis tool for all preprocessing mentioned below. Considering the aim of survival prediction, a binary target variable is created with values 0 (did not survive) and 1 (survived). To calculate this variable, 'SurvivalTimeRecode' field is used which provides number of years and months of survival after diagnosis. Although much of the time in this research work was spent on data cleansing and preprocessing, only a brief description is given here. To adjust the survival rate, those records were removed in which patient died within 5 years after diagnosis and the cause of death was not breast cancer.

SEER used the same database schema for the data of all anatomical sites (like breast, throat, urinary etc.). So there were many attributes which are common to all cancer types and not specific to breast cancer. Moreover, some redundant variables like recodes and overrides were also removed. For instance, Extent_of_Disease variable aggregates tumor size, # of nodes examined, # of positive nodes examined, lymph node involvement and clinical extension of tumor.

Other than this, variables that had more than 70.0% missing values, categorical variables that had a single category accounting for more than 90.0% of cases, continuous variables that had standard deviation less than 0.1%, and continuous variables that had a coefficient of variation (SD/mean) less than 0, were also removed. For input variable selection, we tried to limit the number of variables and selected only the clinically relevant variables. But for some variables like Stage, 40% of records contained missing values. Since this variable is an important predictor of survivability, instead of deleting whole column, only records containing missing values for this variable were removed.

After an exhaustive pre-processing, final data set with 162500 records, 16 predictor variables and 1 target variable, was constructed. The target variable 'IsSurvival' is a binary categorical variable with possible values '0' (did not survive) and '1' (survived). Table-1 shows the predictive variables and their descriptions, used in our work:

Field	Description
Stage	**Defined by the size of cancer tumor and its spread**
Grade	**How does the tumor looks like and its resemblance to more or less aggressive tumors.**
Lymph Node Involvement	**None, (1-3) Minimal, (4-9) Significant etc**
Race	**Ethnicity like White, Black, Chinese etc.**
Age at Diagnosis	**Actual age of patient in years**
Marital Status	**Married, Single, Divorced, Widowed, Separated**
Primary Site	**Presence of tumor at a particular location in body. Topographical classification of cancer**
Tumor Size	**2-5 cm, at 5cm prognosis worsens**
Site Specific Surgery	**Information on surgery during first course of therapy whether it was cancer directed or not.**
Radiation	**None, Beam Radiation, Radioisotopes, Refused, Recommended etc.**
Histological Type	**The form and structure of tumor**
Behavior Code	**Normal or aggressive behaviour of tumor have been defined in codes.**
# of Positive Nodes Examined	**When the lymph nodes are involved in the cancer, they are called "positive."**
# of Nodes Examined	**Total nodes (positive/negative) examined**
# of Primaries	**Number of primary tumors (1-6)**
Clinical Extension of Tumor	**Defines the spread of tumor relative to breast**
IsSurvival	**Target binary variable defines the class of survival of patient.**

Table 1: Breast Cancer Predictive Factors used for Survivability Prediction

Following table shows the important statistics related to above mentioned prognostic factors in training data. Here symbol is assigned to recognize each feature, in the order it appears in training record i.e. Age (A) comes first and number of primaries (P) appears last in training record.

Symbol	Nominal Variable Name	Num of Distinct Values
B	Race	28
C	Marital Status	9
D	Primary Site	9
E	HistologicTypeICD	44
F	Behaviour Code	2
G	Grade	5
I	Extension of Tumor	12
J	Node Involvement	10
M	Site Specific Surgery	22
N	Radiation	9
O	Stage	9

Symbol	Numeric Variable Name	Mean	Std.Dev	Range
A	Age at Diagnosis	61.105	14.165	20-106
K	Num of Pos Nodes	24.376	41.238	0-99
H	Tumor Size	103.168	273.144	0-999
L	Num of Nodes	14.033	16.89	0-99
P	Num of Primaries	1.302	0.565	1-6

Table 2: Statistical Description of Predictor Variables

3.3 Decision Trees

Decision tree techniques have always been popular for extracting rules from domain knowledge to classify objects. As mentioned above, to generate fuzzy rules, we opted to use decision trees in the first step of modeling. We used binary C4.5 for all the experiments mentioned in this manuscript. A brief description of its working is given here. To partition the data at each stage of tree, a test is performed to select an attribute with lowest entropy. Information gain (IG) is used as a measure of entropy (H) difference when an attribute contributes the additional information about class C (Witten and Frank 2005).

$$H(C) = -\sum p(c)\, logp(c) \quad , c \in C \qquad (1)$$

$$H(C|X_i) = -\sum p(x) [\, \sum p(c|x) log\, p(c|x) \,], x \in X_i, c \in C \qquad (2)$$

$$IG_i = H(C) - H(C|X_i) \qquad (3)$$

In equation (1), p(c) is the probability that an arbitrary sample belongs to class 'c'. Equation (2) shows the entropy after observing the attribute X_i for the class 'c' and 'p(c|x)' is the probability that a sample in attribute branch X_i belongs to class 'c'. We used binary decision tree because it has been proved in previous research (H. Al-Attar 1996, J.R. Quinlan 1990) that they usually outperform the multi-branch trees generated by the original ID3 algorithm. To cater for over fitting problem, the constructed tree is optimized in size using pruning. During experiments, we generated different decision tree models, and analysed the following trend.

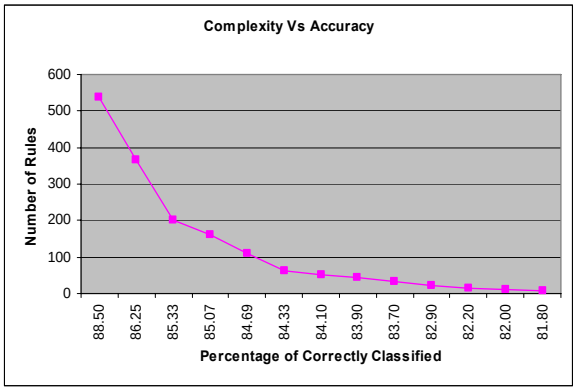

Figure 3: Rules per Model and Accuracy

As shown in figure-3, with a maximum complexity we got maximum accuracy. But for a model with 20 rules to a model with 8 rules, we got the almost similar value of accuracy i.e. around 82%. This trade-off is explained below.

3.3.1 C4.5 Limitations, Interpretability and Model Selection

As described earlier, the fundamental weakness of crisp C4.5 decision tree is that the induced tree will have sharp decision boundaries at each node. In case of continuous attributes, even small changes in attribute values may result in misclassifications. In (Quinlan 1990

R1: IF ClinicalExtensionOfTumor <= 40.0 AND AgeAtDiagnosis <= 73.0 THEN IsSurvival = YES

R2: IF ClinicalExtensionOfTumor <= 40.0 AND AgeAtDiagnosis > 73.0 AND AgeAtDiagnosis <= 83.0 And LymphNodeInvolvement <= 1.0 THEN IsSurvival = YES

R3: IF ClinicalExtensionOfTumor <= 40.0 AND AgeAtDiagnosis > 73.0 AND AgeAtDiagnosis <= 83.0 And LymphNodeInvolvement > 1.0 AND NumOfPositiveNodesExamined <= 5.0 THEN IsSurvival = YES

R4: IF ClinicalExtensionOfTumor <= 40.0 AND AgeAtDiagnosis > 73.0 AND AgeAtDiagnosis <= 83.0 And LymphNodeInvolvement > 1.0 AND NumOfPositiveNodesExamined > 5.0 THEN IsSurvival = NO

R5: IF ClinicalExtensionOfTumor <= 40.0 AND AgeAtDiagnosis > 73.0 AND AgeAtDiagnosis > 83.0 AND AJCCStage3ed <= 10.0 AND AgeAtDiagnosis <= 85.0 THEN IsSurvival = YES

R6: IF ClinicalExtensionOfTumor <= 40.0 AND AgeAtDiagnosis > 73.0 AND AgeAtDiagnosis > 83.0 AND AJCCStage3ed <= 10.0 AND AgeAtDiagnosis > 85.0 THEN IsSurvival = NO

R7: IF ClinicalExtensionOfTumor <= 40.0 AND AgeAtDiagnosis > 73.0 AND AgeAtDiagnosis > 83.0 AND AJCCStage3ed > 10.0 THEN IsSurvival = NO

R8: IF ClinicalExtensionOfTumor > 40.0 THEN IsSurvival = NO

Figure 4: Least Complex Model for Interpretability

and Carter et al. 1987), some threshold softening approaches are considered for categorical and continuous attributes. But the results related to splitting of continuous attributes do not show significant improvement (Quinlan 1996). The trade-off presented in figure-3, gets harder due to sharp decision boundaries. For a physician involved in prognostic decision making, both accuracy and interpretability are a must. So we decided to choose interpretability first and leave the accuracy and decision confidence for the second stage.

Although it is difficult to give a precise definition of interpretability, many researchers like (Ralf et al. 2005, Bodenhofer and Bauer 2002, Cordon and Herrera 2000, Jin et al. 1998) have agreed on interpretability involving following aspects:

1. Number of rules should be small enough to be comprehensible.
2. Rule antecedents and consequents should be in easy in structure and it should contain only few features.
3. Rule base should be consistent i.e. similar antecedents should produce similar consequents.
4. Fuzzy system should use features familiar to users.
5. The inference mechanism should produce technically and intuitively correct results.

Based on above recommendations, out of different models generated during experiments, we have chosen the 8-rules least complex model shown in figure-4. This is because from models with 20 rules to model with 8 rules, accuracy remained almost same, as shown in figure-3. Hence, we decided to choose simplest model with 8 rules and an acceptable accuracy 81.5%, to act as base model for FDT and wFDT in next section. Note that, this model contains most decisive factors ranked using Information Gain i.e. extension of tumor, stage, lymph node involvement and positive nodes examined as shown in figure-2. Age also gives a strong idea about survival since it is easier to recover in young age.

3.4 Weighted Fuzzy Decision Trees (wFDT)

In continuation of previous section, we already have an induced crisp decision tree, which partitions the input and output space into n-dimensional space where 'n' is total number of attributes. To convert a sharp transition at decision node into a gradual one, fuzzification is applied to both branches of a decision node (since we are using a binary decision tree). Figure-5 gives a simple and clear idea of crisp and fuzzy classes visually.

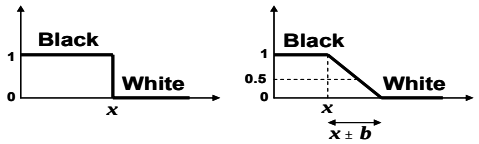

Figure 5: Difference between Crisp & Fuzzy Sets

In figure-5, '$x \pm b$' is a relaxation applied to crisp threshold. To do this, an attribute or decision node is represented by a fuzzy set using a pair of complimentary membership functions M_1 and M_2, as elaborated in figure-6.

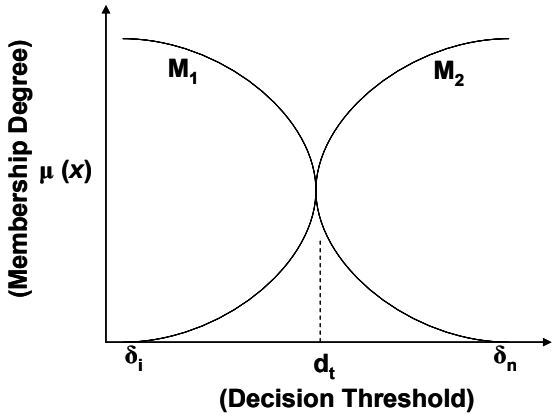

Figure 6: Complimentary Membership Functions over domain δ_i and δ_n

Fuzzy region is defined around a crisp threshold 'dt', already defined at each decision node or attribute by C4.5

algorithm. A membership function defines degree of membership μ (*x*), of a particular input value '*x*', into a fuzzy set. This degree lies in the range 0 to 1, with μ(*x*)=0 means 'no membership' and μ(*x*)=1 means 'full membership'. Membership degree or value is a key concept which ensures that sharp transition concept ceases to exist. Some examples of membership functions will be explained later in this section. Although there can be many smart ways to initially specify the domain, lower bound 'δ_i' and upper bound 'δ_n' of a membership function, we stick to a common and simplified domain specification. Since decision threshold 'd_t' is already generated at each node of DT and remains fixed, domain delimiters can be calculated as:

$$\delta_i = d_t - f * \sigma \quad \text{and} \quad \delta_n = d_t + f * \sigma \qquad (4)$$

here δ_i and δ_n are lower and upper bounds of membership function, respectively. 'σ' is the standard deviation of the domain attribute. It determines how tightly an attribute's values are bound around its mean value. It helps in guessing what proportion of an attribute would be assigned partial membership degree. '*f*' is the fuzzification applied around decision threshold 'dt'. Studies have shown empirically that '*f*' is usually chosen in the domain '0-5'. This is because larger values of '*f*' would introduce too much fuzzification and decision making process would become unclear. For our experiments, '*f*'=2 gave the optimal results.

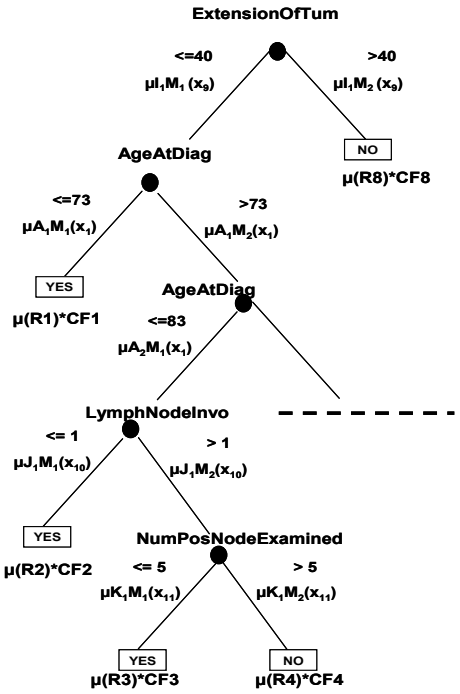

Figure 7: Fuzzified Decision Tree

A portion of fuzzified decision tree is shown in Figure-7 It describes how an attribute at each node is fuzzified using left and right complimentary membership functions (M1, M2). Crisp rules R_1 and R_8, in figure-4, can be obtained in fuzzified form by traversing the left most and right most paths of the tree in figure-7. An important feature of this approach is that, it preserves the decision thresholds and symbolic structure obtained from induced tree.

3.4.1 Fuzzy Inference

The approach used is very simple and interesting in a way that, for classifying an example, all the rules contribute their knowledge to some degree. A brief description is given below: (for classification of an incoming record)

1. For each path (or rule) of the fuzzy decision tree, a cumulative membership grade is calculated by applying an intersection operator (like Yager or Zadeh operator) to the set of individual membership function values at each branch on that path. For example, cumulative membership grade of first rule R_1(left most path from root to leaf, in figure-7) is computed as:

$$\mu(R_1) = \cap \{\mu I_1 M_1 (x_9), \mu A_1 M_1 (x_1)\} \qquad (5)$$

Here '$\mu I_1 M_1(x_9)$' is the membership function of fuzzy set "ExtensionOfTumor <=40". 'I_1' means first occurrence of 'ExtensionOfTumor' node in tree. 'M_1' means this function is left complimentary function. 'x_9' is 'ExtensionOfTumor' value in input record. This membership function will compute membership value of the input record, for a particular branch in rule path. Now we have 8 cumulative membership grades. Each of such grades is multiplied by the rule weight or certainty factor CF_i. Section 3.4.3 explains in detail about the computation of CF and its effect on fuzzy rule based classification system. This weight is calculated for each rule using training data. Rule weight has a great significance in fuzzy inference and here it is used as an alternative of Genetic Algorithms for parameter optimization.

2. Finally, all the products ($\mu(R_i)*CF_i$) are combined using *union operator*, and a rule (e.g. Zadeh) or a class (e.g. Yager) with maximum membership grade, will decide the class of incoming record.

$$\text{Decision} = \cup\{\mu(R_1)*CF_1 , \mu(R_2)*CF_2 , \ldots \mu(R_8)*CF_8\} \qquad (6)$$

We used Yager and Zadeh (Witten and Frank 2005) inference operators for intersection and union.

3.4.2 Fuzzy Membership Functions

We have used different types of fuzzy membership functions like Linear, sigmoid, convex and concave membership functions (Earl Cox 1994) to evaluate wFDTs. A brief description of these membership functions is given below:

$$\text{Linear}(\delta_i, \delta_n, x) = \left\{ \begin{array}{l} 0 \rightarrow x \le \delta i \\ x - \delta i \,/\, \delta n - \delta i \rightarrow \delta i \le x \le \delta n \\ 1 \rightarrow x \ge \delta n \end{array} \right\}$$

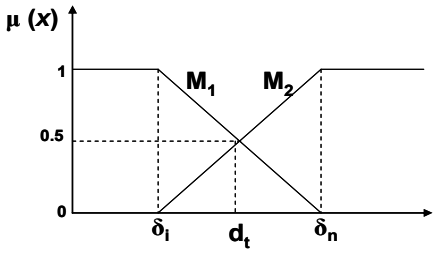

Figure 8: Linear Membership Function

Here '*x*' is the input value of an attribute. 'δ_i' generates zero membership while 'δ_n' generates maximum membership i.e. '1'. Both 'δ_i' and 'δ_n' are computed from equation-4.

$$\text{Sigmoid}(x;\ \delta_i, ,\delta_n, \beta) = \begin{cases} 0 \rightarrow x \le \delta i \\ 2\left(\left(x-\delta i\right)\Big/\left(\delta n-\delta i\right)\right)^2 \rightarrow \delta i \le x \le \beta \\ 1-2\left(\left(x-\delta n\right)\Big/\left(\delta n-\delta i\right)\right)^2 \rightarrow \beta \le x \le \delta n \\ 1 \rightarrow x \ge \delta n \end{cases}$$

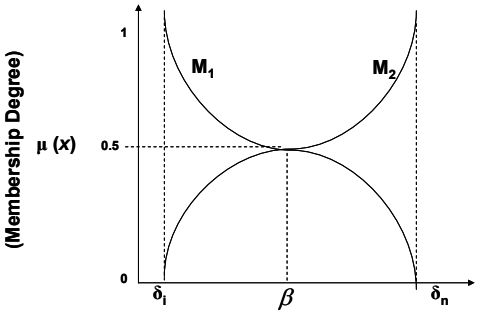

Figure 9: Sigmoid Membership Function

Here β is defined as half membership point ($\delta_i + \delta_n$)/2. It represents known distribution of sample space, and assumed to be 'dt = β'. Fuzzification gets to its maximum as '*x*' gets closer to 'δ_n'.

$$\text{Convex}(\delta_i, \delta_n, x) = \begin{cases} 0 \rightarrow x < \delta i \\ 1-\left[\ 2*\left(x-\delta n\right)\Big/\delta n\ \right] \rightarrow \delta i \le x \le \delta n \\ 1 \rightarrow x > \delta n \end{cases}$$

$$\text{Concave}(\delta_i, \delta_n, x) = \begin{cases} 0 \rightarrow x < \delta i \\ \left(x-\delta i\right)\Big/\delta n - \delta i \rightarrow \delta i \le x \le \delta i \\ 1 \rightarrow x > \delta n \end{cases}$$

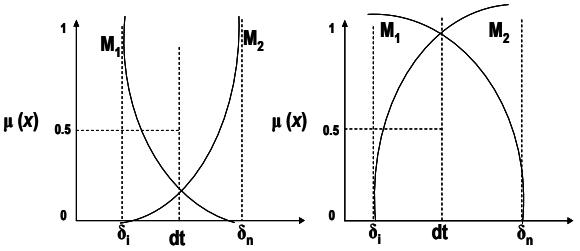

Figure 10: Convex (Left) and Concave (Right) Membership Function

In convex membership function, the membership grade at point of intersection is significantly less than 0.5. As the value of '*x*' equals 'dt', membership would be low in both M_1 and M_2 representing low confidence in both child branches of decision node in binary tree.

Concave membership function assigns higher membership grade to the branching threshold, which implies strong confidence in both child branches of decision node.

3.4.3 Effect of Weights on Fuzzy Rules

In this section, we have discussed a very important and significant aspect of wFDT modelling. Effect of weights, learnt from data, on fuzzy rule base is analysed. For the performance enhancement of fuzzy rule based systems, there has always been a room for membership function optimization through learning or other adjustment techniques. This analysis is based on an argument that learning of certainty grades (rule weights) can partially replace the adjustment of membership function. A few aspects of this analysis are referenced from (Hisao and Tomoharu 2001, Nauck and Kruse 1998) in which they discussed rule weighing for linguistic fuzzy modelling.

This concept is based on an assumption that modifying a membership function can deteriorate the comprehensibility of fuzzy classification system. It can also introduce a gap between modified membership function and expert's knowledge about that function. On the other hand, learning single real number is a relatively easier task, and it improves the classification accuracy of fuzzy rule based system. Another significant importance is that it represents the strength of each rule, in other words the confidence in rule's decision. This would help physician in establishing his confidence in a particular rule.

In fuzzy rule based systems where inference is based on one winner rule classification, if certainty grades are not used, the rule with maximum compatibility grade (membership value) for a record to be classified, decides the class (detailed inference mechanism is described in next section). Following expression formulize this concept.

$$\mu_{j*}(X) = \max\left(\mu_j(X) \mid j=1,2\ldots N\right) \qquad (7)$$

This expression simply shows that the rule with maximum membership value for an input record 'X', will decide its class. 'N' is total number of rules. Based on this, each rule has a particular decision area as shown in figure-11.

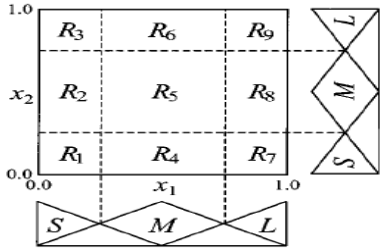

Figure 11: Decision area of fuzzy rules (Hisao et al. 2001)

Decision areas of rules without certainty grades are proved to be rectangular (Kenchuva 2000). These decision areas can be modified and adapted, by learning certainty grades (rule weights) from data, to alternatively affect the membership functions without explicitly modifying them. Modified decision areas will automatically result in modified class boundaries. Figure-12 shows the effect of certainty grade on fuzzy sets.

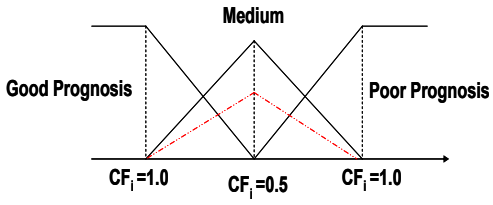

Figure 12: Effect of Certainty Grade on Fuzzy Set

The red dashed line shows the product of compatibility grade and certainty factor (CF) of a rule which modified class boundary. Now the rule with maximum of this product will be the winner as expressed by the following equation:

$$\mu_{j*}(X)CF_{j*} = \max(\mu_j(X)CF_j \mid j=1,2\ldots N) \quad (8)$$

Certainty grades or rule weights are calculated for each rule as:

1. When consequent class of Rule is YES (or 1)

$$CFi = \left(\beta_{ClassYES}(R_i) - \beta_{ClassNO}(R_i)\right) \Big/ \left(\beta_{ClassYES}(R_i) + \beta_{ClassNO}(R_i)\right) \quad (9)$$

2. When consequent class of Rule is NO (or 0)

$$CFi = \left(\beta_{ClassYES}(R_i) - \beta_{ClassNO}(R_i)\right) \Big/ \left(\beta_{ClassYES}(R_i) + \beta_{ClassNO}(R_i)\right) \quad (10)$$

Where $\beta_{ClassK}(R_i) = \sum_{x \in ClassK} \mu_i(x), k = YES, NO$

In simple words, for each rule 'R_i' its combined membership value for all the training patterns of class 'YES' ($\beta_{ClassYES}(R_i)$), and its combined membership value for all the training patterns of class 'NO' ($\beta_{ClassNO}(R_i)$) is computed to get its certainty grade using above equations. Certainty grade values lies in the range $0 \le CF_i \le 1$, which means when all compatible patterns with rule R_i (those with $\mu_i(x)$ >0 for R_i) belong to the same class, CF_i equals one.

This analysis, resulted in significantly increased performance, as mentioned in next section.

4 Performance Evaluation

Experiments were performed using WEKA, Matlab and Java on a Pentium PC at 1.7GHz with 1.5GB RAM. Execution time for calculating decision tree with different kernel functions varied for 6 to 12 seconds. Out of 162500 records, 30000 records as training and 10600 as test data were obtained using uniform random selection, taking into account the overlapping factor (stratified sampling).

4.1 Accuracy, Sensitivity and Specificity

In this study, we have used following three performance measures:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (12)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (13)$$

where TP, TN, FP and FN denotes true positives, true negatives, false positives and false negatives, respectively.

4.2 10-Fold Cross Validation

k-Fold cross validation is used to minimize the bias associated with random sampling of training and test data samples in comparing predictive accuracy of two or more methods (Dursun et al. 2004). Here the whole data set is randomly split into 'k' mutually exclusive subsets of approximately equal size. Classification model is trained and tested k times. Each time it is trained on all but one fold. For example, we used 10-fold cross validation because empirical studies (Kohavi 1995, Breiman et al. 1984) have shown that 10 folds are appropriate to optimize the testing time and minimize the bias and variance associated with validation process. In this case, data is split into 10 mutually exclusive subsets (using stratified sampling). Each of these 10 folds is used once to test performance of classifier, while other 9 are used for training.

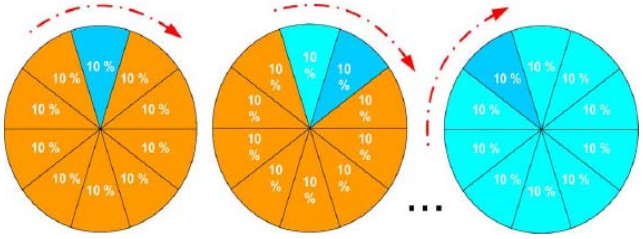

Figure 13: 10-Fold Cross Validation (Dursun et al.)

Cross validation estimate of classifier's overall accuracy is calculated by simply taking the mean of 'k' individual accuracy measures. Table-3 shows the 10-fold cross validation estimates of crisp decision trees, fuzzy decision trees (FDT) and weighted fuzzy decision trees.

Fold #	Crisp Decision Tree					FDT					wFDT				
	Confusion Matrix		Accuracy	Sensitivity	Specificity	Confusion Matrix		Accuracy	Sensitivity	Specificity	Confusion Matrix		Accuracy	Sensitivity	Specificity
1	7810	1451	0.812	0.85	0.607	7943	1318	0.8313	0.8576	0.666	8356	1090	0.8914	0.885	0.944
	551	850				481	920				68	1148			
2	7755	1496	0.8028	0.838	0.571	7923	1328	0.8279	0.8564	0.641	8346	1100	0.8895	0.8835	0.94
	606	805				506	905				78	1138			
3	7705	1526	0.7962	0.8346	0.5485	7938	1314	0.8299	0.8579	0.646	8359	1082	0.8913	0.8854	0.9369
	646	785				499	911				77	1144			
4	7825	1421	0.8111	0.846	0.581	7954	1290	0.8324	0.86	0.6495	8347	1094	0.8890	0.8841	0.926
	593	823				497	921				90	1131			
5	7741	1471	0.7988	0.84	0.535	7950	1293	0.8329	0.86	0.655	8367	1079	0.8923	0.8858	0.9433
	674	776				489	930				69	1147			
6	7680	1502	0.79	0.836	0.51	7940	1321	0.831	0.8574	0.654	8380	1066	0.8947	0.8871	0.9539
	724	756				485	916				56	1160			
7	7845	1421	0.8169	0.8466	0.6196	7905	1351	0.8295	0.8540	0.6685	8320	1116	0.8914	0.8850	0.9666
	531	865				466	940				41	1185			
8	7621	1645	0.7837	0.8224	0.5265	7976	1280	0.8397	0.8617	0.6948	8335	1101	0.8877	0.8833	0.9217
	661	735				429	977				96	1130			
9	7820	1446	0.8157	0.8439	0.6282	7923	1328	0.8279	0.8564	0.641	8390	1046	0.8972	0.8891	0.9592
	519	877				506	905				50	1176			
10	7700	1566	0.7831	0.8310	0.4656	7966	1285	0.835	0.8611	0.664	8359	1082	0.8913	0.8854	0.9369
	746	650				474	937				77	1144			
Mean			0.8010	0.8388	0.5592			0.8318	0.8583	0.6579			0.8916	0.88537	0.94285
St Dev			0.0121	0.0078	0.0492			0.00338	0.00229	0.0154			0.00261	0.00164	0.01329

Table 3: 10-Fold Cross Validation Estimation of Three Models

Results in table-3, describe significant and consistent performance enhancement using wFDT as compared with crisp decision trees. Figure-13, describe the Receiver Operating Characteristics (ROC) for FDT (blue) and wFDT (red).

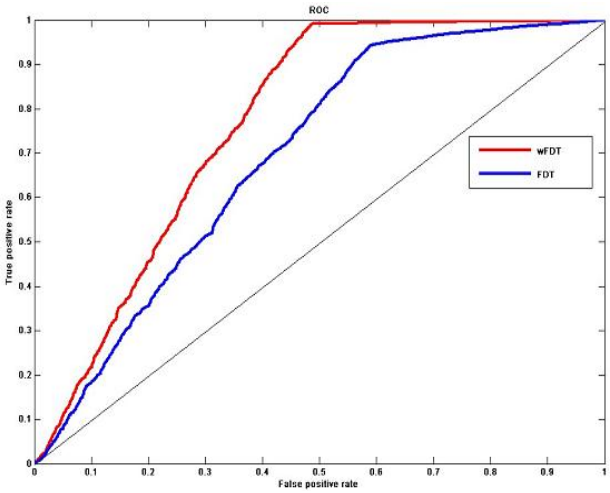

Figure 14: ROC Curve Analysis of FDT and wFDT

	AUC
FDT	**0.69**
wFDT	**0.77**

Table 4: AUC Measures

In the results presented in table-3, the variance of crisp decision tree and weighted fuzzy decision trees needs to be analysed. There is an obvious uncertainty in crisp decision tree performance. On the other hand, the estimations of FDT and wFDT are consistent through out the 10 folds. This high variance of decision tree estimations is due to sharp, inflexible decision boundaries.

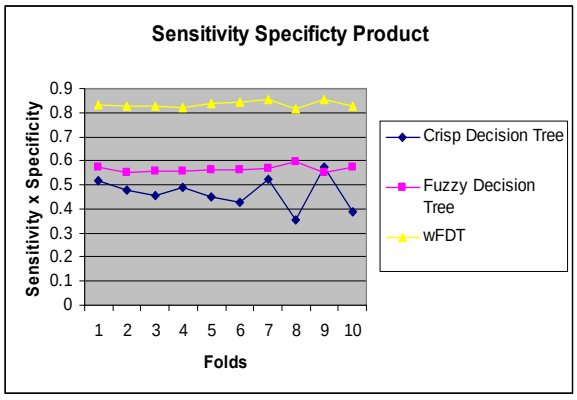

Figure 15: Sensitivity Specificity Product

Figure-14 graphically depicts the value of wFDT to a clinician who is now confident about the survival chances of patient, with such a high sensitivity specificity product. Again consistency (variance) of three curves should also be noticed.

The effect of weights on the performance of fuzzy decision trees has become obvious. There is a significant increase in accuracy of wFDT as compared to FDT. High performance of wFDT in terms of specificity proves its robustness specially when there is some bias like 'class imbalance problem' or variance due to sampling bias. Table-5 shows that wFDT performed best for Yager inference and sigmoid membership function.

Inference Tech.	Linear	Sigmoid	Convex	Concave
ZADEH	**11.5**	**10.40**	**12.56**	**13.01**
YAGER	**10.45**	**10.02**	**12.07**	**12.75**

Table 3: Average Error Rate on Unseen Date for wFDT

Performance comparisons suggests that weighted fuzzy decision trees have good compatibility with all the requirements and features of an accurate and comprehensible prognostic decision making system, mentioned in introduction and related work sections.

5 Conclusion

In this paper, we have shared our experiences of investigating intelligent machine learning techniques for breast cancer prognosis analysis. We analyzed the possible potential of fuzzy logic based classifiers, and came up with a conclusion that they are fit to act as natural allies of a physician involved in predictive medicine. Moreover, they can proficiently manage contrasting requirements of accuracy, interpretability and balance in decision. When we say balance, obviously it is not crisp. Interesting cooperation between DTs and Fuzzy theory helps to realize this aim.

After these experiments, we outlined some future dimensions which can help wFDTs to prove their potential as a strong classifier and predictor in cancer prognosis. Optimization through rule weights or genetic algorithms; an analysis is required, since rule weights, domain delimiters and inference parameters are the key players affecting accuracy. Cooperation among rules in decision making process can also be a good area research in this perspective.

We are committed to explore the strengths of wFDTs for personalized predictive medicine, which is indeed a growing trend in personalized healthcare.

Acknowledgment

This research is supported by Foundation of ubiquitous computing and networking (UCN) project, the Ministry of Knowledge Economy (MKE) 21st Century Frontier R&D Program in Korea and a result of subproject UCN 08B3-S2-10M. Moreover, Hyunjung Shin would like to gratefully acknowledge support from Post Brain Korea 21 and the research Grant from Ajou University.

6 References

Amir A, Evans DG, Shenton A, Lalloo F, Moran A, Boggis C, Wilson M, Howell A (2003): Evaluation of breast cancer risk assessment packages in the family history evaluation and screening programme. *J Med Genet 40: 807–814*

Brenner H, Gefeller O, Hakulinen T. A computer program for period analysis of cancer patient survival. Eur J Cancer 2002;38(5):690—5.

Bundred NJ. Prognostic and predictive factors in breast cancer. Cancer Treatment Rev 2001;27:137—42

Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Monterey, CA: Wadsworth & Brooks/

Cole Advanced Books & Software; 1984.

Cox DR. Analysis of survival data. London: Chapman & Hall;1984.

Carlos Andres, Pen a-Reyes, Moshe Sipper: A fuzzy-genetic approach to breast cancer diagnosis. Artificial Intelligence in Medicine 17 (1999) 131–155

C. Carter, J. Cartlett, Assessing credit card applications using machine learning, IEEE Expert, Fall Issue (1987) 71–79.

Dursun Delen*, Glenn Walker, Amit Kadam (2004): Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial Intelligence in Medicine (2005) 34, 113—127*

D. Nauck and R. Kruse, "How the learning of rule weights affects the interpretability of fuzzy systems," in Proc. 7th IEEE Int. Conf. Fuzzy

Systems, Anchorage, AK, May 4–9, 1998, pp. 1235–1240.

Earl Cox (1993) The Fuzzy Systems Handbook: A Practitioner's Guide to Building, Using, and Maintaining Fuzzy Systems

H. Brenner, O.Gefellor (2002): A computer program for period analysis of cancer patient survival. *Eur J Cancer 2002;38(5):690—5.*

H. Al-Attar, Improving the performance of decision tree induction in non-deterministic classification domains, M.Phil. Thesis, Manchester Metropolitan University, Manchester, England, 1996.

Ilias Maglogiannis, Elias Zafiropoulos and Ioannis Anagnostopoulos (2007): An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers. *Journal of Applied Intelligence*.

Ian H. Witten and Eibe Frank. Data Mining: Practiacal Machine Learning Tools and Techniques, 2nd Edition. San Fransisco:Morgan Kaufman; 2005

J. Quinlan, Induction of Decision Trees, Machine Learning, vol. 1, Kluwer Academic Press, Dordrecht, 1986 pp. 81–106

Joseph A. Cruz, David S. Wishart (2006): Applications of Machine Learning in Cancer Prediction and Prognosis. *Cancer Informatics*

J.R. Quinlan, Probabilistic decision trees, in: Y. Kockatoft, R. Michalshi (Eds.), Machine Learning, vol. 3: An AI Approach, 1990, pp. 140–152.

J.R. Quinlan, Improved use of continuous attributes in C4.5, J. Artificial Intelligence Res. 4 (1996) 77–90.

Keeley Crockett, Zuhair Bandar, David Mclean, James O'Shea: On constructing a fuzzy inference framework using crisp decision trees, Fuzzy Sets and Systems 157 (2006) 2809 – 2832

Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Wermter S, Riloff E, Scheler G, editors. The Fourteenth International Joint Conference on Artificial Intelligence (IJCAI)San Francisco, CA: Morgan Kaufman; 1995. p. 1137—45.

L. Sison, E. Chong, Fuzzy modeling by induction and pruning of decision trees, IEEE Symposium on Intelligent Control U.S.A., 1994,pp. 166–171

Lotfi A. Zadeh The role of fuzzy logic in the management of uncertainty in expert systems," Fuzzy Sets Syst., vol. 11, pp. 199–227, 1983.

L. I. Kuncheva (2000): How good are fuzzy if-then classifiers, IEEE Trans. Syst., Man, Cybern. B, vol. 30, pp. 501–509, Aug. 2000.

M. Umano, H. Okamoto, I. Hatono, H. Tamura, Generation of fuzzy decision trees by fuzzy ID3 algorithm and its application to diagnosis by gas in oil, Japan–U.S.A. Symposium, 1994, pp. 1445–1450.

Mamdani EH (1974) Applications of fuzzy algorithms for control a simple dynamic plant. In: Proceedings of the IEEE 121(12):1585–1588

National Cancer Institute USA (2008): Breast Cancer Statistics http://www.cancer.gov

O. Cordón, F. Herrera, L. Magdalena (Eds.), Trade-off between Accuracy and Interpretability in Fuzzy Rule-Based Modelling, Studies in Fuzziness and Soft Computing, Physica, Heidelberg, 2002.

O. Cordón, F. Herrera, A proposal for improving the accuracy of linguistic modeling, IEEE Trans. Fuzzy Systems 8 (3)(2000) 335–344.

U. Bodenhofer, P. Bauer, A formal model of interpretability of linguistic variables

Yijun Sun, Steve Goodison, Jian Li, Li Liu and William Farmerie (2007): Improved breast cancer prognosis through the combination of clinical and genetic markers. *Bioinformatics, Vol. 23 no. 1 2007, pages 30–37*

Y. Jin, W. von Seelen, B. Sendhoff, An approach to rule-based knowledge extraction, in: Proc. IEEE Conf. on Fuzzy Systems, Anchorage, Alaska, 1998, pp. 1188–1193.