Predictive Dynamic Risk Mapping and Modelling of Patients Diagnosed with Bladder Cancer

Department of Automatic Control & Systems Engineering

The University of Sheffield

By:

Olusayo Obajemu

Supervisor:

Prof. Mahdi Mahfouf

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

September 2016

Contents

List of Figures

List of Tables

Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Networks
AUC	Area under *ROC* curve
BCa	Bladder Cancer
CI	Computational Intelligence
CIF	Cumulative Incidence Function
CoxPH	Cox Proportional hazard model
CV	Cross validation
DA	Discriminant Analysis
DSM	Disease Specific Mortality
FP	False positive

FCM	Fuzzy C-Means
MF	Membership Function
OCM	Other Cause Mortality
RMSE	Root Mean Square Error
ROC	Receiver operating characteristics
SVM	Support Vector Machines
T1 FS	Type-1 Fuzzy Set
T2 FS	Type-2 Fuzzy Set
T1 FLS	Type-1 Fuzzy Logic System
T2 FLS	Type-2 Fuzzy Logic System
TP	True positive
UTI	Urinary Tract Infection

Abstract

An estimate (both quantitative and qualitative) of the level of risk of death a patient has after being diagnosed with bladder cancer (*BCa*) is often beneficial to both the clinician and the patient. The estimate can help to investigate the effects of different therapies on this risk level thereby aiding risk management decisions. The estimate can also help the patients plan their lives as well as to understand the specific lifestyle changes needed to lower the risk. But estimating this risk is often very difficult because of factors such as censoring, high dimensionality and evolving treatments during follow ups as well as erratic disease behaviour. Traditional artificial intelligence techniques do not handle these challenges adequately whilst allowing for the integration of the vast amount of expert knowledge in *BCa* management. This thesis addresses these concerns and challenges by proposing efficient methods to elicit fuzzy models from data which will allow one to better handle these complexities

with a view to more accurate and interpretable predictions.

Fuzzy logic technology is investigated throughout because fuzzy models are relatively transparent and, as humans, can relate to and as a result would facilitate their adoption by clinicians and oncologists. To handle the problem of censoring, the thesis integrates the proposed fuzzy models with the well-known Cox model. This would allow one to take advantage of the mature field of survival analysis (a branch of statistics) and fuzzy modelling so that the mathematical convenience and simplicity of the original Cox modelling framework is exploited.

It is widely understood that relapse from cancer completely changes the disease prognosis so that the 'static' model predictions on initial diagnosis become grossly inaccurate. The thesis also investigates how this dynamic information (relapse from *BCa*) can be incorporated into the modelling framework.

Results suggest that the proposed modelling frameworks successfully handle these challenges efficiently and elegantly and lead to better performances than the traditional artificial intelligence and survival analysis methods.

Acknowledgement

It is with great pleasure to acknowledge the many people that have been of tremendous help and support during the course of my PhD studies.

I would especially like to give thanks to my supervisor Professor Mahdi Mahfouf for his moral and financial support, patience and constant push which has allowed me get the very best out of my PhD studies. We exchanged more than a thousand emails during the course of my PhD studies and I could not have asked for a better supervisor. It has indeed been an honour working with you.

I gratefully acknowledge all students and staff members of the Department of Automatic Control and Systems Engineering, the University of Sheffield not only for the help, guidance and support during the PhD studies but also providing me a departmental scholarship.

I also wish to thank the Nigerian community in Sheffield for helping to

mitigate the homesicknesses while being more than 4000 miles from home.

I wish to thank my wife, Henrietta, my daughter, Ifesewa, my parents and siblings (Mayowa, Modesola and Tolulope) for their patience and understanding that the missed calls and late nights were not borne out of neglect for family virtues but for a greater goal. Thank you.

Chapter 1

Introduction

1.1 Research Background and Motivation

A number of questions immediately come to the minds of both the patient and the clinicians when the patient is diagnosed with bladder cancer (***BCa***). Is the cancer invasive or not? Is the patient of low or high risk? What type of therapy would lower this risk? What is the likelihood of a relapse from the disease during follow-up? Essentially, it is of interest to know how the disease would behave and respond to the many possible, different and dynamic therapies (and perhaps changes in lifestyle[1]) during follow-ups. Answering these questions would allow the clinicians to decide on the *BCa* management

[1] A patient smoker may decide to quit smoking on being diagnosed with ***BCa***.

pathway [1].

Traditionally, an estimate of the risk on diagnosis with ***BCa*** and treatment options are usually derived from a combination of the clinicians prior experience and standard staging mechanisms (for example the Tumour-Node-Metastasis (TNM) staging) [1]. However, this approach can often return misleading conclusions with many patients diagnosed as having aggressive cancers (e.g. *T4*, *N3*, *M1* in the *TNM* staging mechanism) beating the odds and living well beyond the clinicians' expectation and vice versa.

The advent of fast and relatively cheap computing and storage platforms have provided opportunities for keeping track of patients' and disease information during the course of follow-ups. These data, if adequately analysed, can provide vital clues and information from which useful conclusions can be drawn to help clinicians and patients better understand the likely course of the disease. ***BCa*** is one of the most common cancers[2] and arguably the most expensive to manage [1] because it is a recurring and relapsing disease requiring several courses of treatments thus requiring long term surveillance of the bladder. Gleaning out useful information from a *BCa* database can lead to significant cost savings as well help prevent the harmful side-effects of aggressive therapies (such as chemotherapy and radiotherapy) which are not often

[2]fourth most common in UK men and 11th most common in UK women [2]

needed. Such models would allow for investigating and visualising beforehand different hypothetical scenarios (such as the likely effects of different treatment types), thus allowing for a more dynamic and targeted therapy, in itself a detour from the usual generic (and usually static) treatment administered to patient sub-groups.

The wealth of experience that clinicians have gained over the many years of managing *BCa* should, however, not be discounted. Good insights may often be derived from this so called expert knowledge which are capable of integration with information gained from the *BCa* database. Not only would this allow for more robust prediction models but also for facilitating the adoption of such models by clinicians since such models tend to be transparent and easy to relate to [3].

Eliciting prediction models can be very challenging because of the peculiarities of a *BCa* database coupled with the limitations associated with existing machine learning algorithms. Typically, *BCa* databases contain noisy variables and **censored**[3] observations and include high dimensionality with unobserved or dynamic variables [4]. It is not unusual to find that, in a dataset, patients can have identical values of covariates but may surprisingly also have completely different event times. It may be that risk of death for the two pa-

[3]Censoring is a type of missing variable problem where the response variable is not known exactly.

tients is dependent on a covariate not measured at the time of data collection. Additionally, the *BCa*'s possible course is diverse, as shown in Fig. 1.1.and is not a straightforward one-end-point disease.

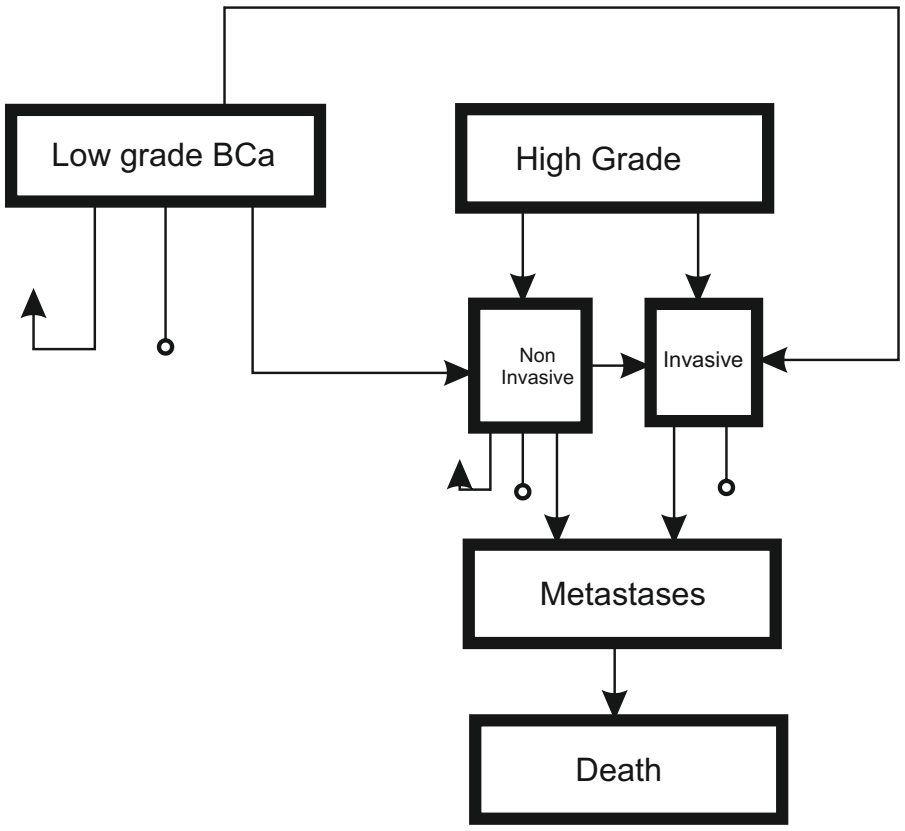


Figure 1.1: Bladder cancer may behave in a variety of ways as shown in this figure. Each line represents the possible course of behaviour after treatment. The cancer grade tells how much different the cancer cells are to normal cells. High grade *BCa* tends to grow faster and is more likely to be invasive. This is explained in greater detail in Chapter 2.

Existing machine learning technologies do not adequately manage these difficulties as well as allow for incorporation of both dynamic and expert information.

This project provides a dynamic risk-model of patients diagnosed with *BCa*. The elicited models are capable of handling the highlighted peculiarities of a *BCa* database. The mapping (or model) provides both qualitative and quantitative risk estimates for a new patient diagnosed with *BCa* and allows for incorporating patients' dynamic information relating, for example, to a relapse from *BCa* during follow up.

The developed algorithms will allow for maximum usage of the limited available information and will result in models which are open to scrutiny and available for usage by clinicians at different time points during follow-up periods .

Treatments administered (if any) on patients have been included in the design of the models. This allows for investigating the effect of these treatments and how they affect the risk of the patient.

This is a collaborative project between the Royal Hallamshire Hospital[4] and the Intelligent Systems Laboratory (*ISL*) of the Department of Automatic Control & Systems Engineering (*ACSE*), the University of Sheffield, United

[4]Sole provider of urological services in the city of Sheffield, United Kingdom.

Kingdom (*UK*). To determine and understand long term consequence of *BCa* patients, the hospital created a database of all *BCa* patients for 16 years between 1 January 1994 and 31 December 2009. This database is analysed extensively in Chapters 5 and 6.

1.1.1 Motivation

Previous and current studies on the *BCa* database have focussed on understanding the data using non-parametric competing risk analysis (see [5] for details). Specifically, the data were stratified into 12 patient groups based on age group, sex of patients and competing mortality risk analysis performed based on the cumulative incidence function (*CIF*) method developed by Scrucca in [6] to these strata in order to identify the risk groups of these strata.The study concludes that older patients tended to have higher cancer specific mortality (*CSM*) meaning that they tend to die sooner from *BCa*. Worse outcomes were reported for female patients but the study concludes that further investigations are needed to ascertain why this is the case. The major criticisms with this approach are:

- Although approximately 12 covariates were recorded for each patient at presentation with *BCa* the study stratifies on only 3 of these covariates: tumour group, age and sex. The other covari-

ates were neglected and there was no adequate accounting for the varying treatments administered to patients on diagnosis with *BCa*. The obvious criticism centres around not including the other covariates in the analysis. The more salient criticism involves the fact the stratification does not allow for personalised analysis since the patients have been grouped and variations within groups not adequately catered for.

- The second major criticism is that the study does not account for the dynamics of the study. For example, intermediate events such as relapse from *BCa* were not adequately accounted for when finding the Other Cause Mortality (*OCM*) and Disease Specific Mortality (*DSM*). As illustrated in [7], these intermediate events can considerably change the risks of a patient during follow-up and it would be beneficial to both clinicians and patient to understand such changes.

This project addresses these concerns and accounts for the heterogeneity of the patients in the database thus allowing for better personalised prognosis, treatment and subdivision into risk groups. To allow for models that are truly interpretable, accurate and systematic, the elicited models incorporate a new type-2 fuzzy logic modelling approach synergistically combined with matured

survival data techniques to result in a powerful risk grouping algorithm. And because treatments (if any) administered to patients on presentation with *BCa* are fed back as inputs to the model, the elicited model can help clinicians investigate which particular treatment would lower the patient's risk of death from *BCa*.

1.1.2 Aims and Objectives

This research aims at finding an interpretable and accurate dynamic predictive risk map for patients diagnosed with *BCa*. The steps by which this is achieved in this research are as follows:

1. The research will develop a new transparent survival analysis framework to understand the risk of death from *BCa* which takes into account missing variables (such as censoring) and complex non-linear interaction amongst the covariates/input variables. The proposed framework will integrate interval type-2 fuzzy modelling with survival analysis in a synergistic manner to exploit the strengths of both modelling paradigms.

2. The research develops a competing risk analysis modelling framework to understand the effects of competing events (death from other causes) on the estimated quantitative risk.

3. The research proposes a transparent framework to extend the above modelling frameworks so that a patient's dynamic information can be used to add knowledge from the database to achieve dynamic risk prediction. This is necessary because *BCa* may evolve and patients will move into different states during follow-up.

4. The research proposes a framework for uncertainty management and quantification in the form of a Bayesian approach modelling.

1.2 Overview of Thesis

The thesis is organised as follows: Chapter 1 provides a background motivation for the research project. It discusses the aims and objectives of the project and concludes with a summary of contributions made to existing knowledge in the course of carrying out the research project.

Chapter 2 provides a synopsis of what *BCa* is, treatment options and provides a literature survey of techniques in the literature for *BCa* prognosis using data-modelling approaches.

Chapter 3 introduces the theory of fuzzy logic. Fuzzy logic is considered in this research as it is able to model very complex systems using easy to understand human-like if-then rules. This is a natural way to accomplish the

goal of developing a transparent and accurate risk maps of *BCa* patients. To handle linguistic uncertainties in the modelling process. Higher order fuzzy logic called type-2 fuzzy logic is considered. The chapter concludes by proposing a new framework for interval type-2 fuzzy modelling elicitation which is computationally efficient and allows for the systematic determination of both antecedent and consequent parameters of the fuzzy logic system.

Chapter 4 incorporates the Bayesian reasoning into the type-2 fuzzy modelling. As will be explained in the chapter, the elicited model is capable of providing a degree of confidence in the predictions as determined by both the random uncertainties and linguistic uncertainties embedded in the modelling process. Additionally, the chapter discusses a new representation for fuzzy sets which facilitates the incorporation of the Bayesian and fuzzy reasoning. The proposed framework is tested on a challenging real-life engineering problem.

Chapter 5 explains how survival analysis techniques particularly lend themselves well to handle noisy and complex cancer data. The survival analysis theory is mature and dates back to the early 1960's when data were restricted to few data points and limited computational power. The advent of fast computing and storage platforms provide opportunities for making use of complex modelling frameworks (such as those developed in this chapter) so that maximum information can be extracted from data. The chapter discusses the

relevant theory of survival analysis and presents a transparent predictive risk modelling framework. The chapter also provides extensive analyses of the *BCa* database used in the research.

Chapter 6 extends Chapter 5 by allowing elicited models to incorporate patients' dynamic information. The proposed framework is based on a new type of interval type-2 fuzzy multistate modelling approach. The chapter also presents extensive analysis of the results from testing of the proposed framework on the *BCa* database.

1.2.1 Summary of Contributions

The summary of contribution is discussed as follows:

1. A systematic and computationally efficient framework for interval type fuzzy model elicitation is developed and discussed in chapter 3.

2. The Bayesian reasoning is hybridised with interval type-2 fuzzy modelling so that the resulting framework can manage linguistic and random uncertainties inherent in both the data and the modelling process.

3. A model (hybrid of survival data and fuzzy modelling) to deter-

mine the risk of a *BCa* patient is elicited. The model also allows for risk management decisions since it allows to investigate how different therapies affect these risks.

4. A dynamic risk model capable of incorporating new information in a bid to improve modelling accuracy is elicited.

Chapter 2

A Review of Risk Modelling in Bladder Cancer

2.1 Introduction

This chapter reviews the literature on existing data-modelling techniques for understanding cancer. The chapter begins by providing a brief introduction to cancer and defines relevant cancer terms which facilitate understanding subsequent chapters. Existing approaches for eliciting risk models from a cancer database are discussed extensively paying particular attention to bladder cancer (*BCa*) models. How uncertainties (both random and linguistic) make the modelling problem very challenging and how existing techniques may not be

tenable in addressing these challenges are discussed. The concluding parts of this chapter propose and provide the 'rationale' for the new modelling frameworks elicited in the remaining parts of the thesis which are able to handle the peculiarities of *BCa* risk modelling problem.

2.2 Carcinogenesis or the origin of cancer

Cells are fundamental building blocks of life and are created when an old cell divides into two or two cells fuse together (as do sperm cells with eggs). Every cell has a life cycle. During the life cycle, the living cells may divide and make more cells (mitosis and meiosis), go into inactive state (replicative senescence) or the cells may simply die (apoptosis) [8].

Cancer[1] begins when an individual's[2] normal cells deviate from their planned and systematic life cycle (apoptosis) to follow their own proliferation sequence. All cells produced from these abnormal cells inherit this abnormality. Consequently, what follows is a mass of abnormal cells called a **tumour** which may remain in the original tissue (called **in-situ cancer**) or may invade neighbouring tissues (called **invasion**) or spread in order to attack parts and organs

[1] Cancer derives its name more than 2300 years ago when *Hippocrates* observed the surfaces of some cut solid malignant breast tumours resembles a crab. From this observation came its name Carcinoma/Carcinos (Greek word for crab) which was translated to Latin by Celsius into **CANCER** [9].

[2] It is now clear that cancer cells arise from the body's own tissue and are not due to foreign tissues introduced to the patient's body by say an infection.

of the body (called **metastasis**[3]). Tumours which are able to metastasise are called malignant tumours. The stages of tumour development is a multi-step process called tumourigenesis and is described as follows:

1. Genetic damage causes successive mutations in a cell's deoxyribonucleic acid (*DNA*). These mutations alter key cellular functions such as suppressing programmed cell deaths or increasing cell division potentially leading to a cancerous cell [4]. These damages mainly result from an exposure to carcinogens (called somatic mutations) or inherited mutated versions (from the germline [5]) of some genes and this results in the so-called **familial tumours.**

2. These altered cells and their progenies grow and divide rapidly (a condition called hyperplasia).

3. Continuous division of the abnormal cell's progenies to form a tumour (called dysplasia).

4. These tumours are called *in-situ* tumours at first because the tumour resides in the same tissue the abnormal cells started to grow

[3]Metastasis refers to the ability of cancerous cells to invade lymph nodes and use blood vessels as highways to spread to parts of the body to invade normal tissues of organs completely different from where the cancer originate.

[4]For example, some cancers result from decreased apoptosis which means a mass of cell develops overtime.

[5]Germline mutation refers to inheritable mutations in the germ cells. The germ cells are cells destined to become sperm/egg or zygote (at the single cell stage)

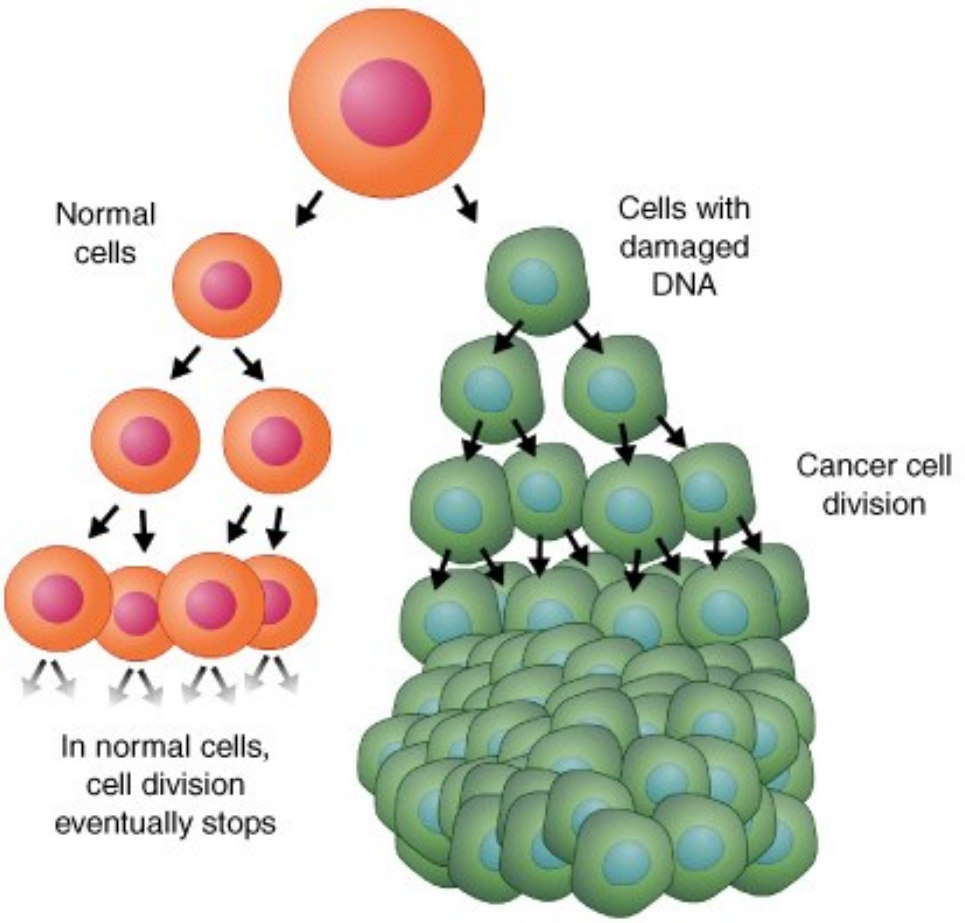


Figure 2.1: Normal Vs. Cancerous Cell Divisions. As can be seen, cancerous cells deviate from normal cells' programmed cell death (apoptosis) which results in proliferation of these (abnormal) cells [10].

in. Though this tumour may remain contained indefinitely in the same tissue, it is not unusual for tumours to invade other nearby tissues.

5. Some of these abnormal cells undergo even more mutations to become malignant which allow them to invade other tissues, lymph nodes and pass through the blood stream to invade, settle, attack and form more tumours in completely different organs and systems of the body. For example, Lymphatic spread of *BCa* involves infiltration of the para-ortic and ilac lymph nodes with the spread involving vital organs such as the lung, liver and adrenal gland. The cancer is said to have metastasised. Metastasised cancers are the most dangerous and carry the worst prognosis!

It has also been observed that nearly all cancers are monoclonal in nature i.e. the cancerous masses have resulted from a single abnormally mutated cell. This means that the cancerous masses or tumours are descendants of a single ancestral cell which underwent the unfortunate mutation. Only a small percentage is poly-clonal where the resulting cancerous tumours have resulted from two or more abnormally mutated cells.

2.2.1 Bladder Cancer - Epidemiology and Aetiology

The bladder is part of the urinary tract and is a hollow muscular organ which stores urine. It has four main walls: inner lining called the urothelium/epithelium, followed by the lamina propria which is a thin layer of connected tissues, nerves and blood vessels. The next layer is a mass of muscle called the muscularis propria. The last layer is the fatty tissue mass which divides the bladder from other organs of the body [5], see Fig. 2.2.

95% of all (*BCas*) are carcinomas[6] (Urothelial Cell Carcinoma (*UCC*)) meaning they develop from the epithelial [7] cells [12] i.e. *BCa* typically starts from the urothelium.

BCa is one of the most common cancers and arguably the most difficult to diagnose and expensive to treat [12]. About 380 000 ($\approx 3\%$ of all cancers) new cases were recorded worldwide in 2008 [1]. Risk factors of BCa typically include prolonged exposure to cigarette smokes, chemicals in workplaces (such as in dye, rubber, textile, leather factories), exposure to arsenic, previous *BCa* diagnosis, genetic predisposition, age, gender and low fluid intake. Smoking tobacco is the most common risk factors of *BCa* and it accounts

[6] Lymphomas, Leukaemias and Sarcomas are some of the other types of cancers which occur in the lymph nodes, bloodstream and fat/bone/muscles respectively.

[7] Epithelium is one of four tissue types in animals. Epithelial cells line the surfaces of all organs in humans. Cancers arising from organs such as bladder, lungs, colon, prostate and breast are typically carcinomas

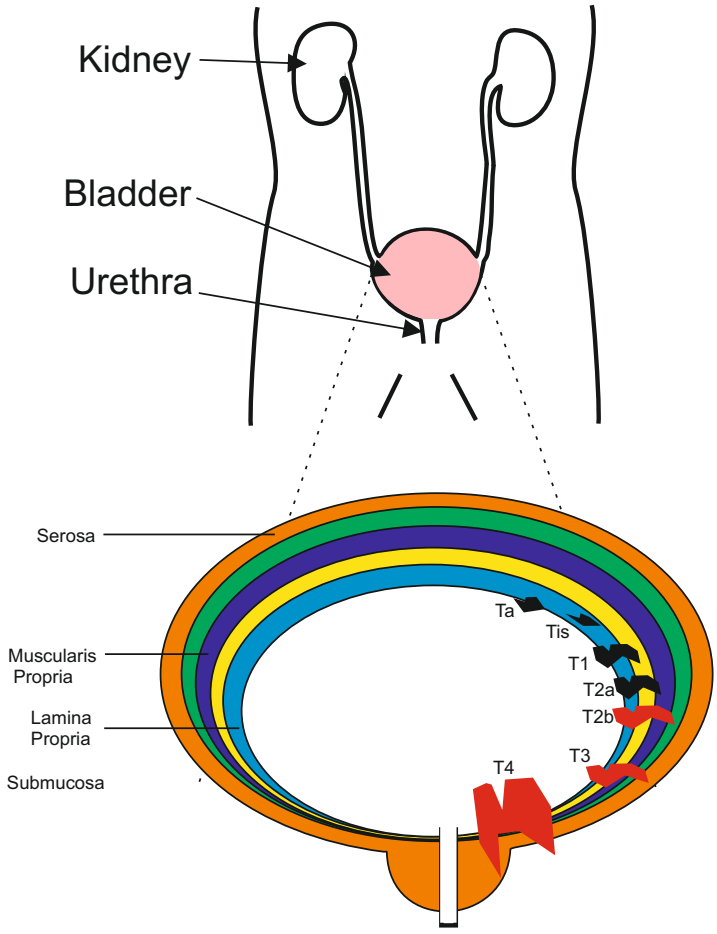


Figure 2.2: The Bladder consists of four main layers [11]. Bladder cancer typically starts from the innermost layer (the urothelium). According to the *TNM* classification [11], non-muscle invasive tumours consist of T1, Ta and Tis. The stages of *BCa* in increasing order of advancement/aggressiveness are: **CIS, Ta, T1 T2, T2a, T2b, T3, T3a, T3b, T4, T4a, T4b**.

for about 50% of *BCas*. The study in [13] shows that fruit and vegetable intake lowers the risk of *BCa* but the link between dietary and *BCa* generally remains controversial [1].

As can be seen in Fig. 2.2, as *BCa* moves from the urothelium into other layers, it becomes more progressive and more dangerous. *BCas* are often described by how far they have progressed into neighbouring layers of the bladder. **Non-Invasive** is where the cancer cells are still within the urothelium and have not grown deeper. Superficial/non-muscle invasive means the diagnosed *BCa* is non-invasive and has not progressed to the muscles of the bladder. **Invasive** means the cancer cells have grown into the lamina propria or perhaps even deeper into the outer muscle layer. **Metastatic** is the most advanced and refers to when the cancerous cells have spread through lymph nodes to attack other organs and tissues of the body.

2.2.1.1 Testing for Bladder Cancer

Diagnosing *BCa* is symptomatic and based on history and clinical investigation of the patient. The most common presenting symptoms are painless and intermittent visible (sometimes invisible) haematuria (blood in the urine) and recurring urinary tract infections (*UTIs*). However, haematuria and voiding symptoms may be present with other bladder diseases, therefore differential

diagnosis such as prostate cancer and urinary tract calculi are usually considered. When *BCa* is suspected, the oncology provider typically performs urinalysis, abdominal investigation, digital rectal examination and cystoscopy. Cystocopy involves inserting a cystocope (a small flexible lens tube) into the bladder through the urethra. If a tumour is suspected present in the area, the tumour is removed and cell samples sent off to be 'biopsied'. Biopsy means looking into these sample tumour cells under a microscope. With advances in technology, imaging tests such as the computerized tomography urogram, are increasingly being used. These advanced forms of diagnosis can locate the site, the size of tumour and any evidence of metastasis or hydronephrosis. It is also normal to test other organs of the body to see if the *BCa* has metastasised. During a biopsy, the pathologist typically searches for poorly defined cell boundaries, disorganised arrangements, irregularly shaped nuclei/ cells as well as variations in cell shape and size. This would inform the pathologist on the grade, the stage,the type of cancer, its invasiveness and if metastasis has occurred.

2.2.1.2 Management and Treatment Options for Bladder Cancer

Treatment decisions are usually made by taking into account several factors which include patient and disease characteristics. The grade of *BCa* tells how

much different the cancer cells/tissues are to normal cells/tissues. High grade tumours tend to grow faster than low grade tumours. The *BCa* stage refers to how deep the cancer cells have progressed from the urothelium or whether they have spread to other organs of the body. The *TNM* staging is the staging model of choice by most oncology providers. It tells the size of the tumour ***(T)***, whether the cancerous cells have spread to nearby lymph nodes ***(N)*** and if metastasis ***(M)*** is present. The *T* is usually found from the grade after biopsy (Fig. 2.2) [12]. Some of the treatments [14] typically administered include:

- **Transurethral resection** refers to the surgical removal of the tumour and the surrounding tissue. This is a more common treatment for the earlier stages of *BCa*.
- **Cystectomy**, which refers to *total* or *partial* removal of the bladder, sometimes along with the lymph nodes. This treatment is typically administered in more advanced *BCas* and is the recommended treatment for muscle-invasive tumours.
- **Adjuvant intravesical Chemotherapy** (mytomin C) is performed mainly on patients with muscle-invasive *BCa* to prevent it metastasising. Systemic chemotherapy is used to manage distant metastatic

BCas (*M1*).

- **Radiotherapy** uses high energy beams aimed at the cancer cells with the intention of killing them. Often used when cystectomy is not desired and the patients whose disease is still in infancy. Sometimes also used after surgery to kill cancer cells that may 'go' undetected.

Sometimes, when *BCa* has been treated, it may reoccur usually with a significantly worse prognosis. Then, the *BCa* is said to have relapsed.

2.3 Modelling Literature

The availability of electronic medical records and technological advances in data capture has presented opportunities and challenges for understanding *BCa* using artificial intelligence and machine learning. This offers numerous potentials to:

- Identifying in advance the section of the population more susceptible to *BCa* thus helping in early diagnosis [15].
- Identifying, after diagnosis, the patients at greater risk of death so that, perhaps, the patient is made to undergo a more aggressive therapy [16].

- Providing recommendations for personalised/precision medical interventions [17, 18, 19].

The next subsections discuss the opportunities that these techniques present, survey the literature on how they have been used and discuss the challenges of eliciting models especially from a *BCa* database.

2.3.1 Machine Learning and Data Mining in Medicine and Healthcare

The use of data mining and machine learning techniques has grown significantly recently due to availability of very fast and relatively inexpensive computing platforms. These techniques have helped to uncover hidden meanings and in understanding of large biomedical datasets. This has consequently helped in uncovering new information and providing recommendations for decisions as well as in generating new hypotheses from relatively large experimental biomedical datasets. The following survey discusses the most widely used machine learning techniques applied on the most important aspects of medicine.

Broadly, machine learning and data mining techniques applied to medical sciences may be classified as either descriptive (unsupervised) or predictive (supervised) [20, 21, 22, 23]. The former is exploratory in nature and at-

tempts to cluster the data so that observations are grouped according to how similar they are. This leads to discovering previously unknown patterns and relationships in the database thus helping in knowledge extraction. Descriptive techniques include clustering, summarization, sequence discovery and association [24]. The latter attempts to garner insights from labelled data so that prediction is possible from the elicited models [24]. Predictive techniques are further broadly classified into two: classification and regression. Classification is for when the explanatory variable (output variable) is discrete and the objective of the process is to elicit models capable of predicting which group an arbitrary future observation belongs to (an observation can only belong to one of K groups). For example, consider the objective of finding if a patient diagnosed with *BCa* is either of low or high risk of death from the disease. The output variable being the risk of the patient could be one of two groups; low or high. Regression, in contrast allows, for the possibility of the output variable being a continuum. For example, consider the problem of predicting the time until a patient relapses from BC. Time being a continuous variable makes the problem a regression problem [25]. Descriptive methods have been widely used in the medical literature. Clustering, in particular, has been the technique of choice in micro-array analysis because very little is known about the gene data. Co-expressed genes are usually clustered/grouped such

that they are co-regulated and in many cases share biological functions [26]. The co-expressed genes consequently help in identifying meaningful information which was previously unknown [27]. One of the seminal applications of understanding micro-array data using clustering is Van't Veer's analysis of 98 primary breast tumours using hierarchical clustering [28] where the goal was to establish the likelihood of relapse. A thorough survey of the applications of machine learning techniques in medicine is given in [29, 20].

Predictive techniques (classification and regression) are also widely used in medical applications. Classification, for example, has been used in the diagnosis of diseases [30, 31, 32, 33, 34] and to predict outcomes of diseases in medicine [35, 36, 37]. The artificial neural network (*ANN*) is arguably the most popular predictive algorithm used in both classification and regression (see Fig. 2.3). It has very powerful interpolation capabilities especially when there is a large amount of data. However, many health professionals find understanding how the decisions are reached perplexing because of its relatively opaque nature which has resulted in limited adoptions in practice. Another popular classification algorithm is the decision tree because it is very intuitive and fast. A decision tree algorithm, introduced by Ross Quinland in 1979, also known as the Iterative Dichotomiser (*ID3*) [38, 20] and its variants including the *C4.5* algorithm [39], remain the most popular decision tree algorithms

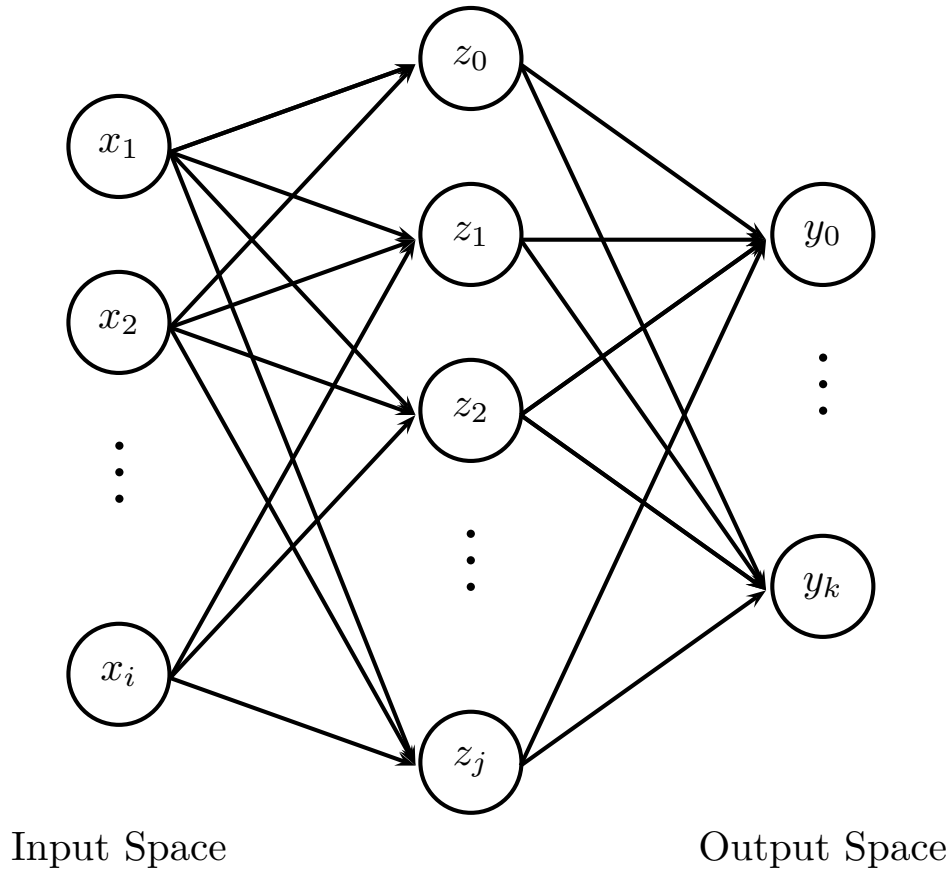


Figure 2.3: Artificial neural network block diagram.

[20] (Fig. 2.4). Decision trees are ideal candidates for data with small features. However, when the data include very large number of features, then decision trees may often result in overly complex models which are difficult to understand. This makes them unsuitable for analysing the so called 'big data'.

Support vector machines (*SVMs*) are perhaps the most popular and pow-

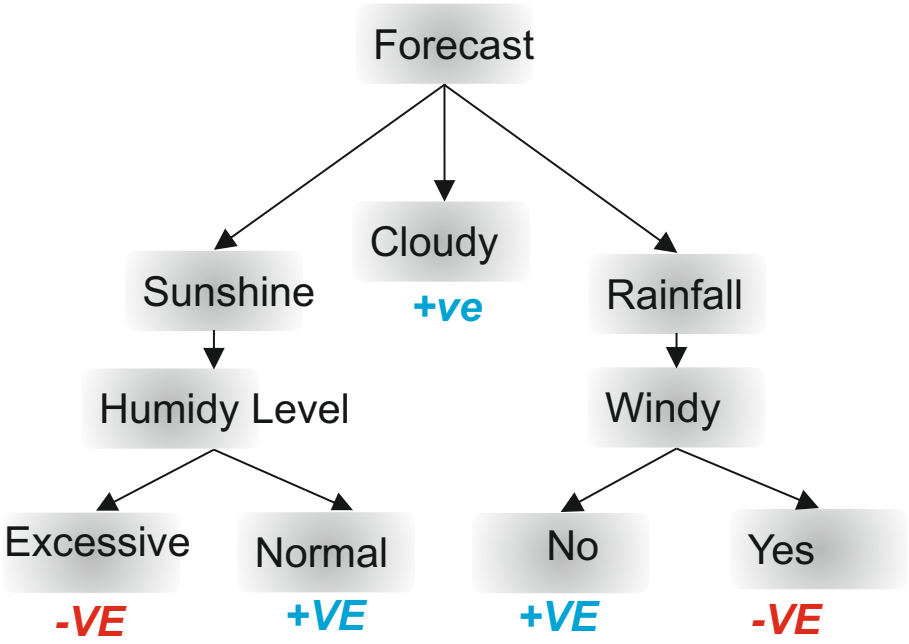


Figure 2.4: Should I play football or not? The decision tree algorithm can help answer this sort of question based on data and some selected attributes (the weather in this case). See [38] for detailed discussions of the popular *ID3* decision tree algorithm. -ve = No, +ve = Yes.

erful classification algorithm. In the 1990's, Vladimir, Vapnik and co-workers at the AT&T Bell Laboratories introduced two powerful *SVM* algorithms which use statistical learning theory to solve a binary classification problem by finding the optimal hyperplane [40]. The first [41] allows for a linear kernel function only while the second allows for more accurate classification perfor-

mances by including polynomial and radial basis function (*RBF*) [42] kernels. *SVMs* have been used to predict the muscle invasiveness of *BCa* [43] and in many other medical applications [44] [45] and are particularly suited to imbalance data.

Many cancer studies have attempted to compare various classification and regression algorithms. For example, Hu and colleagues performed a comprehensive comparative study on different classification algorithms applied to micro-array data analysis [46]. *SVMs*, decision trees and ensemble methods were performed on seven publicly available micro-array data sets (Kent Ridge biomedical dataset repository http://datam.i2r.a-star.edu.sg/datasets/krdb) with *SVM* producing the best results on most data sets after 10 fold cross validation for all the techniques. The Discriminant analysis (*DA*), logistic models, *CART* and *ANNs* were also compared in the study by Harper in [47].

A comprehensive survey of regression modelling approaches in healthcare is given by [48].

AI permeates all aspects of medicine as studied by [49] where analysis of papers submitted to the international conference on Artificial Intelligence in MEdicine (*AIME*) showed the application of *AI* techniques on a wide range of research themes ranging from ontology and terminology, planning and scheduling, uncertainty management in diverse medical fields.

2.3.2 Artificial Intelligence in Cancer Studies

For many cancer studies, a wide range of *AI* techniques has been used to infer how prognostic factors or input variables (such as age, tumour size, lifestyle) influence the outcome of the disease. Definition of this outcome is dependent on the study in question and could be subdivided into risk groups [50, 51, 52, 53], prediction of time to death [54] or time to relapse [55, 56] and status (alive or dead, relapsed) after a time period [57, 58, 59].

The next subsections review the most influential techniques for cancer prognosis with particular attention on *BCa*. The techniques are broadly divided into two categories: statistical techniques and computational intelligence techniques which will be discussed in turn.

2.3.3 Explicit Statistical Techniques

In the 17th century, the prospects of applying probability theory was widely debated amongst physicians who believed treatment interventions given to patients require only professional judgement and should not be influenced by quantitative analysis. The argument against the use of quantitative judgements in medicine was that each patient was unique and it was impossible for quantitative analyses to capture the special peculiarities of the patient and the

intended interventions[8].

However, French Mathematician, Laplace (1749-1827) who had been working on probability theory and had published many books at that time disagreed. Laplace believed that the correct treatment method would become apparent with increasing observations [61, 62, 60]. Laplace argued that probability theory provides the foundation for understanding ALL uncertainties and the peculiarities of patients can accurately be captured from the viewpoint of probability theory [63, 64]. Physician Pierre-Charles-Alexandre Louis (1787-1872) is perhaps the first recorded prominent use of probability theory in medicine which was applied to the study of typhoid fever. In his study from 1822-1827, patient data were recorded and used to determine how survival time of typhoid fever patients living in Paris differed from those living outside of Paris. Louis' conclusion that the subgroup who resided in Paris lived longer triggered further debates in France and cultivated more interests in using statistical techniques to understand diseases. It was not until the 1950's, however, that the use of statistical science recorded tremendous growth through the use of random trials (due to the father of modern statistics Sir Ronald A. Fisher) for understanding disease outcomes. It was at this point

[8]Indeed, Francois Double (1776-1842) was famously quoted as saying Denis Poisson's (1781-1840) attempt at mathematicising medical decisions were useless and Lambert Adolphe Jacques Quetelet (1796 -1874)'s use of the 'average man' akin to a "shoemaker who after having measured the feet of a thousand persisted in fitting everyone on the basis of the imaginary model" [60].

that survival analysis began to gain prominence in cancer studies particularly due to the influential surveys of Berkson and Gage [65] and Cutler and Ederer [66]. It soon dawned on oncology researchers that survival analysis, a technique which had been the mainstay of 17th century actuary and demography, could solve the particular troublesome problem of censoring[9] as exemplified by Berkson's 1950 [65] study on the survival rates of cancer patients.

Models based on survival analysis can broadly be divided into non-parametric, semi-parametric and Parametric as discussed next.

2.3.3.1 Non-parametric, Semi-Parametric and Parametric Models

Kaplan and Meier in [67] popularised the non-parametric approach to survival data modelling. This technique now popularly known as the Kaplan-Meier estimator (*KMe*) is perhaps the most widely used survival analysis technique for analysing a homogeneous population. It is usually called non-parametric because no assumption is made about the distribution of the data and the models and analysis are based on non-parametric maximum likelihood method. The Kaplan-Meier method has been used in several studies to determine *BCa* outcomes. For example, [68] used it to determine muscle invasive *BCa* subtypes' sensitivity to frontline chemotherapy. See [69] and [70] for other excellent recent applications to *BCa* prognosis. The Kaplan-Meier method was one of

[9] This is discussed extensively in chapters 5 and 6

the selected techniques used to analyse the *BCa* data in this research as discussed in Chapter 5. The use of parametric models became more pronounced in the mid 20th century in medical cancer applications through the revolutionising works of Boag [71] where the study investigated how cancer therapies influenced the five-year survival rate for different cancers using lognormally distributed models. The advantage of the parametric methods is that covariates can be seamlessly introduced to allow for a more robust analysis of a heterogeneous population. However, one has to make the strong assumption that survival times follow a particular distribution. Cox, not satisfied with this limiting assumption, introduced a technique that is arguably the most important in the medical cancer literature in [72] where the so-called semi-parametric regression model is used on the hazard[10]. The Cox modelling framework is discussed extensively in [73, 74]. The Cox model was particular influential because assumptions are kept to the minimum and this has resulted in many researchers adopting the technique particularly in *BCa* studies [75, 76, 77, 78, 79].

2.3.3.2 Bayesian Methods

The *KMe* and Cox methods are based on the 'frequentist approach' where the the goal of analysis is to find a set of parameters that best explains the data.

[10]The hazard is the instantaneous risk of death from a particular disease.

The Bayesian approach takes the view that there is an uncertainty involved in these parameter sets which should be quantified. Additionally, by using the Bayes' rule, subjectivity can be introduced in the form of a prior. The use of Bayesian analysis in *BCa* models is rather limited perhaps due to the fact that the complexities of the *BCa* database would inevitably lead to analytical intractability of the resulting Bayesian models. Numerical methods like the Markov Chain Monte Carlo [80] and Variational methods [81] could also be prohibitively expensive. However, as discussed in [82], the Bayesian paradigm offers the interesting and important capability to quantify and understand the uncertainties in predictions. Additionally, in many cases, the frequentist approach is a special case of Bayesian view, with the Bayesian model offering a more general approach to data analysis. In Chapter 4 of this thesis, the Bayesian paradigm is combined in a new manner with interval type-2 fuzzy modelling resulting in a model capable of handling and quantifying both linguistic and random uncertainties embedded in the modelling process.

2.3.3.3 Other Statistical Approaches

Statistical methods in medical studies and especially in cancer studies continue to evolve. An active research area is the use of counting processes to survival data modelling introduced in Aalen's PhD thesis in 1975. This approach

remained unexplored for many years but enjoying attention in recent years in the medical literature [83]. The approach allows for simplified analysis and particularly extends well to multistate processes (the subject of Chapter 6 of this thesis). The counting processes methods and general stochastic processes methods are excellently discussed in books by Fleming in [84], Andersen *et al* in [85] and in particular Aalen *et al* in [86].

2.3.4 Computational Intelligence Techniques

Computational intelligence (*CI*) refers to a set of a nature-inspired algorithms developed to address very complex problems [87]. These algorithms, especially those based on fuzzy logic [88], artificial neural networks (*ANN*) [89] and evolutionary computation [90], have been applied successfully to a wide range of applications ranging from engineering [91], bio-medicine [92, 93] and finance [94]. They have proved particularly successful in medical applications because of their ability to work with incomplete knowledge or uncertainty. In urological oncology, for example, they have been successfully applied in optimising diagnosis [95], staging [96], determining novel biomarkers [55, 97, 98] and prognostic prediction [99] as well as performing data mining on large micro-array data sets [100]. In *BCa* prognosis, *CI* (fuzzy logic) has been used to predict *BCa* behaviour [101]. Catto *et al* in [55] com-

pared *ANN*, Logistic regression (*LR*) and an Neuro-Fuzzy Modelling (*NFM*) techniques in order to determine prognosis after a 5-year period. They found that the two *AI*-based (*ANN* & *NFM*) methods outperformed the *LR* but noted that the *NFM* method may be more beneficial since it provides a transparent model. In their subsequent studies, using larger set of variables, they found the *NFM* to be more accurate [101] than both *ANN* and *LR*[11]. In terms of comparison expert prior knowledge with *AI* techniques, Qureshi *et al* in [91] found *ANN* to be more accurate for muscle invasive disease (62% vs 82% mortality accuracy) but was less accurate (72% vs 79% recurrence mortality) in predicting superficial cancer behaviour. It would, however, be interesting to compare the accuracies of these approaches with an hybridised version of expert knowledge and *ANN* say in the framework of fuzzy modelling. *CI* techniques have also been successfully applied to urological imaging and treatment planning. For example, Feleppa *et al* in [102] developed an automated image analysis which is similar to those for diagnosis and radiotherapy planning in histopathology which bypasses the dependence of error-prone human operators. Maclin in [103] with 99% accuracy trained an *ANN* to accurately recognise renal ultrasound images. The *NFM* approach has also been used in automated tumour diagnosis using spectral imaging parameters [104] with

[11] Although, *NFM* is a special case of *ANN*.

more than 10% improvement in accuracy over the conventional method (nearest neighbour classifier).

In medicine as a whole, fuzzy systems have found important applications in many areas ranging from neurology, pediatry and pharmacology. Abbod *et al* in [105] have surveyed comprehensively the utilisation of fuzzy systems in medicine and healthcare.

In summary, *CI* techniques continue to be pivotal in helping to understand cancer databases because of the strengths earlier discussed. Algorithms based on fuzzy logic, in particular, relate well with these databases and medical applications because of extensive expert knowledge for which fuzzy sets allow for easy integration. The type-2 approach to fuzzy modelling can help handle linguistic uncertainties (see Chapter 3). However, in the presence of random noise (as is typical in *BCa* modelling), existing fuzzy technologies may not adequately handle these uncertainties which may result in wrong inference leading to catastrophic effects on the patients lives. It is argued in this thesis, that a synergistic combination of statistical approaches with fuzzy models would lead to significant lead to more robust and accurate models. There exist some approaches at this as discussed in the following section.

2.3.4.1 Hybrid Models

The use of techniques that attempt to hybridise statistical methods with computational intelligence methods is plentiful in the literature.

Some of these techniques include local fitting methods [106, 107, 108], where the linear model $\beta^T x$ is replaced by a locally linear function. The use of splines in conjunction with the Cox model is also ubiquitous which allows for more flexible modelling of parameter β such as the work in [109] which models hazard functions directly without the use of smoothing splines as opposed to Gu's use of smoothing splines [110]. Additive models have also been used widely where the linear parts of statistical models are replaced by a summation of functions $\sum_i f_i(x_i)$ in a similar fashion as a fuzzy model under simple assumptions and are described in detail in [111, 112, 113, 114, 115]. Extending adaptive models, [116, 117, 118] allowed f_i to be a function of time in a manner similar to time dependent variables of the Cox model [119].

2.4 Adequacy of Existing Techniques - Uncertainties and Challenges

Unlike physical systems, the complexities of medical systems make the development of intelligent decision systems not so straightforward. This is even

more so in eliciting intelligent decision support systems for *BCa* database because of inherent uncertainties, high dimensionality, complex variable interactions, missing data, censored observations. This can effectively render traditional quantitative techniques grossly inadequate and inappropriate. For example, the widely and usually robust logistic regression would fail when applied to survival modelling of *BCa* [120]. Similarly, traditional fuzzy modelling approaches though having been shown to be able to represent incomplete and approximate human knowledge seamlessly will lead to biased models if censored[12] variables are not adequately accounted for. For example, in a typical *BCa* database, it has been shown in [7] that the risk of a patient changes with time due to the new information obtained during patient follow-ups examinations. For example, initial risk estimates become invalid for a patient who has been found to have relapsed from *BCa* after undergoing the initial treatment i.e. intermediate events alter likely future behaviour the disease. Therefore, medical data typically contain time-series attributes which would bias elicited models if they were not adequately handled. Medical data are usually gathered during the course of caring for a patient. And this makes the medical data less reliable and of inferior quality compared with those from other fields. The reason for this is because patients treatments are di-

[12]This is explained in Chapter 5.

verse and patients have unique lifestyles which are not usually recorded in the databases. Additionally, in many medical data gathering process, the culture of patients and clinicians filling details of sicknesses, diseases and patient histories is still prevalent. To facilitate data mining, these paper databases must be transformed digital forms. This process presents challenges in gathering highly reliable data as follows:

- Without adequate patient data gathering process, patients and clinicians can easily not enter key details. For example, patient filling personal history during follow can easily forget or even reluctant to provide significant events during follow-ups.
- Significant data preparation is often required, many data imputation stages are usually carried out by humans and it is not impossible to find a particular person required to impute 100 of thousands patient records into the computer. Human errors are fairly common and this can corrupt the data.
- Privacy and legal issues are major concerns in the medical field. Many of legal and ethical considerations must be taken into account. The need to protect the confidentiality of patients often contrast the ability of clinicians and researchers to get highly re-

liable and quality data. Also, there is still considerable debate about the legal implications of actually implementing artificial intelligence-based treatment recommendations [20]. This can cause reluctance in implementing these models. This point further buttresses the need for elicited models to be transparent and open to scrutiny which is the main reason for considering fuzzy logic models in this research project.

The above consideration combined with the peculiarities of the *BCa* database makes necessary developing new algorithms. As will be seen in subsequent chapters of this thesis, the *BCa* database involved patient histories of disease information such as relapse. Because this information is dynamic in manner (i.e. obtained during patients' follow-ups at different time points), an algorithm capable of incorporating this dynamic information to dynamically update the predictions is, hence, needed; multistate approaches allow for this seamless integration. However, because of computational infeasibility and transparency problems, many of the approaches use simple linear models. These linear models may not be tenable when covariates influence prognosis in a complex non-linear manner. Existing fuzzy modelling approaches (though resulting in non-linear transparent models) and *CI* approaches unfortunately may fail to adequately handle and incorporate this important dy-

namic information. This thesis proposes and projects to implement a new modelling framework that synergistically combine fuzzy models and multistate approaches in analysis the *BCa* database via a dynamic risk map for new patients diagnosed with *BCa*. The elicited models exploit the strengths of the transparency and non-linear inference capabilities of fuzzy models combined with the dynamic prediction capabilities of multistate models.

2.5 Summary

This chapter has provided an extensive literature survey on existing methodologies for cancer risk modelling with particular emphasis on bladder cancer models. The challenges of eliciting such models which include problems relating to censoring, high dimensionality and complex disease behaviour has been discussed. The justification for selecting fuzzy modelling framework in a bid to address these problems has also been discussed. The next chapter will review the literature on existing fuzzy modelling frameworks with particular emphasis on the type-2 fuzzy modelling configuration. A new method for the systematic elicitation of fuzzy models from data will also be proposed.

Chapter 3

Interval Type-2 Fuzzy Logic Systems - An Approach for Systematic Elicitation from Data.

Our understanding of many complex systems is imprecise and sometimes dependent on the individual. Such subjective knowledge lacks precise mathematical basis (in contrast to objective knowledge). For example, consider the effect of systemic chemotherapy on a distant metastasised *BCa* patient: as

explained in chapter 2, the patient is considered high risk with bad prognosis. This therapy (systemic chemotherapy) is known, through experience, to lower the risk and improve prognosis for this patient. However, it is difficult to explain exactly in mathematical terms how this therapy affects the risk of the patient. Only a linguistic explanation, based on expert knowledge is available. The expert knowledge may, for example, be of the following form: *adjuvant intravesical chemotherapy tends to lower 5-year death risk on distant metastastic BCa patients*. It is such systems that fuzzy logic particularly lends itself to. This is because Fuzzy logic, introduced by Zadeh in 1965 [121], allows one to seamlessly represent such imprecise information in a principled manner as discussed in succeeding subsections. Fuzzy logic extends conventional bivariate logic by allowing partial truth. In relation to set theory, this means that an element may belong to a set partially. This contrasts with classical sets where an element can only either belong to the set or not all. The heart of fuzzy logic systems (*FLSs*) is the fuzzy set (*FS*) which declares with certainty if and to what degree an element belongs to that particular set . Debates about how this degree of membership should be represented have resulted in two broad categories of *FLSs*: the Type-1 *FLS* (*T1 FLS*) and Type-2 FLS (*T2 FLS*). These categories of *FLSs* as well as the justification for adopting the interval *T2 FLS* (*IT2 FLS*) approach in this research is discussed in this chapter. The chapter

concludes by providing a comparative study of the proposed framework with existing techniques of the same type.

3.1 Type-1 Fuzzy Logic Systems

The notion of set membership is pivotal to representing subjective knowledge using fuzzy logic. The central object in fuzzy logic is this *FS* which describes the subjective knowledge an individual has about the meaning of a word. Mathematically a *FS* in the universe of discourse X is set of ordered pairs of an element $x \in X$ and its *MF* defined as follows:

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

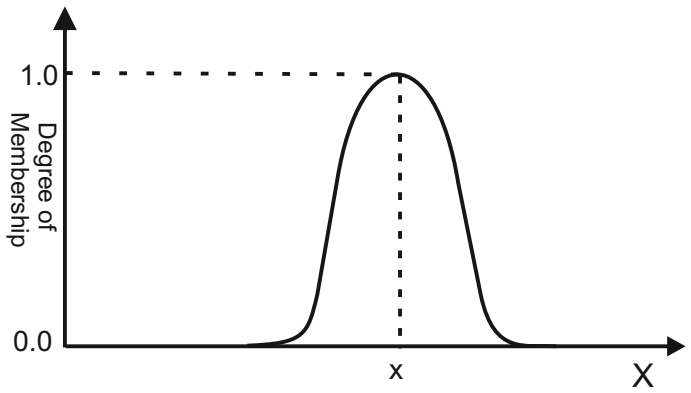
Consequently, X can viewed as the space of a linguistic variable u which may take linguistic values such as *moderately high*. The elements x in X are sometimes used interchangeably to mean a linguistic variable. Typical linguistic values for, say, the linguistic variable *height* may include *low, medium high, tall, short* which are labels easily interpreted by us humans. Therefore, it can be seen that a *FS* A generalises the *crisp set* by allowing elements of the set A defined in the universe of discourse X to have a degree of membership defined by the membership function (*MF*) $\mu_A(x)$. The *MF* is a measure that shows the degree of similarity of an element in X to the *FS* A. The concept of

a *FS* may also be represented diagrammatically as shown in Fig. 3.1.

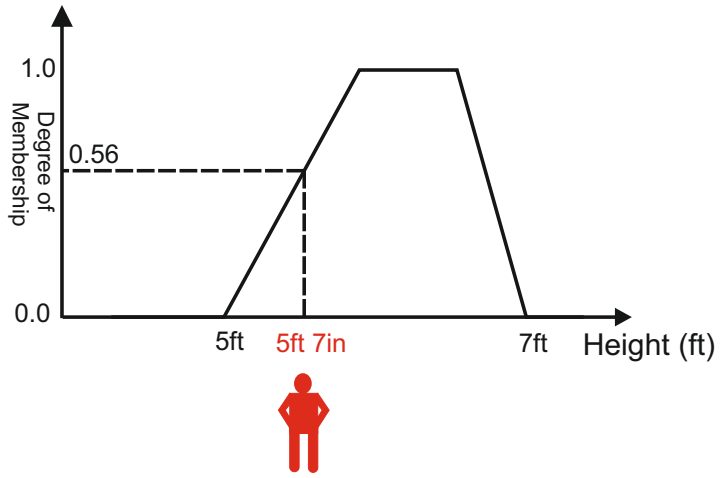
To be able to incorporate fuzzy logic and *FSs* in engineering systems, *FSs* are typically used to represent the variables of the system with inference carried out using the so-called fuzzy inference engine (*FIS*). This arrangement results in the fuzzy logic system (*FLS*) which is shown in Fig. 3.2.

The *FLS* provides a mapping from the input space $\mathcal{X}$ to the output space $\mathcal{Y}$ so that the block diagram of Fig. 3.2 can be seen as $y = f(x)$. The details relating to all component of this block diagram are as follows:

- **Fuzzification:** This process makes the crisp input quantities fuzzy. The premise behind fuzzification is that many quantities which are presumed crisp actually carry a significant amount of uncertainty. For example the quantity *about 6 feet tall* carry uncertainty in terms of vagueness of information. The ammeter usually outputs crisp current values but this could be subject to several errors which would result in imprecision. Fuzzification provides a mechanism to represent and consequently handle these types of uncertainties. Fuzzification is the mapping of a crisp value of the ith input variable, i.e. $x_i = x_i^*$, into a fuzzy number (similar to that of Fig. 3.1a). In many applications, singleton (as opposed to non-singleton) fuzzification is the preferred choice since it is



(a) Typical Fuzzy Set. The elements (x) of the universe of discourse (X) have degree of membership in that particular *FS*. This degree of membership is called a membership function (*MF*).



(b) Fuzzy Set representation of *moderately tall*. A person of height $5ft7in$ would have a degree of membership of 0.56 in this *moderately tall FS*.

Figure 3.1: Type-1 Fuzzy Sets

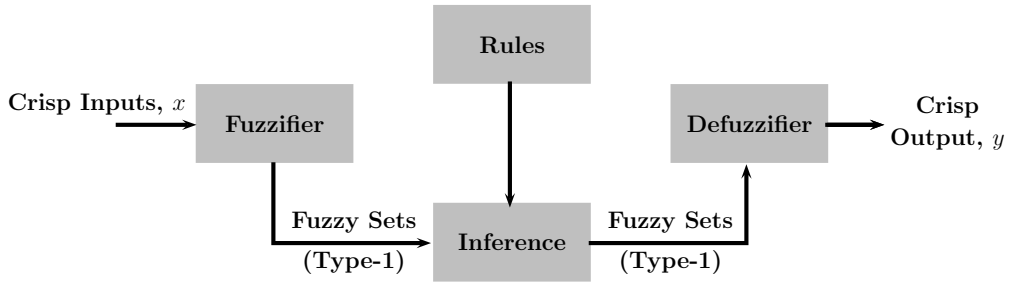


Figure 3.2: A fuzzy Logic system

more mathematically tractable and results in a significant computational cost saving. Singleton fuzzification is one where the membership value is unity at x_i^* and zero everywhere else which contrast to nonsingleton fuzzification where the membership is unity at x_i^* but slowly decreases as one moves farther away from x_i^* (again in a manner to that obtained in Fig. 3.1a). Singleton fuzzification is adopted in this research. Readers are referred to [122] for detailed analyses of the different fuzzification types.

- **Rule Base:** The fuzzy rule base provides a linguistic mapping from the input variables to the output variable. This is what makes fuzzy systems relatable to humans since the rules are IF-THEN rules. A sample rule is of the following form:

$$R^l : \text{IF} \quad x_1 \text{ is } \mathbf{A}_1^l \text{ and } x_2 \text{ is } \mathbf{A}_2^l \ \cdots \text{ and } x_n \text{ is } \mathbf{A}_n^l, \quad \text{THEN } y^l \ is \ \mathbf{B}^l \tag{3.2}$$

where A_l^l and B_l are fuzzy sets (Fig. 3.1a), $x_i, \forall i = 1 : n$ represents the ith input domain and y is the output. IF M is the number of fuzzy rules, then $l = 1, 2, \ldots, M$. The *IF* part is called the antecedent and the **THEN** part is referred to as the consequent.

- **Inference Engine:** The inference engine or the fuzzy inference engine (*FIS*) is the heart of the *FLS* and it is where fuzzy logic principles are incorporated into the fuzzy IF-THEN rules and the *FSs* (both input and output). For a *FLS* with M rules with $R^{(l)}$ being the lth rule for $l = 1, 2, \cdots, M$ and n inputs $x_1, \ldots, x_n$ (henceforth represented by the vector $\mathbf{x}$), the steps involved in the *FIS* mechanism is as follows:

  1. For each fuzzy IF - THEN rule, determine the *MF* $\mu_{A_1^l \star \cdots \star A_n^l}(\mathbf{x})$ for $l = 1, 2, \ldots, M$ where $\star$ is the t-norm operator which is usually selected to be the minimum or product. The product formulation for the t-norm operator is used in this project. The t-norm operator represents

how the **AND** part of the fuzzy IF-THEN rules are combined.

2. Calculate $\mu_{B^l}(\mathbf{x}, y) = \mu_{A_1^l \star \cdots \star A_n^l \rightarrow B^l}(\mathbf{x}, y)$. This is the THEN part of the fuzzy rule and is called **IMPLICATION**. As is the case for calculating the AND part for multiple inputs, the implication is calculated by using the t-norm which is typically a product or a minimum. Again, the product t-norm is used in this project. At this point, one is left with M *FSs* (i.e. $\mu_{B^l}(\mathbf{x}, y)$ for $l = 1, 2, \ldots, M$.

3. Combine the M *FSs* using the s-norm operator which results in a composite *FS* i.e.

$$B = \cup_{l=1}^{M} B^l = \mu'_B(y) = \mu_B^1(y), \odot \cdots, \odot \mu_B^M(y) \quad (3.3)$$

where $\odot$ is the snorm or t-conorm operator commonly implemented using max of probor [123, 124].

- **Defuzzification:** For many engineering systems, the output is usually crisp. Defuzzification produces this crisp output. Many types of defuzzification techniques exist in the literature such as the mean of maxima, centroid, center of sums and modified height

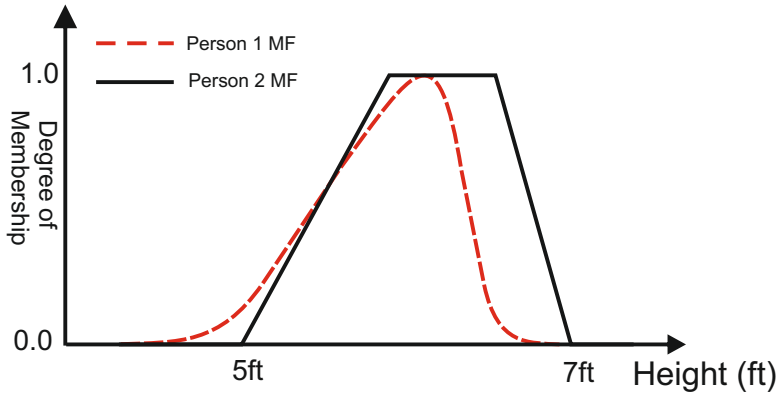
[122, 124]. The centroid defuzzification method is adopted since it is easily differentiable *FLSs* and results in smooth surfaces [125]. The centroid defuzzifier is given as follows:

$$y_c = \frac{\sum_{i=1}^{N} y_i \mu_b(y_i)}{\sum_{i=1}^{N} \mu_B(y_i)} \tag{3.4}$$

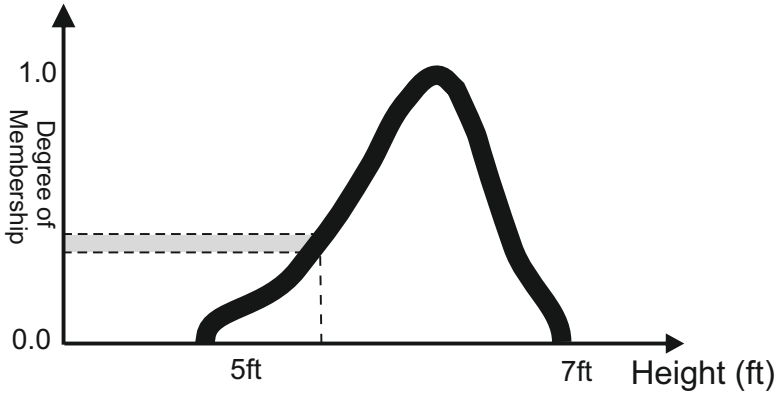
where the *MF* for the composite output *MF* calculated in step 2 above has been discretised into N points. It is worth noting that y_c is a function of the inputs $(\mathbf{x})$.

3.1.1 Challenges with Using the Conventional Fuzzy Logic System

Using the type of *MFs* shown in Fig. 3.1a for complex and highly uncertain systems may not be tenable. This is because, as extensively discussed in [122] and [126], this representation suffers from the severe limitation of oversimplifying real life scenarios. For example, it is entirely possible for two persons to have different meaning and indeed different representation for the word 'tall' in Fig. 3.1a, such that there are different degrees of membership for all or some of the elements of the *FS* in the universe of discourse. Fig. 3.3a illustrates this point. This type of uncertainty is referred to as inter-person uncertainty [122]. It is also possible for a person's meaning about a word to be ambiguous such that the crisp value of *MF* values in conventional *FS*



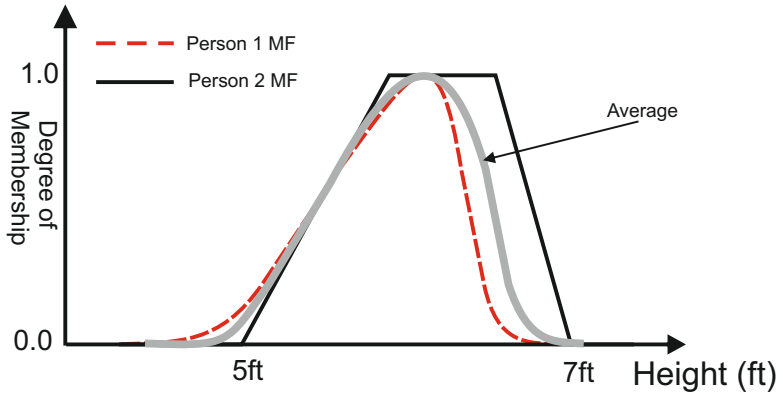
(a) Two persons asked to provide an *MF* for the word 'tall' could lead to the figures above where there is no consistency in the representation.



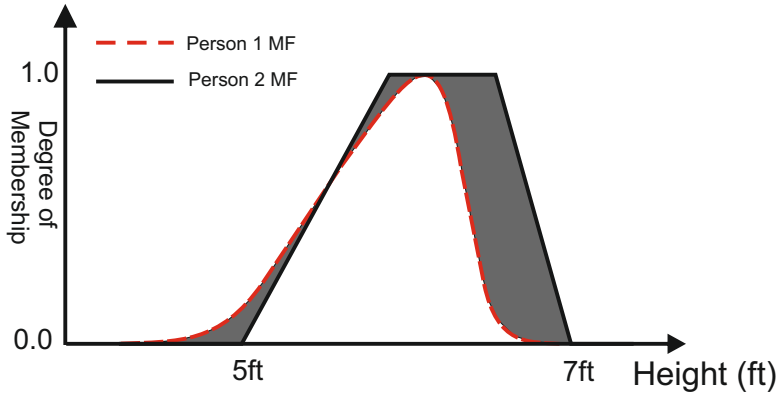
(b) Intrapersonal Uncertainty: An individuals *MF* might be ambiguous which is akin to the blurring of these *MFs* or drawing the *MFs* with a big marker. The degree of membership is no longer a crisp value but an interval.

Figure 3.3: Uncertainties about a word.

is no longer tenable. This is usually referred to as intra-person uncertainty because it represents the doubts that a person has about the meaning and representation of a word and the idea is illustrated in Fig. 3.3b.



(a) Averaging results in loss of information.



(b) *A T2 FS* allows for capturing this uncertainty about the word tall by means of the so-called footprint of uncertainty (*FOU*). The *FOU* is the region covering the primary *MF*. More region covered means more uncertainty about the word.

Figure 3.4: A Type-2 Fuzzy Set allows for capturing the uncertainties.

Earlier approaches to solving this problem include averaging [122] of the *MF* values at specific points (See Fig. 3.4a).

However, this can lead to considerable loss of information and degradation of performance in systems with high uncertainties since variations in *MF* val-

ues are not accounted for. To remedy this problem, Zadeh introduced a new type of *FS* in 1975 [88] called the type-2 fuzzy set (*T2 FS*) (Fig. 3.4b). This *FS* has a third dimension which allows for representing and handling such linguistic uncertainties. The use of *T2 FS* in applications became increasingly popular because of advances in computational power and extensive developments of its theories in the late 1990s and early 2000s [127, 128, 129, 130]. In line with the literature, the conventional *FSs* are hitherto called type-1 fuzzy sets (*T1 FSs*) with a *FLS* incorporating the *T1 FS* in all their *MFs* called Type-1 fuzzy logic systems (*T1 FLS*). Type-2 fuzzy logic systems (*T2 FLS*) refer to those *FLSs* which use at least one *T2 FS* in their *MFs* [129] and are discussed further in the following sections.

3.2 Type-2 Fuzzy Logic Systems

Mathematically, a *T2 FS*, $\tilde{A}$, is given as follows:

$$\tilde{A} \coloneqq \{(x,u), \mu_{\tilde{A}}(x,u) | \forall x \in X, \forall u \in J_x \subseteq [0,1]\} \tag{3.5}$$

where X is the primary domain of the *MF* of A and J_x is the secondary domain. u is sometimes called the primary *MF* while $\mu_{\tilde{A}}(x,u)$ is called the secondary *MF*, $0 \leqslant \mu_{\tilde{A}}(x,u) \leqslant 1$. At each value of $x = x'$, a 2D plane, which is *T1 FS* is called the vertical slice (Fig. 3.4b). The wavy slice or embedded *T1*

FS of $\tilde{A}$ is a *T1 FS* that results when only one value of u for all elements x in the universe of discourse. This representation is similar to that drawn for the average *MF* in Fig. 3.4a.

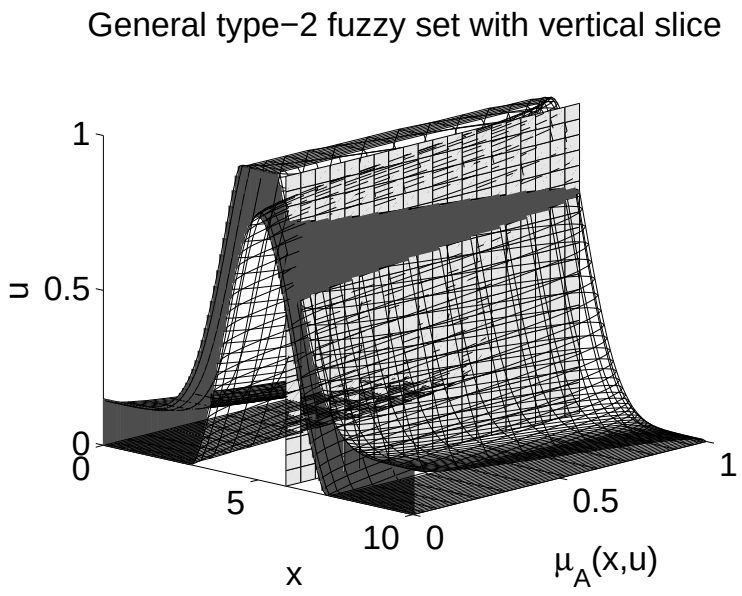
Other representations of *T2 FSs* such as the z-slice/α-slice exist. Readers are referred to [131] for a detailed analysis of the different representations of the *T2 FS*.

There are two major differences between the *T1 FLS* and the *T2 FLS* (Fig. 3.5b). The first, as already mentioned, is the incorporation of the *T2 FSs* as the *MF* of the *FLS*. The second is the presence of of type reduction (*TR*) stage which reduces the *T2 FSs* at the output side of the *FLS* into a *T1 FS* for subsequent defuzzification[1]. The *TR* stage can be computationally taxing because it involves enumerating all embedded *T1 FSs* and then computing the α tuples and computing the centroid of the centroid accordingly. α can be very large depending on the level of discretization[2], see [128] for details. This computational cost has caused a reluctance in the 'fuzzy community' to using the *T2 FLS* in practice[3]. Thankfully, there exist an efficient, precise and

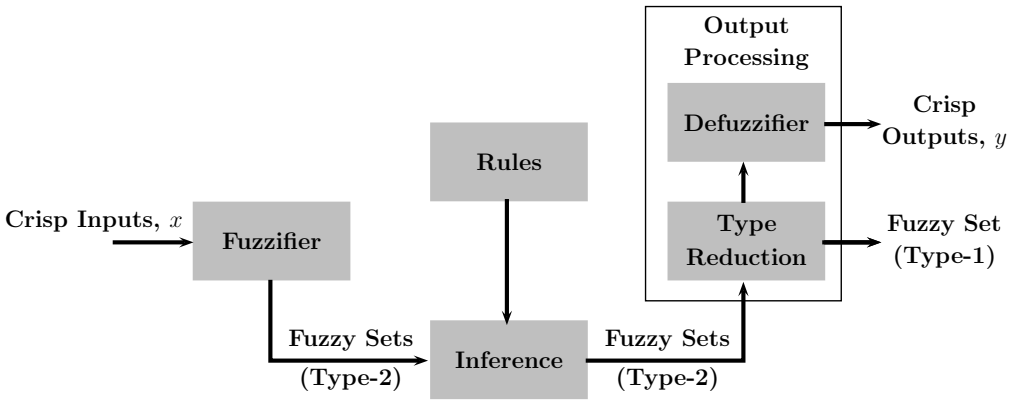
[1]The *TR* stage is a consequence of the Extension Principle [88] and is an extension of the type-1 defuzzification

[2]if the primary domain is discretized into N points and the secondary domain discretized into M_j points for $j = 1, 2, \cdots N$, then $\alpha = \prod_{j=1}^{N} M_j$. Hence, for a modest level of discretization, say $M_j = M = 10$ and $N = 10$, then $\alpha = \prod_{j=1}^{10} 10 = 10^{10} = 10$ billion.

[3]Although, attempts at reducing this computational cost is burgeoning and showing good promise, many of the developed algorithms return only approximate results. See [122] for an informative survey of *T2 FLSs* in practice.



(a) *T2 FLS* showing the secondary *MF* in *3D* and the vertical slice.



(b) *T2 FLS*

Figure 3.5: *T2 FS* and *T2 FLS*.

accurate algorithm for a special type of *T2 FSs* called the interval type-2 fuzzy set (*IT2 FS*) which is discussed in the following section.

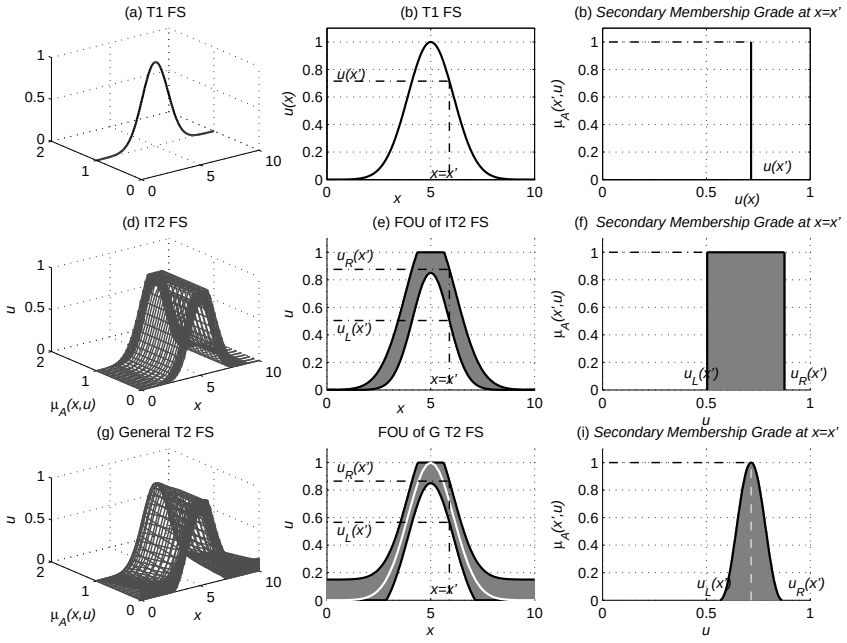


Figure 3.6: Fuzzy Sets

3.3 Interval Type-2 Fuzzy Sets

As already mentioned, the computational complexities of the *T2 FLS* may be significantly reduced by following the so-called *IT2* approach [132]. In *IT2 FSs*, the secondary *MF* is a type-1 interval as shown in Fig. 3.6 so that the secondary membership grades are all unity.

This is expressed mathematically as follows:

$$\tilde{A} \coloneqq \{(x,u), \mu_{\tilde{A}}(x,u) = 1 | \forall x \in X, \forall u \in J_x \subseteq [0,1]\} \tag{3.6}$$

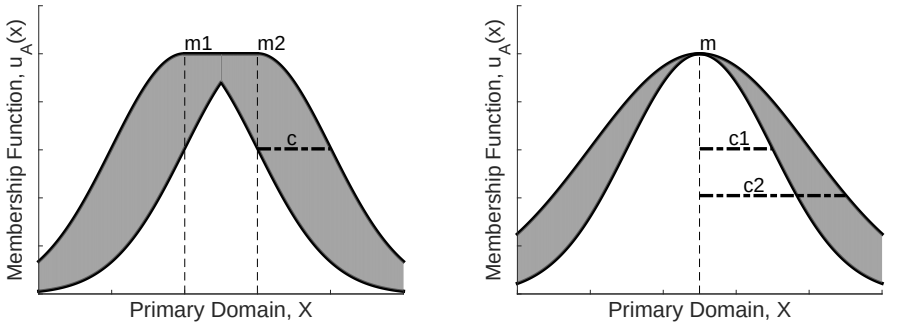
It is worth noting at this point that in an *IT2 FS*, $\mu_{\tilde{A}}(x,u) = 1 \ \forall (x,u)$ while

for the general *T2 FS*, $0 \leqslant \mu_{\tilde{A}}(x,u) \leqslant 1 \ \forall(x,u)$. The Karnik-Mendel (*KM*) algorithm allows for accurate and efficient type-reduction of the *IT2 FS*. Consequently, one can see that using the *IT2 FSs*, there is a trade-off between uncertainty handling/representation and high computational costs. It is for this reason that this *IT2* approach is considered in this thesis. To be able to deploy such *FSs* in applications, the *FSs* must be represented in a parametric form as shown in Fig. 3.7.

All other parts of the *T2 FLS* remain unchanged in the *IT2 FLS* and inference proceeds exactly as described in Sections 3.1 and 3.2.

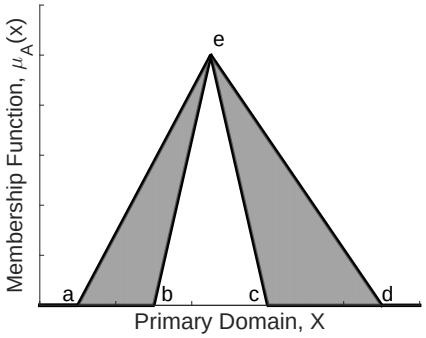
3.4 Fuzzy Logic Systems Modelling

In the absence of or limited expert knowledge, *FLSs* are usually designed using input-output data in a manner similar to how the artificial neural network is designed. The resulting fuzzy model then provides a linguistic representation of the data and the system in question. Automatic model elicitation from data of *T1 FLS* and has been well studied in the literature, see [133, 134] for a thorough review. The clustering method is the preferred method of choice because it provides a rapid, transparent and accurate prototyping mechanism [125]. Often, the model elicited from clustering is tuned further to produce a more accurate model. In the case of designing the *T2 FLS*, it is imperative



(a) Gaussian *MF* with uncertain centre (m_1 and m_2).

(b) Gaussian *MF* with uncertain width (c_1 and c_2).



(c) Triangular *MF*. a, b, c, d and e completely define this *MF*.

Figure 3.7: Examples of Interval *T2 FSs*.

that the optimisation algorithm does not incur high computational costs since the *T2 FLS* is already a costly arrangement. It is for this reason that the gradient descent algorithm has been the method of choice in tuning the *T2 FLS* from data. The gradient descent algorithm is, however, a local optima algorithm and the common method of random initialisation of parameters [135] may result in the algorithm repeatedly trapped in a local minimum. This sec-

tion presents a more systematic method for selecting these initial parameters. The section also provides a gradient descent algorithm for a new kind of *IT2 FLS* which allows for a new and relatively more efficient form of computation. The first step is to determine the structure of the system by clustering the data following a similar approach used in [136]. However, here, a source of uncertainty involving the fuzzifier, which regulates the degree of overlap amongst the clusters when fuzzy c-means (*FCM*) clustering technique, is used to elicit the initial structure of fuzzy models is taken into consideration. To achieve this, a similar procedure introduced in [137] is followed which involves making use of the interval type-2 fuzzy clustering algorithm developed [138] to elicit the initial structure of the antecedent structure of the *IT2 FLS*. This is a detour from the popular method of randomly initialising the width of the IT2 FSs after having used *FCM* for the initial structure determination with the fuzzifier value set to a constant value (usually set to 2) [135]. It is argued in this research that this systematic approach of finding these initial parameters has the potential to helping one build a more optimal fuzzy system especially when no further parameter tuning is performed. Even when the initial system is further optimised, local optima methods may return solutions that are equal or not far off from global optimum solutions [139] i.e. by making use of *IT2* fuzzy clustering, good clustering results are returned

and invariably good modelling performances. The research focusses on the Takagi-Sugeno-Kang (*TSK*) *T2 FLSs* because these are are more general than the Mamdani *FLSs*. The presented approach follows that obtainable in the method proposed in [136, 140] which involves first finding the structure of model using *FCM* clustering and further optimising these parameters. They applied this technique to a *T1 FLS* while ours is implemented on an *IT2 FLS*. The modelling paradigm involves first selecting the initial antecedent structure using *IT2* fuzzy clustering and then the least-squares approach is used to derive initial consequent parameters by making use of the Nie-Tan (*NT*) defuzzification method [141]. This approach bypasses the computationally demanding *KM* algorithm. Finally, the whole system is optimised using the gradient descent optimisation algorithm. In building this framework, some necessary assumptions have been made and this will be made clearer by some sets of equations in the succeeding subsections.

3.4.1 Methodology

As seen in Fig. 3.8, the proposed methodology has three stages which consists of initial structure determination by using a fuzzy clustering approach to determine the antecedent parameters, followed by the use of a the least-squares algorithm to determine the consequent parameters and finally optimising all

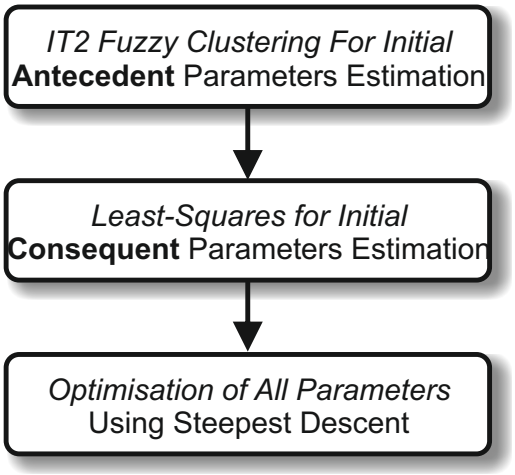


Figure 3.8: The various stages involved in the *IT2 FLS* Modelling framework.

the parameters of the *FLS* using a gradient descent approach. Each of the stages is discussed further in the following subsections.

Using the Nie-Tan defuzzification method, the output of a *IT2 FLS* may be expressed as follows:

$$f_s = \left(\sum_{i=1}^{c} h_i \overline{f}_i + \sum_{i=1}^{c} h_i \underline{f}_i \right) \Big/ \left(\sum_{i=1}^{c} \overline{f}_i + \sum_{i=1}^{c} \underline{f}_i \right) \tag{3.7}$$

where h_i is the consequent value of the ith rule which is defined as $h_i = \beta_{i0} + \beta_{i1}x_1 + \beta_{i2}x_2 + \cdots \beta_{in}x_n = \beta^{\mathsf{T}}\mathbf{x}$. β_{ip} are the consequent parameters for $p = 0, 1, 2, \cdots n$ and $i = 1, \cdots, c$. For the Gaussian primary *MF* of fixed mean and uncertain spread such as that shown in Fig. 3.7b, and using the product $T-$norm, $\overline{f}_i$ and $\underline{f}_i$ are defined as follows:

$$\overline{f}_i = \prod_{j=1}^{n} \overline{\mu}_{ij} \quad \text{and} \quad \underline{f}_i = \prod_{j=1}^{n} \underline{\mu}_{ij} \tag{3.8}$$

$\overline{\mu}_{ij}$ and $\underline{\mu}_{ij}$ are the upper and lower firing strengths respectively for the jth antecedent *MF* of the ith rule for $j = 1, 2, \cdots, n$ and $i = 1, 2, \cdots, c$.

In matrix form, equation 3.7 may be reformulated as follows:

$$fs = \frac{\bar{\mathbf{F}}^{\intercal}\mathbf{H} + \underline{\mathbf{F}}^{\intercal}\mathbf{H}}{\bar{\mathbf{F}}^{\intercal}\mathbf{r} + \underline{\mathbf{F}}^{\intercal}\mathbf{r}} \tag{3.9}$$

where $\mathbf{r}$ is a $c \times 1$ vector of $\mathbf{1}$'s and $\mathbf{F}$ is a $c \times 1$ vector. It is worth noting that the *MFs* are only associated with the antecedents since there are no consequent *MFs* for a *TSK FLS*.

3.4.2 Clustering Methodology

As mentioned earlier, in this stage of the modelling process, the initial structure and parameters of the antecedent *MFs* are obtained through the use of the *IT2 FCM* algorithm. The parameters of these antecedent MFs are found by projecting the *UMF* and *LMF* found from the fuzzy clusters to each input subspaces. This approach follows directly from that obtainable in [140] which uses this approach to find the antecedent initial parameters of a *T1 FLS*. The centres and widths (consequently the variances) are derived from the fuzzy means and fuzzy variances respectively of each input dimension after clustering. For detailed discussions of how these antecedent parameters may be found from the input subspaces projections, the reader is referred to [140] and [136]. The flowchart for the clustering algorithm, is shown in Fig. 3.9.

This is based on the interval type-2 fuzzy c-means (*IT2 FCM*) algorithm developed by Hwang in [138].

Fig. 3.10 shows a synthetic data set including two dimensions and the projections in each dimension for a fixed mean and uncertain standard deviation *IT2 FS* after using the *IT2 FCM* to cluster the data (2 clusters, 2 fuzzy rules).

3.4.3 Least-Squares for Consequent Determination

The least-squares algorithm is used to determine the initial values for the consequent part of the *FLS*. Recall from equation 3.7 the output of an *IT2 FLS* using the Nie-Tan defuzzification method, the consequent parameters are the βs. Now let

$$f_i = \underline{f}_i + \overline{f}_i \quad \text{and} \quad \psi = x_p f_i \Big/ \sum_{i=1}^{c} f_i \tag{3.10}$$

for $p = 0, 1, \cdots, n$, where n is the number of input variables. Then the output of a *IT2 FLS* can be re-expressed as follows:

$$fs = \mathbf{B}^{\intercal}\boldsymbol{\psi} \tag{3.11}$$

where $\boldsymbol{\psi}$ is a vector of ψ_{ip} for $i = 1, \ldots, c,\ p = 0, 1, \ldots, n$ and $\mathbf{B}$ is a vector of associated parameters. Least-squares parameter estimation leads to

$$\hat{\mathbf{B}} = (\Psi^{\intercal}\Psi)^{-1}\Psi^{\intercal}Y \tag{3.12}$$

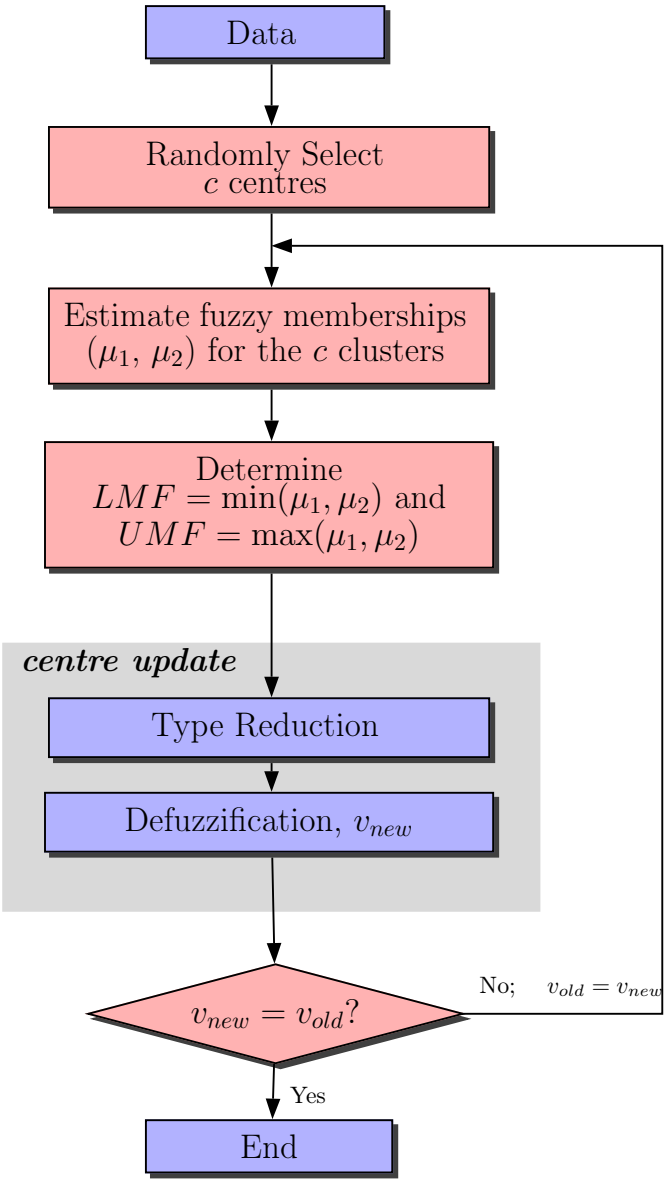


Figure 3.9: Flow Chart of the clustering algorithm.

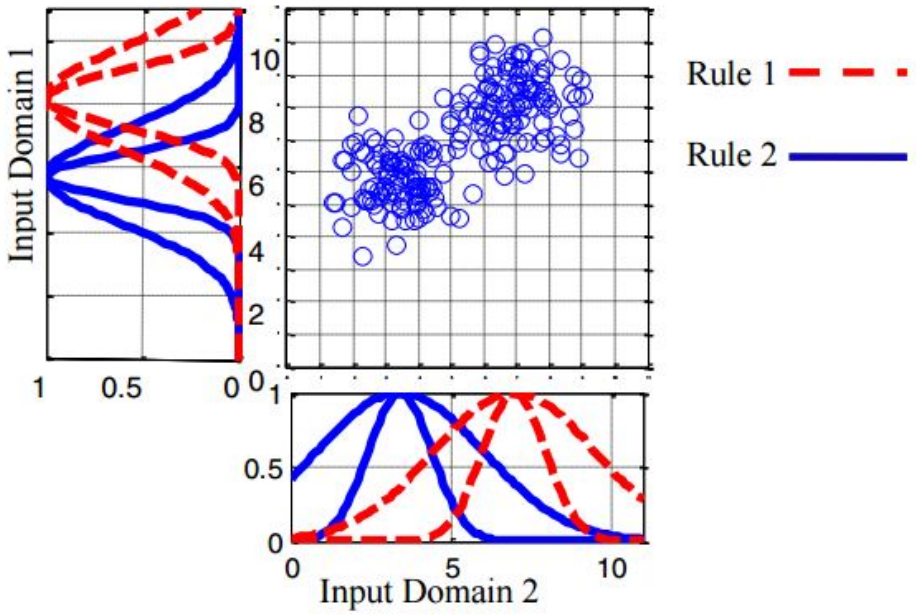


Figure 3.10: Synthetic data projection example.

where Y is a column vector of measured output data and Ψ is the design matrix.

3.4.4 Optimisation

In [142], the derivatives for an *IT2 FLS* when making use of the *KM* algorithm to type-reduce the *FLS* were derived. In this paper, following similar procedures, the steepest descent algorithm is derived for an *IT2 FLS* but this time using the *NT* method of defuzzification. The *NT* method bypasses the computationally demanding type-reduction stage when using the *KM* algorithm.

For the particular antecedent or consequent parameter τ, the update gradient descent algorithm tries to move 'downhill' as given by the following equation:

$$\begin{aligned} \tau(t+1) &= \tau(t) - \alpha_\tau \frac{\partial e_i}{\partial \tau} \\ e_i &= \frac{1}{2}[f_{s_i} - y_i]^2 \end{aligned} \tag{3.13}$$

where t is the iteration count and y_i is the ith output for $i = 1, \ldots, N$. The challenge is to find the derivatives of e_i with respect to all of the parameters of *IT2 FLS*.

3.4.4.1 Antecedent Parameters

For the antecedent parameters (Gaussian primary *MF* of fixed mean and uncertain spread), the derivatives are as follows:

$$\frac{\partial e}{\partial \theta_{ij}^l} = (f_s - y)\left(\frac{h_i - f_s}{\bar{\mathbf{F}}^\intercal \mathbf{r} + \underline{\mathbf{F}}^\intercal \mathbf{r}}\right)\left(\left[\prod_{q=1, q\neq j}^{n} \overline{\mu}_{iq}\right]\frac{\partial \overline{\mu}_{iq}}{\partial \theta_{ij}^l} + \left[\prod_{q=1, q\neq j}^{n} \underline{\mu}_{iq}\right]\frac{\partial \underline{\mu}_{iq}}{\partial \theta_{ij}^l}\right) \tag{3.14}$$

where θ_{ij}^l is the l^{th} parameter of the jth antecedent of the ith rule. For $j = 1, 2, \ldots, n$, $i = 1, 2, \ldots, c$ and $l \in v, \underline{\sigma}, \bar{\sigma}$.

3.4.5 Consequent Parameters

The derivative of the consequent parameters is given as follows:

$$\frac{\partial e}{\partial \beta_{ip}} = (f_s - y)\left(\frac{\overline{f}_i + \underline{f}_i}{\bar{\mathbf{F}}^\intercal \mathbf{r} + \underline{\mathbf{F}}^\intercal \mathbf{r}}\right)\tilde{x}_p \tag{3.15}$$

where β_{ip} is the pth consequent parameter of the ith fuzzy rule, for $p = 0, 1, 2, \ldots, n$ and $i = 1, 2, \ldots, c$.

3.5 Case-Study Steel prediction

This section presents the application of the proposed modelling framework on real life engineering modelling problem involving the mechanical property prediction of hot rolled steels. The performance index chosen in this preliminary stage is the widely used root mean-square error (*RMSE*) defined by the following equation:

$$RMSE = \sqrt{\frac{\sum_{k=1}^{N}(y_k - \hat{y}_k)^2}{N}} \tag{3.16}$$

where N is the number of testing data,y_k and $\hat{y}_k$ are the kth actual output and corresponding model output respectively. As will be discussed in Section 5, this performance index is not suitable for the *BCa* data because of presence of the censored observations[4]. The goal of testing the proposed framework on the selected problem is so as to compare it with existing models of the same type on a highly complex prediction problem. The proposed modelling framework is further tuned so as to be suitable for the *BCa* problem. The heat treatment of hot-rolled steel is a complex and highly non-linear process. Data-based models have been developed in the past to aid metallurgists design steel

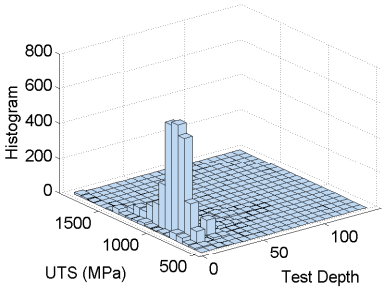
[4]Explained extensively in chapter 5.

alloys [135]. One of the goals of these models is to predict the ultimate tensile strength (*UTS*) which is a common measure of metal strength. A total of 3760 data samples collected from the industry are used to elicit the models in this research work. These data samples include 15 inputs and one output (which is the *UTS*). 12 new data points were used for validating the final model. These two data sets are independent (time-wise) but are from the same production cycle and manufacture setting. Consequently, they have identical types and number of input and output variables. 15 input variables and 1 output variable (ultimate tensile strength (*UTS*) of the steel alloys in megapascal (*MPa*)) were included in the study and are listed in Table 3.1. The distributions of some of the inputs of the data are shown in Fig. 3.11.

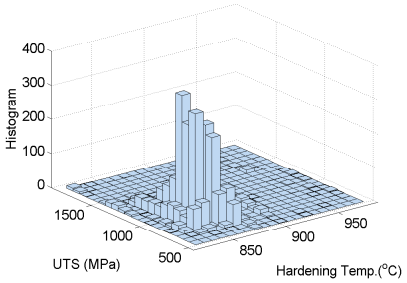
Six fuzzy rules (corresponding to six clusters) were found from data using the elicited framework. Prediction performance of the elicited model is shown in Fig. 3.12. The results are also compared with those available in the literature [143, 144] which are shown in Table 3.2.

3.5.1 Analysis of Results

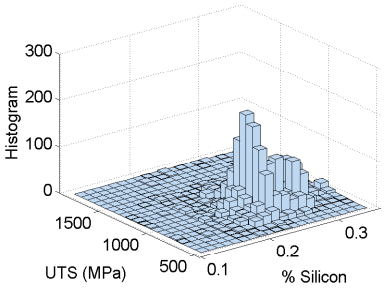
Table 3.2 shows the result of the proposed modelling framework compared to elicited models of the same type in the literature. The proposed modelling framework outperforms the other two models in both training and testing



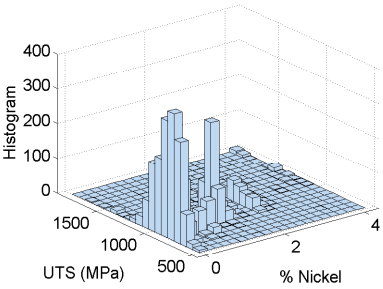
(a) Distribution of Test Depth (mm) and *UTS* (MPa).



(b) Distribution of Hardening Temperature (^oC) against *UTS* (MPa).



(c) Distribution of amount of Silicon (%) against *UTS* (MPa).



(d) Distribution of amount of Nickel (%) against *UTS* (MPa).

Figure 3.11: Distribution of some of the input variables against the output variable (*UTS*).

data as well as generalising well to the previously unseen 12 data points. The six (6) rules for each input variable are shown in Fig 3.13.

3.6 Summary

This chapter has reviewed the literature on existing methods for fuzzy modelling. The chapter has discussed the uncertainty handling advantages of the type-2 configuration over the conventional type-1. A new systematic method

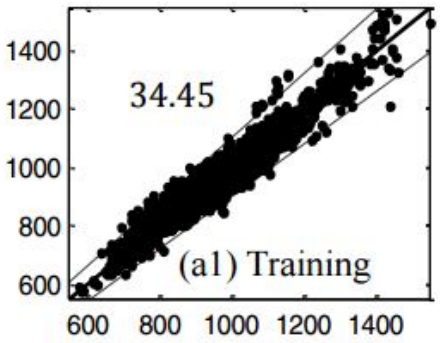
Table 3.1: The input and output variables of the *UTS* dataset.

	Input variable	**Output variable**
1	Test Depth (mm)	Ultimate Tensile
2	Size (mm)	Strength (*UTS*)
3	Site Number	in MegaPascal
4	Carbon (%)	(*MPa*)
5	Silicon (%)	
6	Manganese (%)	
7	Sulphur (%)	
8	Chromium (%)	
9	Molybdenum (%)	
10	Nickel (%)	
11	Aluminium (%)	
12	Vanadium (%)	
13	Hardening Temperature (^oC)	
14	Cooling Medium	
15	Temperature (^oC)	

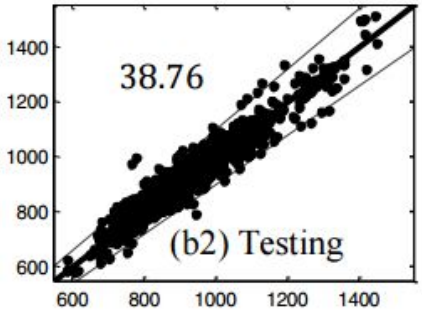
Method	Training	Testing	Validation
IT2-Squared	34.45	38.76	37.34
MOIT2FM	36.33	40.52	34.77
IMOFM_M	46.47	45.12	49.87

Table 3.2: Comparison of proposed method with those found in the Literature for the prediction of *UTS* of steel (*RMSE*).

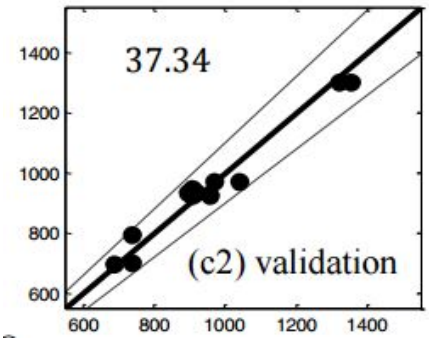
for eliciting interval type-2 fuzzy models from data has been discussed. This method systematically initialises the parameters of the type-2 fuzzy model by making use of the interval type-2 fuzzy clustering algorithm which is a shift from existing method of random parameter initialisation. Compared with models of a similar type, the proposed modelling framework shows significant improvement in modelling performance. The next chapter will discuss



(a) Training Data Results.



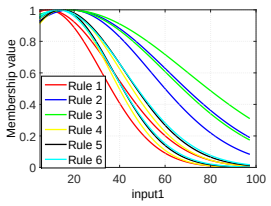
(b) Testing Data Results.



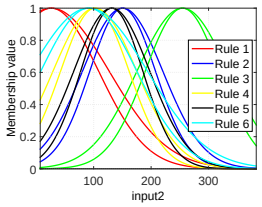
(c) Indepent Data Results.

Figure 3.12: Results on the Prediction of *UTS* of steel.

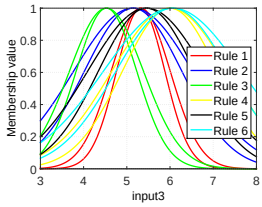
the limitations of existing fuzzy modelling technologies and will propose new a fuzzy-Bayesian integration scheme to address these limitations.



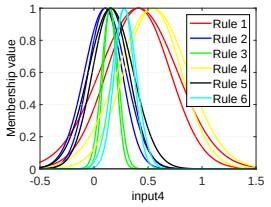
(a) Input 1 (Test Depth)



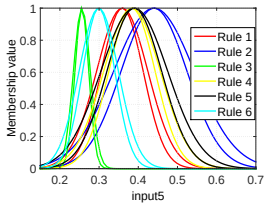
(b) Input 2 (Size)



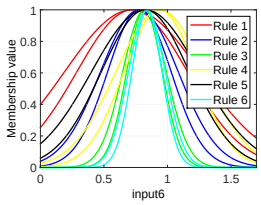
(c) Input 3 (Site Number)



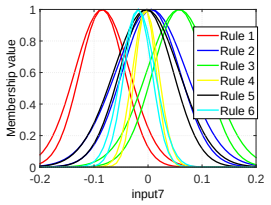
(d) Input 4 (Carbon)



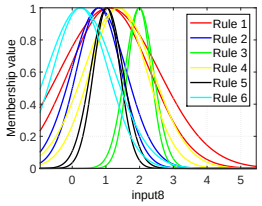
(e) Input 5 (Silicon)



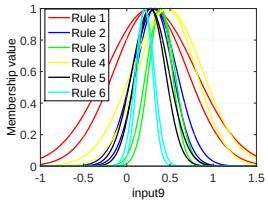
(f) Input 6 (Manganese)



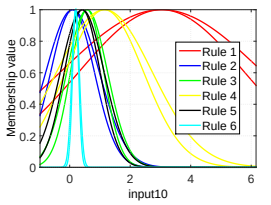
(g) Input 7 (Sulphur)



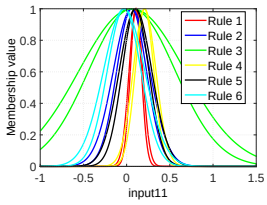
(h) Input 8 (Chromium)



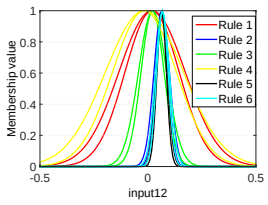
(i) Input 9 (Molybdenum)



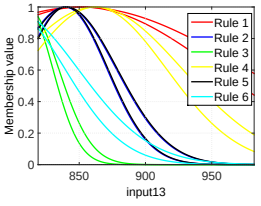
(j) Input 10 (Nickel)



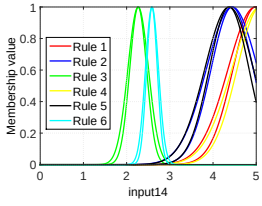
(k) Input 11 (Aluminium)



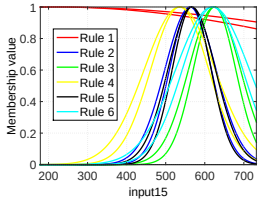
(l) Input 12 (Vanadium)



(m) Input 13 (Hardening Temperature)



(n) Input 14 (Cooling Medium)



(o) Input 15 (Temperature)

Figure 3.13: Rules of the elicited fuzzy model.

Chapter 4

Quantification and Management of Uncertainty

4.1 Introduction

The previous chapter has introduced the different types of fuzzy logic systems (*FLSs*) and how they can help to handle linguistic uncertainties. Using the type-2 equivalent may lead to significant improvements in robustness of applications. However, for modelling problems such as those subject to random uncertainties, as is the case with medical systems, even the use of *T2 FLS* may not suffice in building such a robust model. The main problem with conventional *FLSs* (both $T1$s and $T2$s) lies in the fact they must be represented with

'CERTAIN' parameters. For example, in an *IT2 FS* with fixed mean and uncertain width, such as shown in Fig. 3.7a in chapter 3, only one set of parameters is selected after the optimisation stage. As such the ability of conventional *FLSs* to handle random uncertainties is rather limited. However, one often wishes to elicit these *T2 FLSs* from noisy data. For example, it can often be the case that the same input values give rise to completely different outputs even under constant experimental conditions. As discussed in [145], it is not unusual to find two patients diagnosed with the same cancer and having exactly the same recorded covariates values but having widely varying risk or time of death. Many factors may have accounted for this phenomenon including measurement error , unobserved/unrecorded covariates and precision errors [146].

Probabilistic formalisms, especially those based on the Bayesian paradigm, provide a principled mechanism for handling such challenges. The following sections discuss a new framework for integrating the Bayesian framework with the fuzzy modelling paradigm. This proposed synergy maintains the transparency that fuzzy logic provides in systems design whilst allowing for understanding the uncertainties discussed from the viewpoint of the Bayesian paradigm. The proposed framework is tested on an synthetic and an experimental dataset and the results are discussed in the concluding parts of the

chapter.

4.2 Uncertainty Modelling

Quantifying uncertainties due to measurement errors are well understood and established using the mature fields of probability theory and statistics. However, a unified framework for identifying, quantifying and mitigating the effects of these uncertainties in large scale complex systems with errors in models with a view to making informed decisions on such systems remains an active research area. [146] details 5 areas where the quantifying of uncertainties are critical for understanding the underlying phenomena and making informed predictions which are weather, climate, hydrology and geology, nuclear reactors and biological systems. For example, it is more informative, intuitive, accurate and in-line with how weather phenomena are observed to predict in percentage the likelihood of, say, 'rain tomorrow' since weather systems are so complex that it is impossible to know the certainty of rain. Likewise, in biological systems such as in *BCa* studies, the system is equally as complex, since there is incomplete understanding of the *BCa* process. It is thus more natural, as in weather systems, to predict the likelihood of death after the selected threshold of time say 5 years. The probabilistic formalism can allow for such estimations as will be discussed shortly. However, it is im-

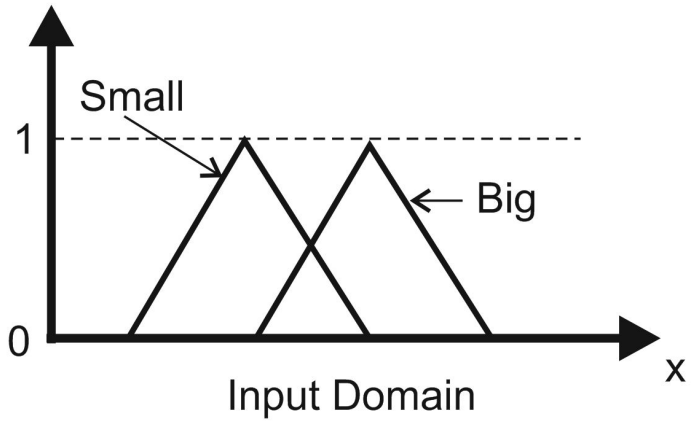
perative to motivate and introduce a new representation of *FSs* proposed in this thesis which is called the non-parametric fuzzy set *NPFS* before discussing this new framework which is discussed in the next section.

4.3 Non-parametric Fuzzy Sets

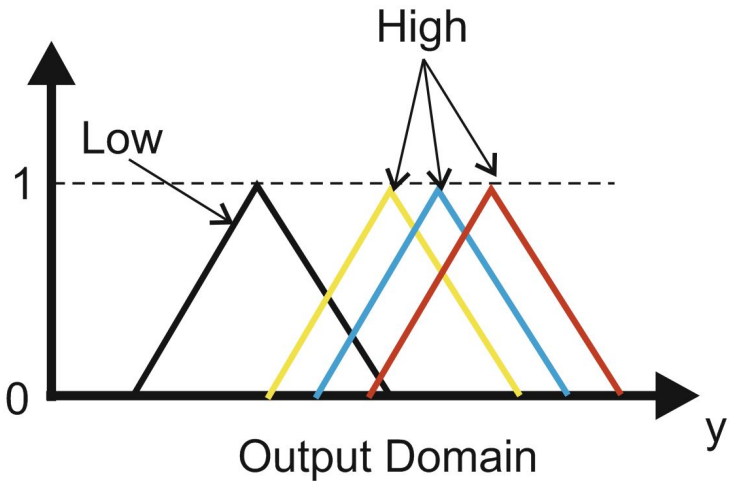
The new representation of an *FS* is introduced via an example. Considering an hypothetical scenario using the following two rules:

```
Rule 1: if the house is small then the price is low
Rule 2: if the  house is big then the price is high
```

The uncertainty embedded in this rule is in defining the *MF* of the linguistic variables *small, big, low* and *high* [127]. Suppose three persons are polled to define the *MFs* for these rules. Different types (shapes and sizes) of *MFs* will most likely result, as discussed in chapter 3. Now, suppose, for simplicity of analysis, that three persons agree on the exactly the same *MFs* for *small, big, low* but disagree on the *MF* for *high*. Fig. 4.1 shows hypothetical *MFs* illustrating this point. A *T2 FS* can help one handle this difference in opinion of the *MF* for *high* (see chapter 3). One way of obtaining a *T2 FS* is to merge the *MFs* as shown in Fig. 4.2 [127], which will transform the existing *T1 FLS* into a *T2 FLS* or a *IT2 FLS*.



(a) Input domain.

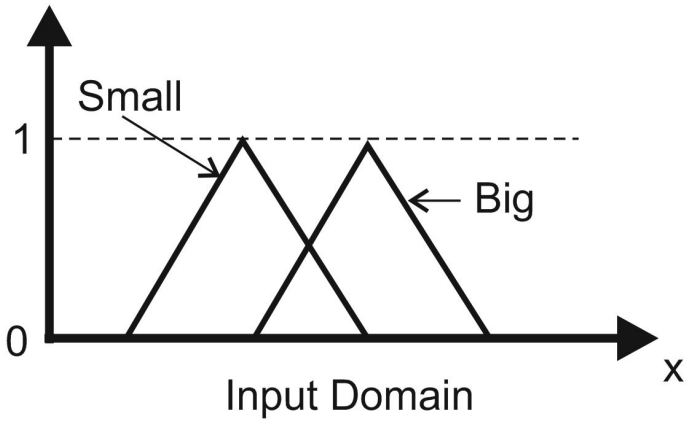


(b) Output domain.

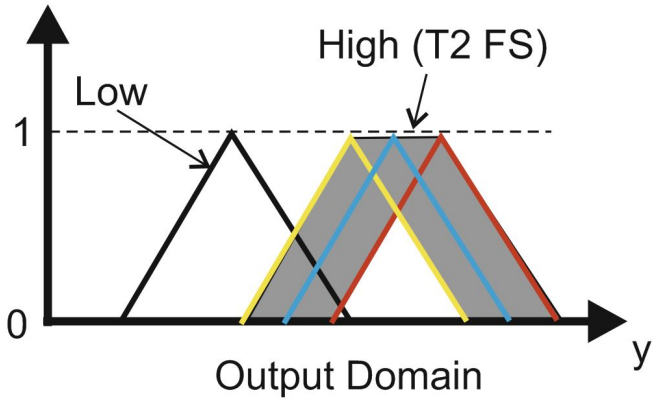
Figure 4.1: MFs for the Rules

The resulting output may be found according to the extension principle [128] which is given by the following equation:

$$C_{\tilde{\boldsymbol{A}}} = \int_{u \in J_{x_1}} \cdots \int_{u \in J_{x_N}} [f_{x_1}(u_1) \star \cdots \star f_{x_M}(u_k)] \Big/ \frac{\sum_{k=i}^{M} x_k u_k}{\sum_{k=i}^{M} u_k} \tag{4.1}$$



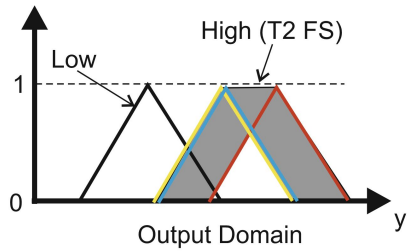
(a) Input domain.



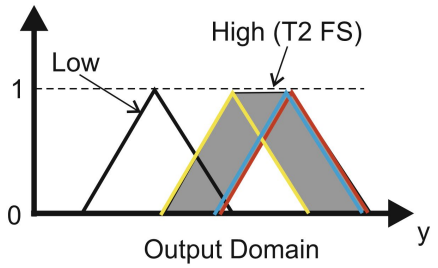
(b) Output domain.

Figure 4.2: Merging of *MFs* results in a *T2 FLS*.

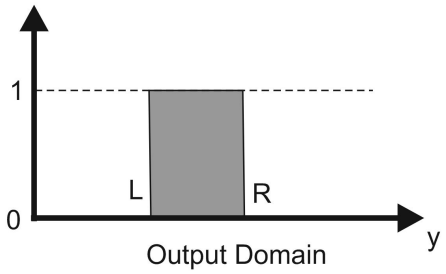
where M is the level of discretisation of the output domain, $f(x_k)$ is the secondary grade of u_k at x_k and $\star$ is the t-norm operator. Equation (4.1) simply states that all embedded *T2 FSs* are first enumerated, then the defuzzified val-



(a) Persons 1 and 2 have similar MF for high.



(b) Persons 2 and 3 have similar MF for high.



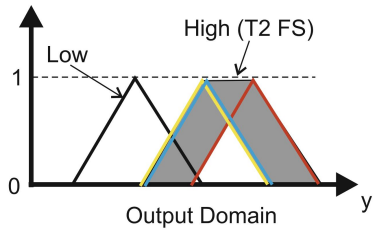
(c) Resulting T1 interval after type reduction of a *IT2 FS*.

Figure 4.3: There is no change in the *LMF* and *UMF* of (a) and (b). Consequently, there is no change in the L and R switch points (c). Uncertainty not well accounted for.

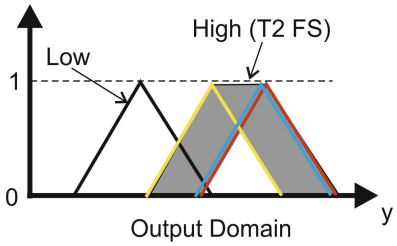
ues are found according to the rightmost side of equation (4.1) and the degree of truth of this new defuzzified value calculated by using the t-norm operator of the corresponding secondary grades.

Hence, for example, if the resulting *MF* is an *IT2 FS*, the *TR* set at the output processing stage is a *T1* interval according to the Karnik-Mendel algorithms (Fig. 4.3c). The *T1* interval basically states that there is an uncertainty in the output which is defined by the left (L) and right (R) end points as shown in Fig. 4.3c. Now, since for each of the merged *MFs* shown in Figs. 4.4a and 4.4b, the values of *UMF* and *LMF* remain unchanged, then the L and R switch points and consequently the defuzzified output remains unchanged even when person 2's *MF* changes. Consequently, the fact that persons' 1 and 2 *MFs* are close is not accounted for as shown in Fig. 4.4. Person 2's *MF* is subsumed in the merged *IT2 FS* and if it is moved even closer to person 3, there are no changes in the resulting *T1* interval after *TR*. Evidently this does not correctly indicate the type of uncertainty in the *FLS*. If one merges this *MFs* into a *T2 FS*, then the secondary *MFs* must be defined. This process is difficult and may result in the loss of interpretability of the fuzzy system.

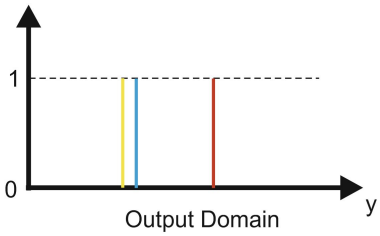
A better approach may be to leave the *MFs* as they are with each *MF* of a particular variable having a corresponding value of antecedent and consequent parts, as shown in Figs. 4.4c and 4.4d. This requires defining a new representation of *FSs* which is called the non-parametric fuzzy set (*NPFS*) in this thesis.



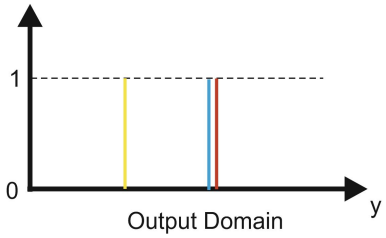
(a) Persons 1 and 2 have similar MF for high



(b) Persons 2 and 3 have similar MF for high



(c) Defuzzified values without merging for (a)



(d) Defuzzified values without merging for (b)

Figure 4.4: By not merging and leaving the *MFs* as they are, one is able to clearly distinguish the uncertainty in each of the defuzzified values. Consequently, uncertainty is well accounted for.

Interestingly, this is equivalent to the Bayesian approach thus providing a powerful mechanism for handling both random and linguistic uncertainties. This approach is akin to having a type of an hierarchical *T1 FLS*, each corresponding to a particular *FLS* of the different persons polled.

4.3.1 Mathematical Representation

The *NPFS* is introduced mathematically in this section. Consider the Dirac delta function, $\delta(x)$, $x \in R$ which is an element in the universe of discourse, X. The Dirac delta function is defined by the following properties:

$$\delta(x) = \begin{cases} \delta(x) = \infty & : x = 0 \\ \delta(x) = 0 & : x \neq 0 \end{cases} \text{ and } \int_{-\infty}^{\infty} \delta(x)dx = 1. \tag{4.2}$$

Then, for any function $f(x)$, the following equation is obtained:

$$\int_{\mathcal{X}} f(x)\delta(a) = f(a) \tag{4.3}$$

where $\mathcal{X}$ is the space of x. Based on the definitions above, it is easy to show that a *T1 FS* may be represented by the following equation:

$$\boldsymbol{A} = \frac{\lambda}{N} \sum_{i=1}^{N} \delta(x^{(i)}) \tag{4.4}$$

where λ is the area of the fuzzy set, $x^{(i)}$ is the ith point of x and N is the number of points used to represent the fuzzy set. To make it more concrete,

it should be recalled that the centroid defuzzification method is given by the following equation:

$$x^* = \frac{\int x\mu(x)dx}{\int \mu(x)dx} \tag{4.5}$$

As a result of the definition of the Dirac delta function, then one can show that the defuzzified value is given by the following equation:

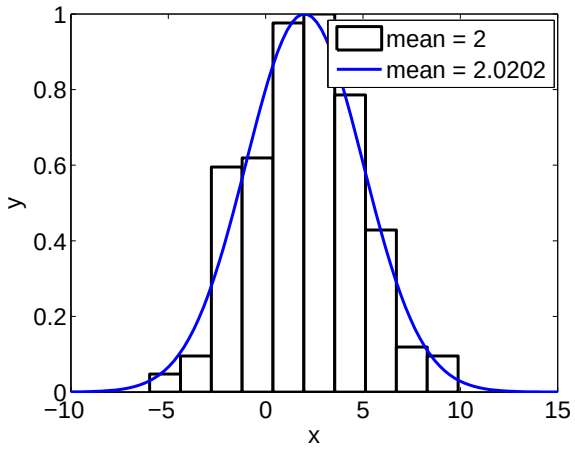
$$\hat{x}^* = \frac{1}{N}\sum_{i=1}^{N} x^{(i)} \tag{4.6}$$

For the maximum defuzzification, the following equation is obtained:

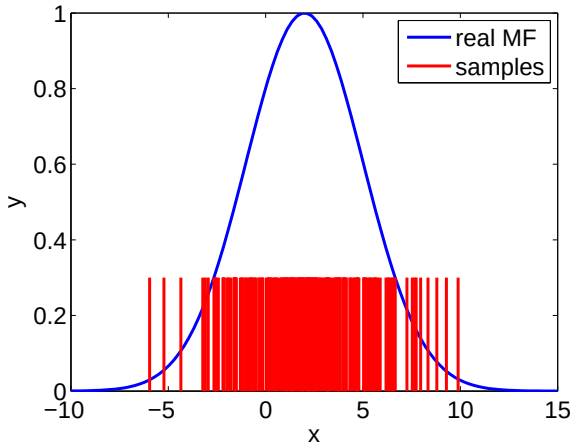
$$\hat{x}^* = \underset{x^{(i)};i=1,\cdots,N}{\arg\max}\ \mu(x^{(i)}) \tag{4.7}$$

other types of defuzzification may be derived following same logic.

How to choose the $x^{(i)}$'s is easy and various computer packages exist for sampling from a one-dimensional function [147]. To show that this method works, examples of representing different fuzzy sets using the *NPFS* is presented shortly. The sampling method employed is the acceptance sampling method. Details of this algorithm is available in [147]. Figs. 4.5 and 4.6 show examples of using the *NPFS* to represent these *MFs*. 10000 samples were taken using the acceptance sampling method and the histogram was found to be able to visualize the form of these samples. The form of the histogram shows that the *NPFS* is a viable and a accurate representation of *FSs*.

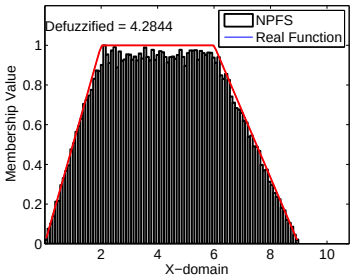


(a) The centroid and histogram

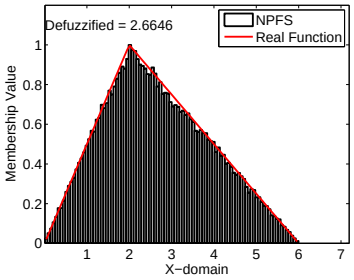


(b) The Samples

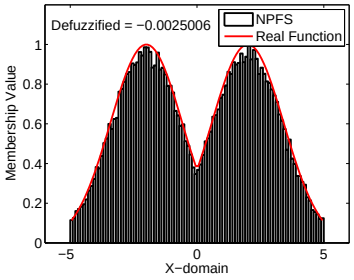
Figure 4.5: Gaussian Membership Function



(a) Trapezoidal Membership Function.



(b) Triangular Membership Function.



(c) Mixtures of Gaussian Membership Function. The centroid of the original function is 0. Centroid of the *NPFS* representation is -0.0025006.

Figure 4.6: The NPFS used in representing different types of *MFs*.

It should be noted from the centroid and histogram that the samples are representative of the *FSs*.

It also worth noting that this method is applicable to both *T1 FS* and *T2*

FS. To represent a *T2 FS*, the secondary MFs (which are *T1 FSs*) are simply replaced with the samples as illustrated in the preceding equations.

4.4 Bayesian and Fuzzy Integration

It is argued that one of the advantages of fuzzy reasoning in modelling is its ability to obtain linguistic models of data without explicitly stating the probability distribution. This statement is not totally true because in building fuzzy models from data, a loss/error function must be defined in the optimisation stage. As shown in [148] and [25], for many models Bayesian reasoning in modelling leads to a generalization from which different loss functions can be derived by defining the type of likelihood and prior i.e. the loss functions can be obtained as a special case of Bayesian inference with different types of priors [82]. It should noted that, by defining a particular error function, the distribution of the data has been implicitly assumed.

In the probabilistic paradigm, given a set of data $\mathcal{D}$, $\mathcal{D} = \{(\mathbf{x}_i, y_i)|i = 1, \cdots, N\}$, where $\mathbf{x} \in R^d$ is a vector of the input and $y \in R$ is the output. The goal of learning from data is to find a function f such that every data point, $f(\mathbf{x}_i) \approx y_i$ for all i. What differentiates learning algorithms from one another is the nature of the function, f. For a *FLS*, the nature of f is as shown in Fig. 3.5b. In the statistical sense, the generative assumption is usually made,

whereby the observed output is the corrupted version of the deterministic output from f as given by the following equation:

$$y_i = f(\mathbf{x}_i, \mathbf{w}) + \epsilon \tag{4.8}$$

where $\mathbf{w}$ is the parameters of f. The simplifying assumption that the observed data are independent and identically distributed (*IID*) Gaussian distribution with zero mean and variance σ^2 is usually made such that $\epsilon \sim \mathcal{N}(0, \sigma^2)$, or equivalently:

$$p(y|\mathbf{x}_i, \mathbf{w}) = \mathcal{N}(f(\mathbf{x}_i, \mathbf{w}), \sigma^2) \tag{4.9}$$

The noise model assumption and the data give rise to the dataset likelihood, $p(\mathcal{D}|\mathbf{w})$, given by the following equation:

$$\begin{aligned} p(\mathcal{D}|\mathbf{w}) &= \prod_{i=1}^{N} p(y_i|\mathbf{x}_i, \mathbf{w}) \\ &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(f(\mathbf{x}_i, \mathbf{w}) - y_i)^2}{2\sigma^2}\right) \end{aligned} \tag{4.10}$$

The Bayesian paradigm involves one defining a prior distribution $p(\mathbf{w})$ over the parameters which expresses of our initial belief on the parameters of the model.

The prior in the presence of evidence (data) is updated to a posterior distribution using Bayes' rule as follows:

$$P(\mathbf{w}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathbf{w})P(\mathbf{w})}{\int P(\mathcal{D}|\mathbf{w})P(\mathbf{w})d\mathbf{w}} \tag{4.11}$$

$P(\mathbf{w}|\mathcal{D})$ is called the posterior distribution of $\mathbf{w}$ in light of evidence $\mathcal{D}$. When a test input $\mathbf{x}^*$ is given, the output y^* is the posterior predictive distribution obtained by marginalising out $\mathbf{w}$ as follows:

$$P(y^*|\mathcal{D}) = \int P(y^*|\mathbf{w})P(\mathbf{w}|\mathcal{D})d\mathbf{w} \tag{4.12}$$

The formation of a predictive distribution by the integration of equation (4.12) is the essence of the Bayesian inference [149] which provides a systematic confidence in the output depending on the degree of random uncertainty embedded in the data.

Prior elicitation is equivalent to expert knowledge incorporation in fuzzy inference and it is a pivotal issue for both Bayesian and fuzzy inferences. However, this information may not be easy to summarise in the form of a prior probability density function (*PDF*). Readers are referred to [149] for the best ways of eliciting these 'informative priors'.

Until the advent of very fast computing platforms, the integrals of equations (4.11) and (4.12) have been a source of considerable controversy. The highly computational demanding Monte Carlo methods are usually employed for sampling from posterior distribution. To understand the idea behind sampling from $P(\mathbf{w}|\mathcal{D})$, it should be noted that equation (4.12) is the same as finding the expectation, E of $P(y^*|\mathbf{w})$ with respect to $P(\mathbf{w}|\mathcal{D})$. As a result of the law of large numbers [147], by taking n *IID* samples from $p(z)$, the integral

such as $\int_{\mathcal{X}} h(z)p(z)dz$ can be approximated with tractable sums $\frac{1}{n}\sum_{i=1}^{n} h(z^{(i)})$ that converge as follows:

$$\frac{1}{n}\sum_{i=1}^{n} h(z^{(i)}) \xrightarrow[n\to\infty]{as} \int_{\mathcal{X}} h(z)p(z)dz. \tag{4.13}$$

where $h(z)$ is a function of variable z whose expectation one wishes to evaluate with respect to $p(z)$; this is equivalent to equation (4.12). Fig. 4.7 shows the results of sampling from a multimodal function ($p(x) \propto 0.3\exp(-0.2x^2) + 0.7\exp(0.2(x-10)^2)$) using the Metropolis-Hastings algorithm for 5000 iterations [147]. [147] and [150] offer excellent expositions on the different methods of Monte Carlo sampling.

It should be noted that $P(y^*|\mathcal{D})$ obtained from equation (4.12) is a function of y. More often than not, one is interested in making a single guess: when point predictions are needed (as is usually the case), it can be shown that the estimator $f_{\mathbf{w}}$ to choose to represent a function of $\mathbf{w}$ can be chosen based on $L(f_{\mathbf{w}}, \mathbf{w})$ [151], that is one finds the estimator $f_{\mathbf{w}}$ which minimises the following cost:

$$J = \int L[f_{\mathbf{w}}, \mathbf{w}]p(\mathbf{w}|y)d\mathbf{w} \tag{4.14}$$

Squared error loss function results in a posterior mean [152], while the absolute loss function estimates the median of the posterior, that is $L[f_{\mathbf{w}}, \mathbf{w}] = |f_{\mathbf{w}} - \mathbf{w}|$ [151].

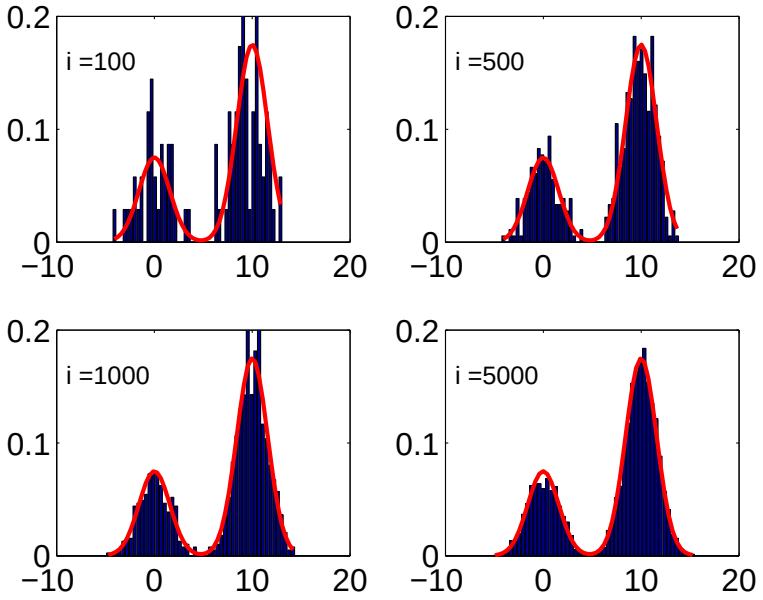


Figure 4.7: Intended distribution with histogram of different sizes of the Monte Carlo samples.

4.4.1 Bayesian Approach

As mentioned earlier, the parameters of the *FLS* are assumed to be random variables and the Bayesian reasoning applied. For a *FLS* with a total of P parameters, for simplicity of analysis, a Gaussian prior with zero mean of the following form is assumed:

$$p(\mathbf{w}) = \frac{1}{Z_{\mathbf{w}}(\alpha)} \exp(-\frac{\alpha}{2} \parallel \mathbf{w} \parallel) \tag{4.15}$$

where α is the inverse variance and is usually called an hyper-parameter because it is a parameter of a distribution of other parameters. The normalisa-

tion constant is given as follows:

$$\frac{1}{Z_{\mathbf{w}(\alpha)}} = \left(\frac{2\pi}{\alpha}\right)^{P/2} \tag{4.16}$$

It should be noted that, choosing the best prior which represents our prior expert knowledge is of crucial importance. The book by Murphy [148] provides an excellent resource for this.

If an *IID* dataset is assumed, the likelihood function is defined as follows:

$$\begin{aligned} p(\mathcal{D}|\mathbf{w}_{\text{FLS}}) &= \prod_{i=1}^{N} p(y_i|\mathbf{x}_i, \mathbf{w}_{\text{FLS}}) \\ &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(f_{\text{FLS}}(\mathbf{x}_i, \mathbf{w}_{\text{FLS}}) - y_i\right)^2}{2\sigma^2}\right) \end{aligned} \tag{4.17}$$

$f_{\text{FLS}}(\mathbf{x}_i, \mathbf{w}_{\text{FLS}})$ is a *T1 FLS* with vector of parameter $\mathbf{w}_{\text{FLS}}$. These parameters are the width and spread of, say, a Gaussian *MF* and the weights of the consequent part of a Takagi-Sugeno-Kang (*TSK*) *FLS*, y_i is the ith observed output for input $\mathbf{x}_i$ for $i = 1 \cdots N$, with N being the number of data points. According to Bayes' rule, the posterior can be written as follows:

$$p(\mathbf{w}_{\text{FLS}}|\mathcal{D}) = \frac{p(\mathcal{D}|\mathbf{w}_{\text{FLS}})p(\mathbf{w}_{\text{FLS}})}{p(\mathcal{D})} \tag{4.18}$$

To elicit a prediction given a new input vector x^*, as discussed earlier, the weights are integrated out as illustrated below:

$$p(y^*|x^*, \mathcal{D}) = \int p(y^*|x*, \mathbf{w}_{\text{FLS}})p(\mathbf{w}_{\text{FLS}}|\mathcal{D})d\mathbf{w}_{\text{FLS}} \tag{4.19}$$

A point prediction that uses the mean of the distribution is given by the following equation:

$$E(y|x^*, \mathcal{D}) = \int \int yp(y|x*, \mathbf{w}_{\text{FLS}})p(\mathbf{w}_{\text{FLS}}|\mathcal{D})d\mathbf{w}_{\text{FLS}}dy \tag{4.20}$$

It is easy to show that equation (4.20) may be rewritten as follows:

$$E(y^*|x^*, \mathcal{D}) = \int f_{\text{FLS}}(x^*, \mathbf{w}_{\text{FLS}})p(\mathbf{w}_{\text{FLS}}|\mathcal{D})d\mathbf{w}_{\text{FLS}} \tag{4.21}$$

where $f_{\text{FLS}}(x^*, \mathbf{w}_{\text{FLS}})$ is the output of the *FLS*. The posterior $p(\mathbf{w}_{\text{FLS}}|\mathcal{D})$is usually sampled from using Markov chain Monte Carlo techniques because the integrals in the equations above are intractable and numerical methods breakdown when the dimension of the parameters is greater than twenty. After sampling, different weight sets can be said to correspond a particular *FLS*. By obtaining samples from the posterior and then making predictions using the sampled weights, what basically results is a set of outputs which corresponds to the non-parametric representation earlier discussed in this chapter (see section 4.3). Every fuzzy set in the antecedent has as many embedded fuzzy sets as there are number of weights. The only difference between the proposed method and the conventional *T2 FLS* is that the proposed method says that for every weight sampled in particular antecedent parameter, there corresponds another weight in the antecedent and consequent parameters which is not necessarily true in the conventional *T2 FS*. Therefore, equation (4.21)

should state that:

Samples of the parameters are taken according to the posterior distribution of the parameters. Each parameter set corresponds to the parameters of the *FLS*. Then the final defuzzified output is calculated using equation (4.7), which is the *NPFS* earlier proposed.

The above definition proves to us that the Bayesian-fuzzy system approach yields a *NPFS* at the output whose form may be approximated via curve estimation methods (kernel density) like those of Fig. 4.6.

4.5 Results

This section presents the results of the proposed modelling framework on two datasets. For both, the *TSK FLS* was used and the initial *FLS* parameters initialised via the use of the fuzzy c-means (*FCM*) clustering algorithm as explained in chapter 3. The first dataset is a synthetic data and four clusters (corresponding to four rules) were generated. The second dataset is a real dataset from the metal industry and eight clusters (corresponding to eight rules) were generated. The centroid defuzzification method was used on the output *NPFS* after type-reduction in both cases.

4.5.1 Synthetic Dataset

This section presents the results of using the Bayesian–fuzzy technique in the modelling of a synthetic dataset obtained from the following non-linear function:

$$y = x_1^2 + x_2^2 \tag{4.22}$$

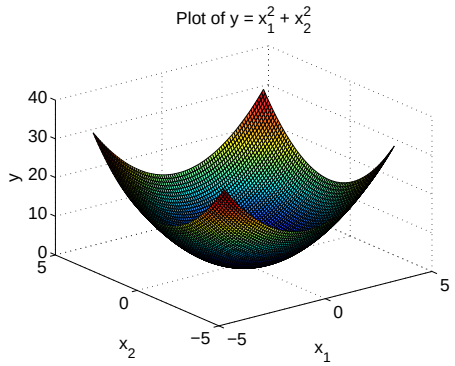
The function includes two input variables (x_1 and x_2) and one output variable (y).

The training dataset is generated between $-4 \leq (x_1, x_2) \leq 4$ with an increment of 1 but some points are removed as shown in Fig. 4.8b. 10 testing data points were generated randomly from the uniform distribution $\mathcal{U}$ such that

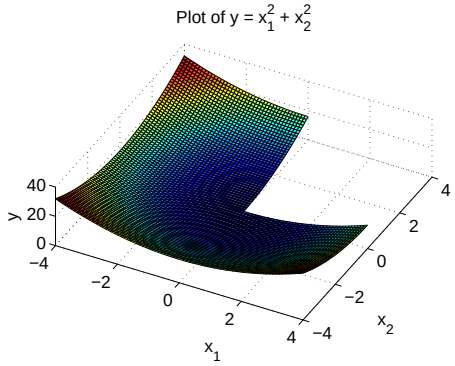
$$-6 \leq (x_1, x_2) \leq 10 \tag{4.23}$$

The testing dataset is labelled $(1 - 10)$ in Fig. 4.8c. The uncertainty of prediction in areas with sparse data such as data points 1 and 4 should be higher than those with less-sparse data such as 5 and 7. 1000 posterior weights of the *FLS* were sampled using the Harmiltonian Monte Carlo technique [153] as explained in the preceding section. Fig. 4.9 shows the result on the testing data with the confidence bands.

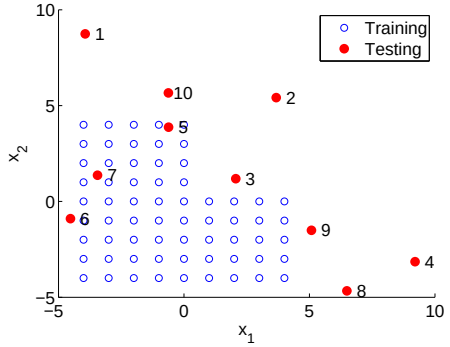
The confidence bands which are represented by a *NPFS* are derived by



(a) The plot of the function.



(b) The plot of the function after the cut.



(c) The plot of the scatter with labels.

Figure 4.8: Sythetic Dataset

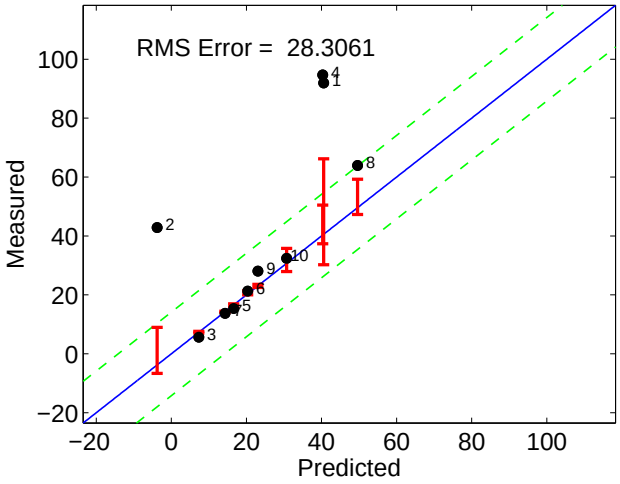
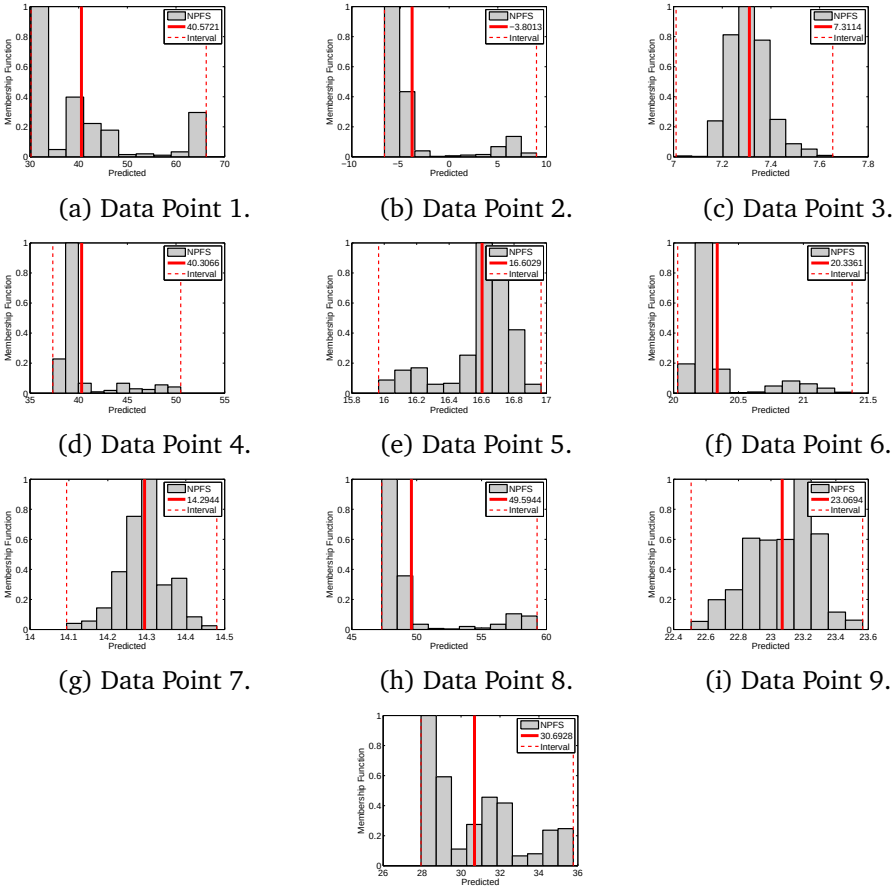


Figure 4.9: Prediction results and confidence on the synthetic dataset.

finding the difference in the minimum and maximum of the resulting type-reduced *NPFS* as shown in Fig. 4.10.

The degree of confidence on the training dataset is expected to be high (the confidence band is expected to be low) and this is exactly the case as shown in Fig. 4.11.

It is worth noting that data points 1 and 4 have confidences that are low compared to data points 5 and 7. The resulting fuzzy rules for the antecedent part is shown in Fig. 4.12.



(a) Data Point 1. (b) Data Point 2. (c) Data Point 3.

(d) Data Point 4. (e) Data Point 5. (f) Data Point 6.

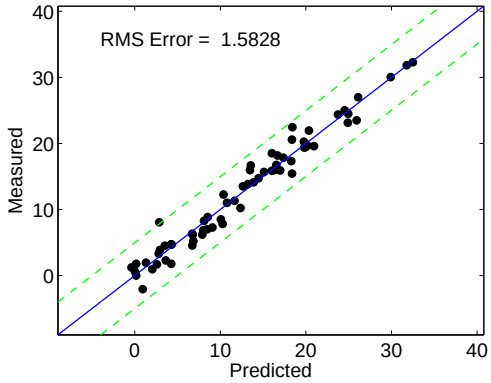
(g) Data Point 7. (h) Data Point 8. (i) Data Point 9.

(j) Data Point 10.

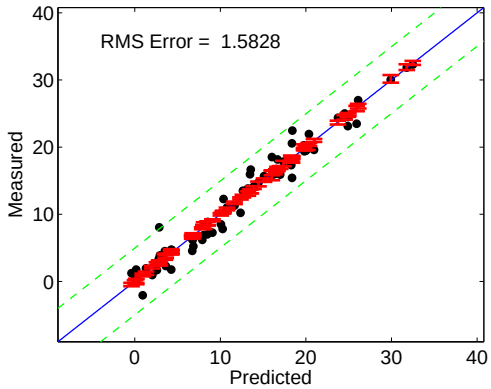
Figure 4.10: Results of the Prediction of Synthetic Dataset.

4.5.2 Ultimate Tensile Strength of Steel Model

The proposed modelling approach is further tested on the steel data used in chapter 3 which is a real data drawn from the metal industry whereby the ultimate tensile strength (*UTS*) of steel alloys are predicted prior to production. This data set has been analysed in chapter 3. As already mentioned, there



(a) Training Dataset Results.



(b) Training dataset performance with prediction confidence.

Figure 4.11: Training dataset performances on the synthetic dataset.

are two sets of datasets each variables as listed in Table 3.1. The decision not to combine these two datasets was taken so as to be able to test how the model would perform on a completely remote dataset. Additionally, a visual inspection of this second dataset revealed that the 7th and 8th data points have exactly the same values for the inputs but different values for the output

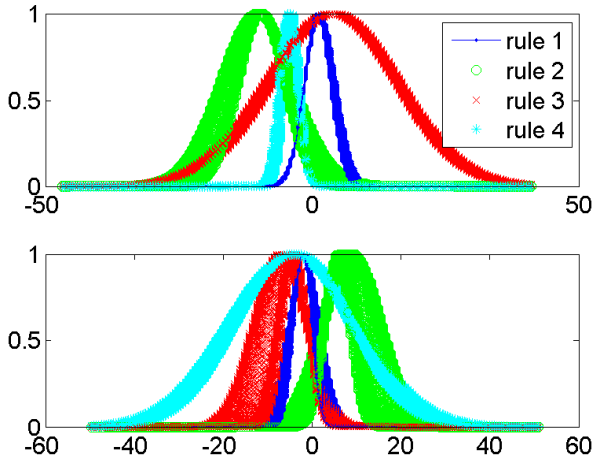


Figure 4.12: Rules generated from the synthetic dataset modelling.

which may perhaps be due to the random uncertainties embedded in the data such as measurement precision or insufficient number of input variables as already discussed. It may even be due to errors in the data entry stage by the operator. These are the types of random uncertainties embedded in data modelling that existing fuzzy models may not handle adequately. It would be interesting to find out how the proposed framework handles this uncertainty though.

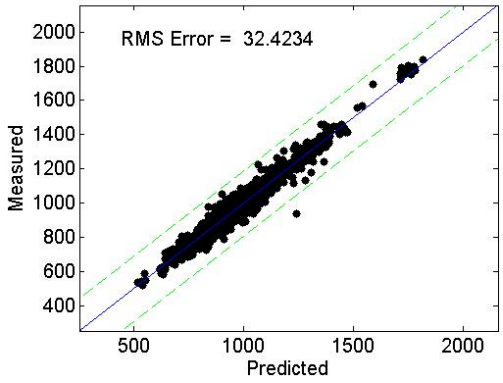
The distribution of some of the input variables against the output variable is shown in Fig. 3.11. It can be observed that the data are highly dense in some areas and highly sparse in others. The first dataset was divided into training (70% of the data) and testing (30%) of the data. The testing dataset

was used to test the generalization performance of the elicited model.

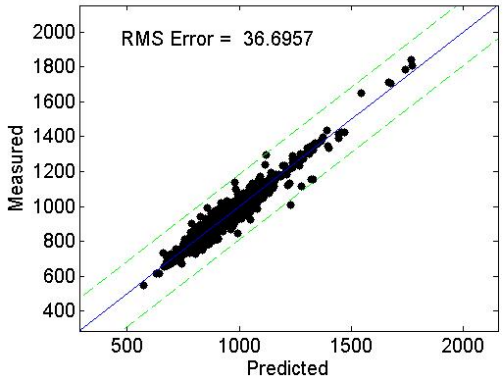
4.5.2.1 Results

The proposed modelling framework can help one overcome this by giving us a visual representation of this uncertainty in the form of an *NPFS* after type reduction. It is able to discriminate between the data points that have been subject to these uncertainties as shown in Figs. 4.14b, 4.14a, 4.15h and 4.15i. The elicited model was able to model the uncertainty in the predictions in these data points.

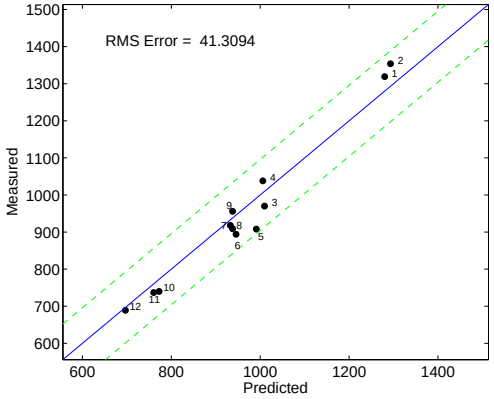
The modelling performances on the training, testing and 12 independent dataset are shown in Fig. 4.13. Compared with existing models of similar type, our proposed framework returned very good results. The model strength, however, lies in its ability to quantify the uncertainty of each predicted output as shown in Figs. 4.14 and 4.15. For example data point 12 has more uncertainty compared to data points 3 and 4. It should be noted that the points close to the middle of the plot tends to have higher confidence than points at the extreme (data points 12, 2 and 1). This is reasonable as when one observes the dataset distribution (Fig. 3.11), it can be seen that the data tend to be denser at the centre of the plot. The model is not as certain for data points with less density than it is for data points with more density.



(a) Training Data Results.

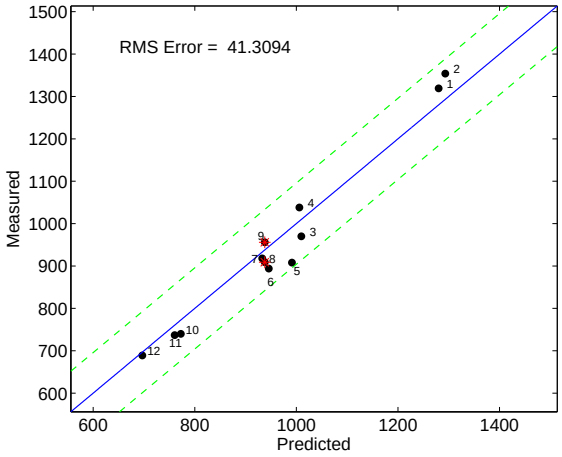


(b) Testing Data Results.

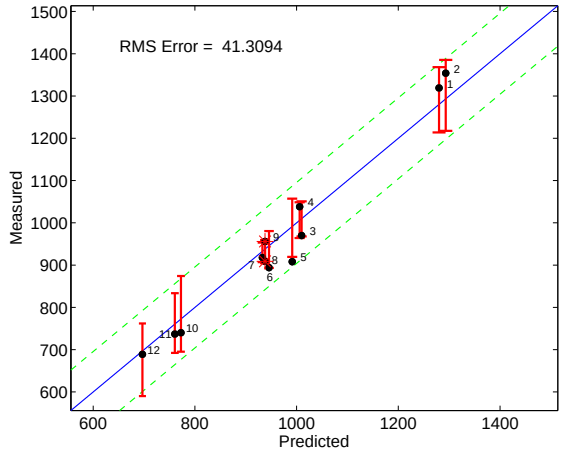


(c) Indepent Data Results.

Figure 4.13: Results on the prediction of *UTS* of steel

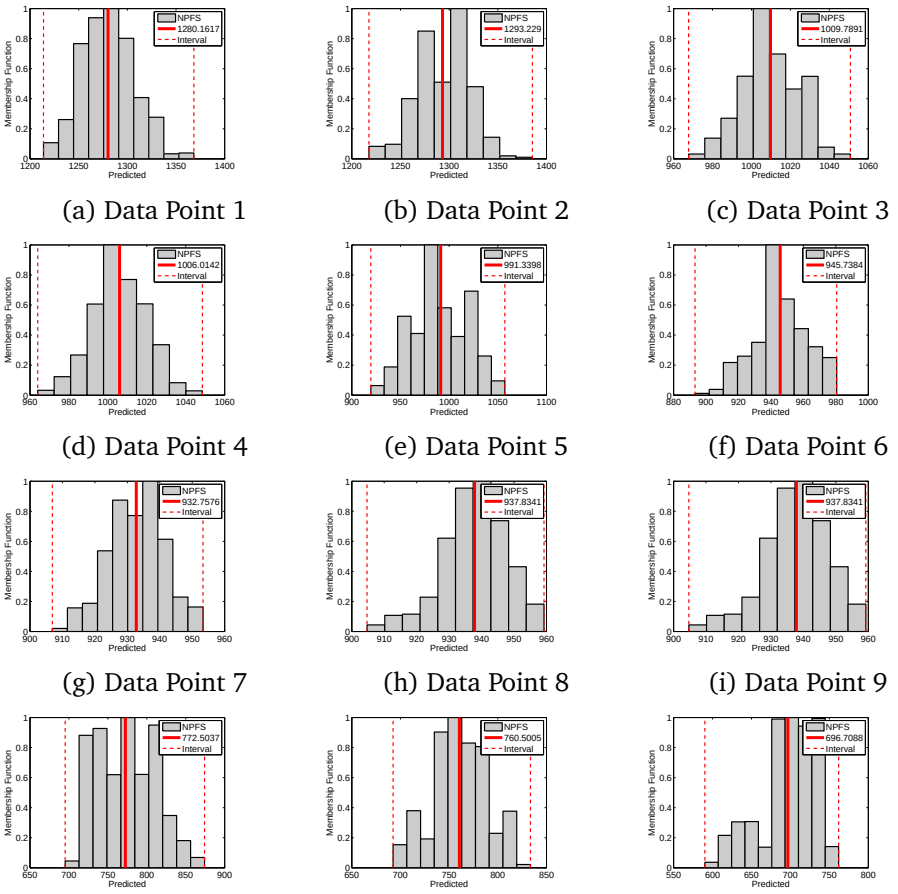


(a) Second Dataset Results highlighting data points 8 and 9



(b) Prediction with confidence

Figure 4.14: Second data set results.



(a) Data Point 1 (b) Data Point 2 (c) Data Point 3

(d) Data Point 4 (e) Data Point 5 (f) Data Point 6

(g) Data Point 7 (h) Data Point 8 (i) Data Point 9

(j) Data Point 10 (k) Data Point 11 (l) Data Point 12

Figure 4.15: Results of the prediction of *UTS* of steel.

4.6 Summary

The limitations of the existing fuzzy modelling technologies have been discussed in this chapter. Specifically, the chapter has emphasized that the capability of conventional type-1 and type-2 fuzzy logic systems to handle random uncertainties is limited by providing examples and representations. To better

handle random uncertainties, the chapter has introduced a fuzzy-Bayesian integration method whereby the parameters of the fuzzy models are taken to be random variables. The proposed method was tested on real and synthetic datasets which provided evidence of its better random uncertainty handling. The next chapter will present extensive analysis of the bladder cancer data used in this research. It will also provide an interval type-2 fuzzy modelling framework which allows one to predict, beforehand, the risk of death of a new patient diagnosed with bladder cancer.

Chapter 5

Static Risk Model using an Integrated Fuzzy Survival Analysis Approach

It has already been highlighted that one of the key challenges in eliciting a predictive model for the *BCa* data is censoring. Survival analysis techniques provide the key to handling censored data. The use of standard survival analysis techniques (such as the Kaplan-Meier estimator and the Cox proportional hazards model) on the *BCa* data as well as their strengths and weaknesses are discussed in this chapter. In the final sections, the fuzzy modelling framework proposed in earlier chapters is integrated with these standard techniques so

that it is able to handle the peculiarities of the *BCa* database. This proposed synergistic framework is tested on *BCa* data as well as on an artificial data. Testing on the artificial data is based on the premise that, unlike in the *BCa* data, it is possible to generate the 'would be' event times had the subject not been censored. The elicited model is a static risk map because it does not allow to incorporate patient dynamic information. However, the elicited model, as will be seen, provides useful insights for clinicians and patients. Results of this new modelling framework are then compared with standard techniques from both the *BCa* and artificial data.

5.1 Bladder Cancer Data Analysis

5.1.1 The Data

As discussed in chapter 1, the *BCa* data were obtained from a study of *BCa* patients at the Royal Hallamshire Hospital in Sheffield, United Kingdom[1]. In order to understand the long term consequence of *BCa* patients, the hospital created a database of *BCa* patients between 1 January, 1994 and 31 December, 2009 [5]. Patients diagnosed with the disease, but who are still alive after the study period, are automatically censored. There were 3634 patients with primary *BCa* but those with insufficient follow up (< 6 months) were ex-

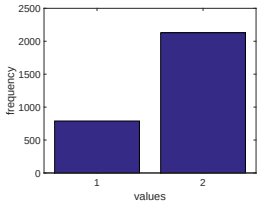
[1]Sole provider of urological services in the city of Sheffield, UK

cluded from such analysis. Additionally, patients with missing covariates were excluded from the analysis leaving 2918 patients in the database. Thirteen (13) explanatory variables (input variables) were recorded for each patient which included details on disease pathology, patient specific characteristics and treatment interventions (if any). The data were anonymised to maintain the privacy of patients in the study. Details of each variable are shown in Table 5.1 and Fig. 5.1.

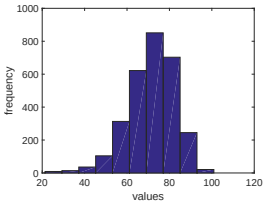
The response variable is time of death from *BCa* in months. The median survival time is 35.26 months (30 days taken equal to 1 month). Of the 2918 patients 2305 were censored (78.99%) due to end of study, loss to follow up or death due to other causes.

The distribution of times is shown in Fig. 5.2. There is a higher proportion of censored observations at longer follow-up times than at lower follow-up times. This is typical in survival analysis and is due to the study having a fixed duration which means that patients with largest follow up times tend to be censored.

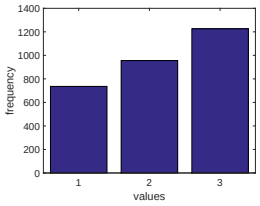
It is worth mentioning at this stage that treatment decisions (cystectomy and/or radiotherapy) were implicitly modelled by including them as part of the input variables. This can provide information on how these treatments affect the risk prognostics index and can help in providing clinicians recom-



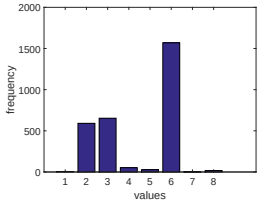
(a) Input 1 (Sex)



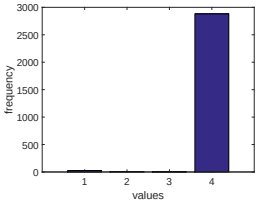
(b) Input 2 (Age)



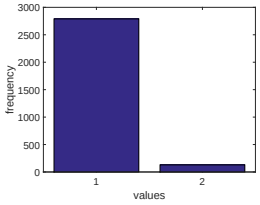
(c) Input 3 (Grade)



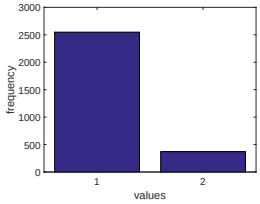
(d) Input 4 (Stage)



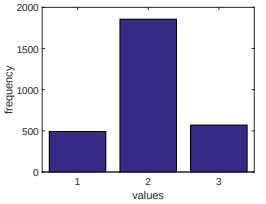
(e) Input 5 (Nodes)



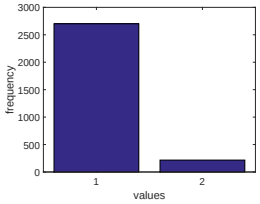
(f) Input 6 (Squamous)



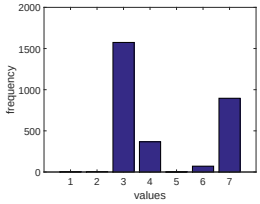
(g) Input 7 (CIS)



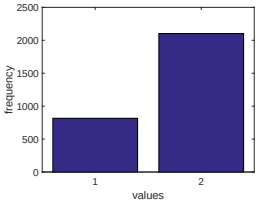
(h) Input 8 (SPB)



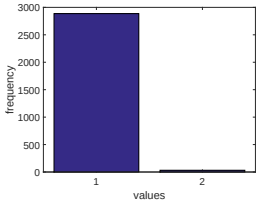
(i) Input 9 (Vascular)



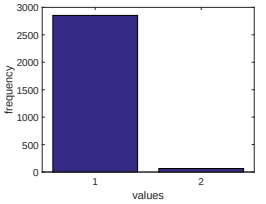
(j) Input 10 (Urothelium)



(k) Input 11 (Muscle)



(l) Input 12 (Cystectomy)



(m) Input 13 (Radiotherapy)

Figure 5.1: Distributions of the explanatory variables

Table 5.1: Input variables in the *BCa* data.

Continuous Variables	**Median**	**Mean**	**Range**
Age	72.7 years	71.6 years	21.3 - 101.0 years
Stage	4.03	4.02	0 - 9
Categorical Variables	**Values**	**Number of Patients**	**Percentage**
Sex	Male	2129	72.96%
	Female	789	27.04%
Tumour Grade	Good	736	25.22%
	Moderate	956	32.76%
	Poor	1226	42.02%
Squamous	No	2789	95.58%
	Yes	129	4.45%
Carcinoma in situ	No	2548	87.32%
	Yes	370	12.68%
Morphology	Solid	492	16.86%
	Papillary	1856	63.61%
	Both	570	19.53%
Vascular Invasion	No	2701	92.56%
	Yes	217	7.44%
Muscle Invasion	No	816	27.96%
	Yes	2102	72.04%
Cystectomy	No	2886	98.90%
	Yes	32	1.10%
Radiotherapy	No	2854	97.81%
	Yes	64	2.19%

mendations for therapy as already discussed.

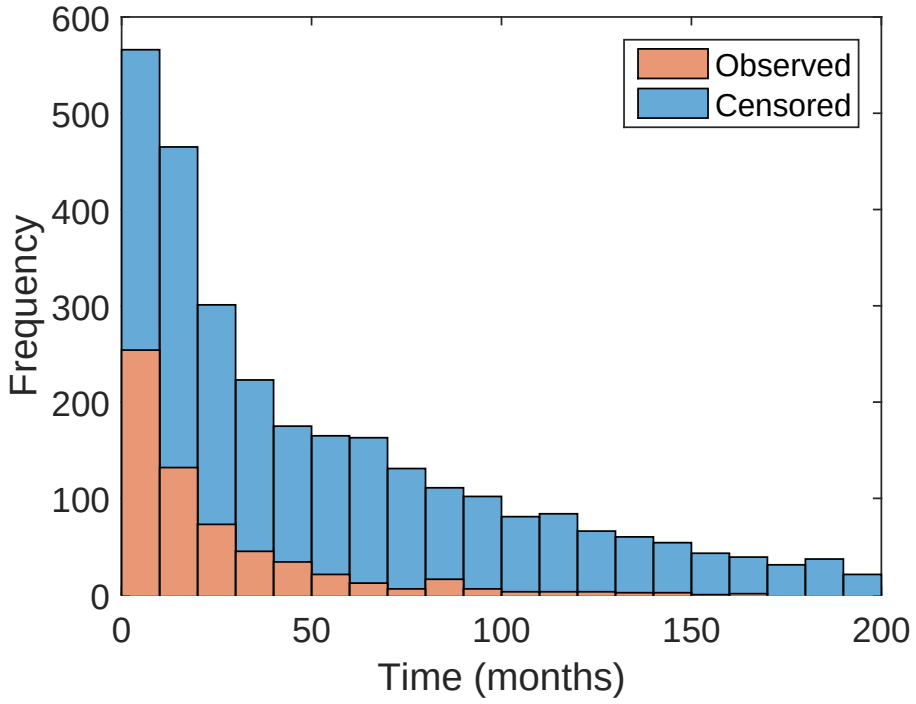


Figure 5.2: Histogram of the times of events for both censored and uncensored observations for the bladder cancer data set. Patients with higher follow-up times tend to be the most censored as is typical in survival data studies. This is is because the study was for a limited amount of time and patients with high event times become automatically censored at the conclusion of the test.

5.2 Survival Analysis Introduction

Survival analysis relates to the analysis of time until an event occurs. Mathematically, if T is a random variable which represents the time an event occurs, survival analysis is concerned with identifying $\mathrm{P}(T > t)$ from data. In survival studies, the quantities of interest are given as follows:

$$
\begin{aligned}
f(t) &= \int_0^t F(t) \\
S(t) &= Pr(T \leq t) = 1 - Pr(T < t) = 1 - F(t) \\
f(t) &= \frac{\partial F(t)}{\partial (t)} = -\frac{\partial S(t)}{\partial (t)} \\
h(t) &= \lim_{\delta t \to 0} \frac{Pr(t \leq T \leq t + \delta t | T \geq t)}{\delta t} = \lim_{\delta t \to 0} \frac{1}{\delta t} \frac{Pr(t \leq T) \leq t + \delta t}{Pr(T \geq t)} \\
&\lim_{\delta t \to 0} \frac{F(t + \delta t - F(t))}{\delta t S(t)} = \frac{f(t)}{S(t)} = -\frac{d}{dt[\log S(t)]} \\
H(t) &= \int_0^t h(u)du = -\int_0^t \frac{d}{du}[\log(S(u))du] = -\log S(t) \\
h(t) &= \frac{\partial H(t)}{\partial t} \\
S(t) &= \exp(-(H(t))
\end{aligned}
\tag{5.1}
$$

where Pr(A) denotes the probability of event A, $h(t)$, $S(t)$, $H(t)$ and $F(t)$ are the hazard function, survivor function, cumulative hazard function and cumulative distribution function for the random variable given as T respectively.

The set of equations in (5.1) shows that deriving one quantity allow one to derive the other. For example, knowing the survivor function allows one to calculate the hazard. The hazard function ($h(t)$) is the quantity most of interest in this thesis as will be seen. The hazard provides the risk an individual face at a given time. Survival analysis is primarily concerned with identifying these functions in the presence of censored observations. The nature of censoring and the different types of censoring will be discussed next.

5.2.1 Censoring

Censoring relates to when the response variable is not fully observed. For example, in the *BCa* data, all the patients have entry points (i.e. when they were diagnosed as having *BCa* i.e disease specific mortality (*DSM*)) but not all have an exit point (when they died from *BCa*). Some patients were 'lost' during follow-up and one is only aware of the last recorded time of follow-up. Also, some patients (38%) died from other causes (this is known as the other cause mortality (*OCM*)). These groups of patients are also considered censored since they **DID NOT** die from *BCa* and it is impossible to know the *BCa* death times since it is impossible to die twice. Fig. 5.3 illustrates this point.

Mathematically, one assumes that there are two type of times for the patient (i). The censored time (C_i) and the event time (D_i) but only one ($T_i = \min(C_i, D_i)$) is observed. The rule for deciding which is observed is given as follows:

$$T_i = \begin{cases} D_i & \textit{if} \quad D_i \leq C_i \\ C_i & \textit{if} \quad D_i > C_i \end{cases} \tag{5.2}$$

Survival analysis techniques allow to model the event times (even if not observed) from the censored observations. In essence, they allow one to answer the following question: *What would have been the time of death from BCa*

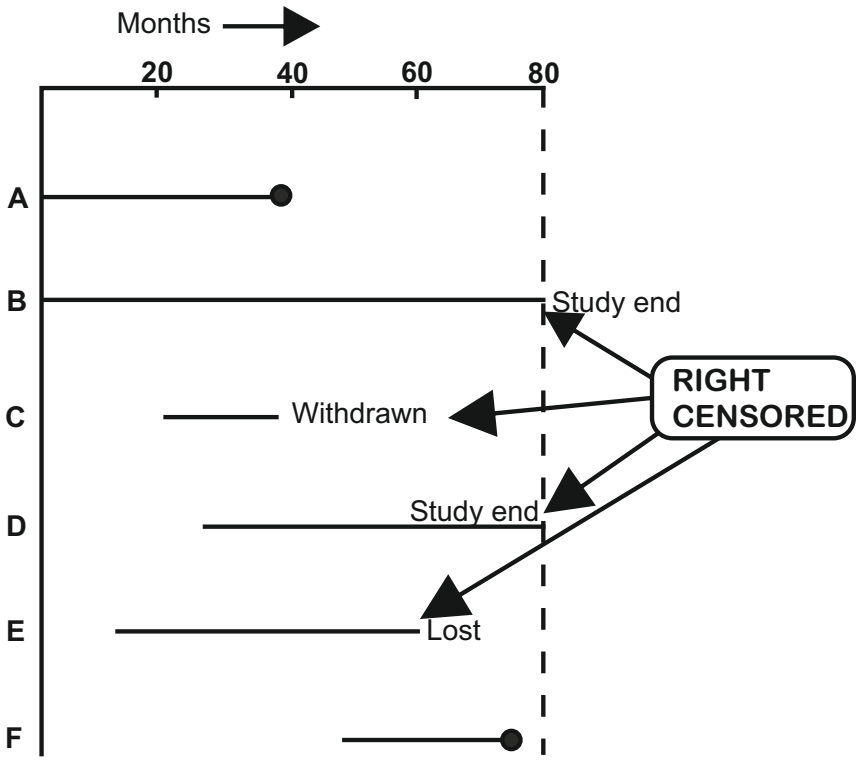


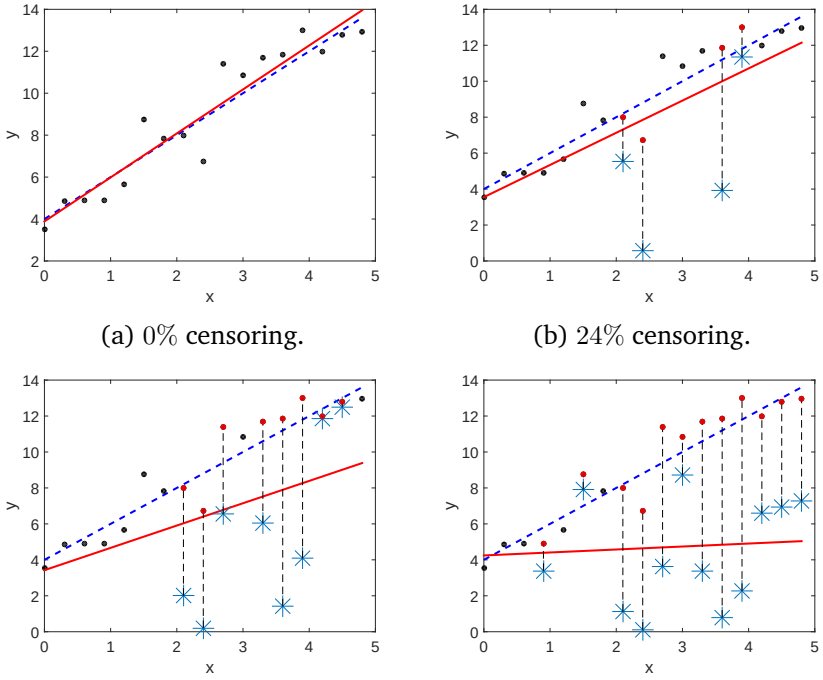
Figure 5.3: Illustration of censoring. The event times of patients B and D are not known since the study ended before they could be observed. Only their censoring times are known. Patients A and F have their event times fully observed as they died from the event of interest at approximately 40 and 75 months respectively. However, the event time of patient F is about 30 months since the patient entry point is about 45 months after the study started. This phenomenon whereby patients do not enter the study at the same time is known as *staggered entry* [120][154]. Patient C withdrew from the study at 40 months (censoring time is 20 months) while patient E was lost to follow-up at 40 months (censoring time is 50 months).

for a patient who is known to have survived past 3 months after being diagnosed but whose event time is not known?. Some questions arise naturally when analysing survival data. Why not just remove the censored observations? Or why not treat censored observations as event times (i.e. set $T_i = C_i \forall i, \delta_i = 0$)? As noted in [154], this is not ideal. In many survival studies, the number of censored data points are typically more than non-censored ones (79.8% censored in the *BCa* data) which may result in the removal of vital data points, thus resulting in biased models. Setting $T_i = C_i$ is even more flawed as illustrated in Fig. 5.4.

There are three broad categories of censoring namely:

1. **Right Censoring**: This censoring type has already been described and is the most common in medical studies. It occurs when the event time is known to exceed a particular time.

2. **Interval Censoring**: This is a situation where the event time is known to fall within a particular interval. This type of censoring is common when there has been a rounding of times of events for example, rounding off times of death from ***BCa*** results in interval censored observations. However, provided that the interval is not too large, analysis may proceed by treating the censored values as event times with little or no loss in model accuracy.



(a) 0% censoring. (b) 24% censoring.

(c) 47% censoring. (d) 71% censoring.

Figure 5.4: The blue dash line is the underlying function. The black dots are the Gaussian corrupted measurements and the red line is the estimated function using linear regression. The blue stars are the observed censored times for the censored observations (red dots). One observes that as the number of censored observations increase, the elicited model becomes more biased.

3. **Left Censoring**: This is very rare in medical practice and occurs when the event times are known to be below a particular time threshold.

Only right censoring mechanisms are considered in this research study.

5.2.1.1 Standard Techniques

Two broad frameworks for survival analysis are the Cox models and accelerated failure time (*AFT*) models. In the Cox models, the hazard is modelled directly. The hazard is defined as the instantaneous risk of an event and is given in the Cox model [72] by the following equation:

$$h_i(t, \mathbf{x}_i) = h_o(t) \exp(\boldsymbol{\alpha}^{\mathrm{T}} \mathbf{x}_i) \tag{5.3}$$

where $\mathbf{x}_i$ represents the values of the input variables (covariates) for the ith subject, $\boldsymbol{\alpha}$ is the vector of parameters and $h_i(t, \mathbf{x}_i)$ the corresponding hazard for $i = 1, 2, \cdots N$. N is the number of subjects under study.

Equation (5.3) simply means that the hazard of failure at any point in time t for a subject with covariates defined by vector $\mathbf{x}$ is a product of two functions. The first function is time-dependent and is called a baseline hazard ($h_o(t)$) and does not depend on the subject's attributes. The second function depends only on the subject's attribute but is not dependent on time $\exp(\boldsymbol{\alpha}^{\mathrm{T}} \mathbf{x}_i)$. It may be easily deduced that the so-called hazard ratio which is the ratio of two subjects' hazards does not depend on time since the common time function which they both share cancels out.

The *AFT* approach models the time of event directly by assuming the times of an event and covariates act multiplicatively on this time scale. This ap-

proach is very similar to ordinary linear regression. The general *AFT* model equation is given by the following equation:

$$\ln(T_i) = Y_i = \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_i + Z_i \tag{5.4}$$

where Z_i is a random variable which is assumed to follow a particular parametric distribution. T_i is thus also a random variable that defines the distribution of event times for subject i. If one defines $\psi = \exp(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$, then it is easy to show the following:

$$T = \psi_i T_o \tag{5.5}$$

where T_0 corresponds to the random variable of the baseline (when $\mathbf{x} = \mathbf{0}, T_0 = \exp(Z_0)$).

Analysis of equations (5.4) and (5.5) quickly reveals that:

$$S(t, \mathbf{x}) = S_o(t\psi_i) \tag{5.6}$$

The interpretation of the *AFT* models is thus as follows: If $\psi_i < 1$, then it appears as if the clock ticks faster for the subject with covariates ($\mathbf{x}_i$) when compared with the baseline. This means that the time scale for that subject is $T_o\psi_i$. Consequently, the survival time is shortened as a result of the failure being 'accelerated'. However, if $\psi_i > 1$, by a similar argument, then the survival time is lengthened as the consequence of failure being 'decelerated'. For this

reason, ψ, which is sometimes called a link function, serves as an *acceleration factor*.

It is also possible to non-linearise the linear part of this this link function using a non-linear model as has already been done in [155] using neural networks and restricted cubic splines [156]. Other types of interpretation in survival data modelling are also possible e.g. proportional odds [157].

Using linear models in the link functions of both the Cox model and the *AFT* models have peculiar intuitive appeals, especially to clinicians. However, in real systems especially, the assumption that the hazards are affected by an exponentiated linear function of the covariates as given by equation (5.3) may not be tenable and there is a reluctance in using black box non-linear models because of the loss of this intuitiveness and interpretability. It is argued in this thesis that by using fuzzy systems, instead, one can maintain interpretability to allow for incorporation of subjective knowledge as well as increase model accuracy. It will be shown in section 5.5 how the interpretability of the fuzzy model output is not affected when incorporated into the link function. Additionally, this output is indicative of the risk a patient faces on outset of a disease since one is directly modelling the hazard earlier defined. Therefore, using *FLS* may be a promising alternative when non-linearity is desired and interpretability is to be maintained. This is illustrated by means of an artificial

data and the *BCa* data. However, it is instructive to first analyse the *BCa* data using the standard survival techniques as presented in succeeding subsections.

5.3 Kaplan-Meier - Non Parametric on the bladder cancer data

It is worth investigating what information can be drawn from the *BCa* data using the Kaplan-Meier estimator [2]. As mentioned in chapter 2, this estimator is most widely used for analysing homogeneous survival data. In the *BCa* case, the data was stratified into different categories based on sex, treatment type and age-group. The Kaplan-Meier estimator is given by the following equation:

$$\hat{S}(t) = \prod_{t_i \leq t} \frac{r(t_i) - d_i}{r(t_i)} \tag{5.7}$$

where t_i are the times at which the events occurred, $r(t_i)$ are the number of events at risk at time t_i and d_i is the number of events at time t_i.

If death times due to other causes are taken as censored[3], then the cancer specific mortality analysis using the Kaplan-Meier estimator of the *BCa* data is as shown in Fig. 5.5. The Kaplan-Meier estimator applied to selected strata are shown from Figs. 5.7 to 5.9.

[2] Also called the product limit estimator.
[3] It will be discussed in the next section why this may not be tenable.

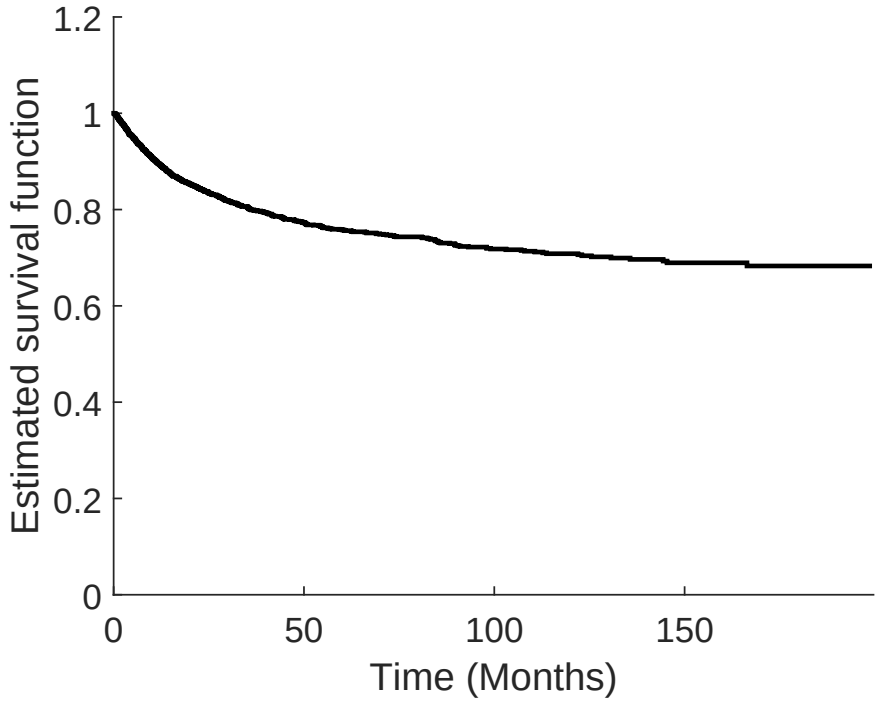


Figure 5.5: Kaplan-Meier analysis of *BCa* deaths.

5.4 Performance Indices

Performance indices allow for comparison of modelling performances of different techniques. Because of the presence of censored observations, using performance indices such as the root mean square error (*RMSE*) typically used in regression or the area under the curve (*AUC*), the receiver operator characteristics (*ROC*) used in classification can be challenging [158]. For this reason new performance indices have been defined in the literature [50] but the concordance-index (c-index), which is one of the used performance indices

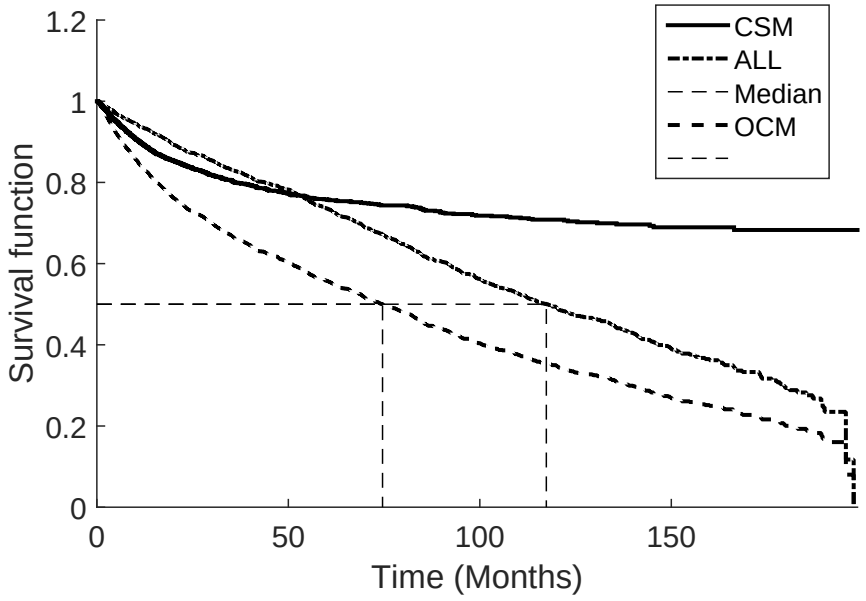


Figure 5.6: Kaplan-Meier analysis of different causes of deaths.

in this thesis appears to have the best intuitive appeal because it is easy to understand.

5.4.1 Concordance Index

The concordance index (c-index) makes use of the fact that a subject with a high risk or prognostic index (p_k), is likely to have a lower survival time than the one with a lower prognostic index (p_l). So, if $p_k > p_l$ and $t_k < t_l$, then the subject pair (p_k, p_l) is said to be in concordance if the assertion holds true. The c-index is useful when one is not strictly interested in the times of failure but rather in grouping the observations into risk groups according to

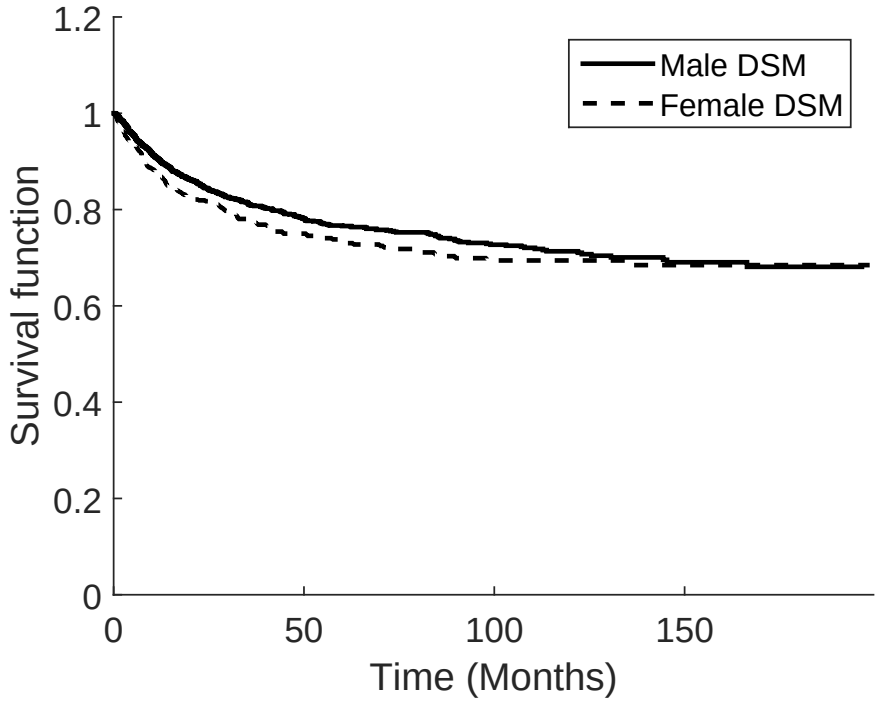


Figure 5.7: *BCa* sex strata Kaplan-Meier estimation.

their risk ranks as is the case in many medical studies and in this thesis. The concordance index is defined as follows:

$$\text{c-index} = \frac{1}{N_p} \sum_{k,l} I(p_k, p_l) \tag{5.8}$$

Where N_p represents the number of unique usable pairs in the observations. A pair is usable if and only if both survival times are observed or the subject with the censored observation has a censored time greater than the subject whose event time is observed. The pair will be unusable if both observations

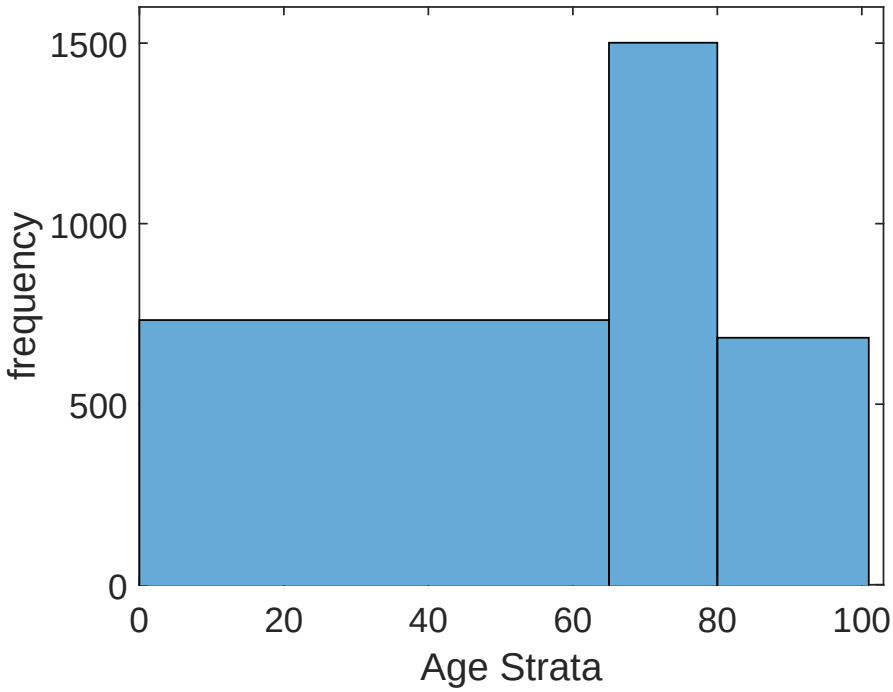


Figure 5.8: *BCa* age strata.

are censored. I is an indicator function which is defined as follows:

$$I(p_k, p_l) = \begin{cases} 1 & \text{if } p_k > p_l, \\ 0 & \text{otherwise} \end{cases} \tag{5.9}$$

The interpretation of this equation is that a higher risk/prognostic index indicates a tendency for a lower survival time.

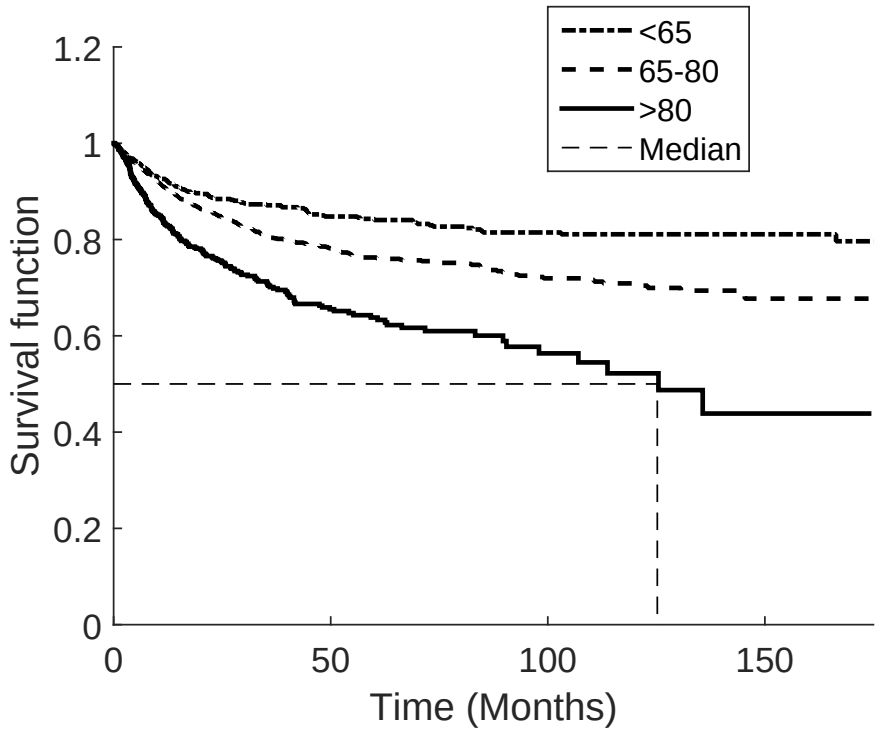


Figure 5.9: Kaplan-Meier estimation of different age strata of the *BCa* data.

5.4.2 Modified ROC

When the prognostic indices for the population has been found, existing methods of finding the *ROC* cannot be used because some observed times are censored. It is not possible in practice to know if an individual with a censored failure time lower than the chosen threshold time (60 months in this study) would survive past this threshold i.e. is a low risk individual. To obtain a pictorial performance index and to be able to use the *ROC* analysis, a new c-index is defined at every stage of calculating the *ROC* curve (false positive *FP* rate vs.

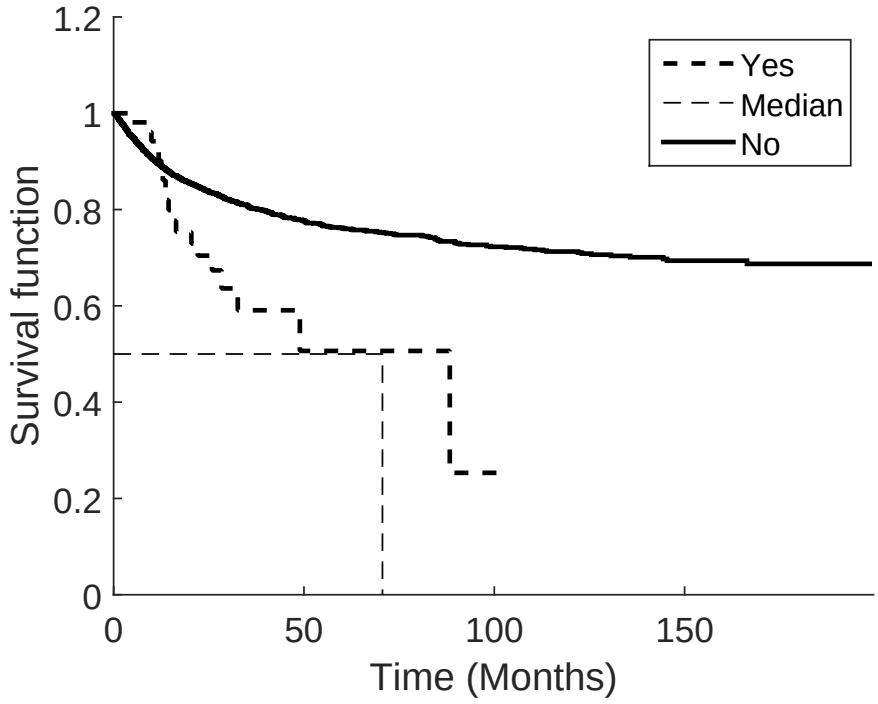


Figure 5.10: Kaplan-Meier Analysis on radiotherapy treatment.

true positive *TP* rate). However, a pair is redefined to be usable by comparing each calculated prognostic indices (p_k) with a constant prognostic threshold (p_{thresh}). Finding the *FP* rate or the*TP* rate is achieved by using this new set of usable pairs. The *ROC* can then be constructed in the usual iterative manner [120]. Also, the Breslow estimator [120] allows for calculating the baseline hazards and baseline survival functions which consequently facilitates calculating median survival times for specific covariate values. This was explored in this study so that the predicted median of survival times can be compared with

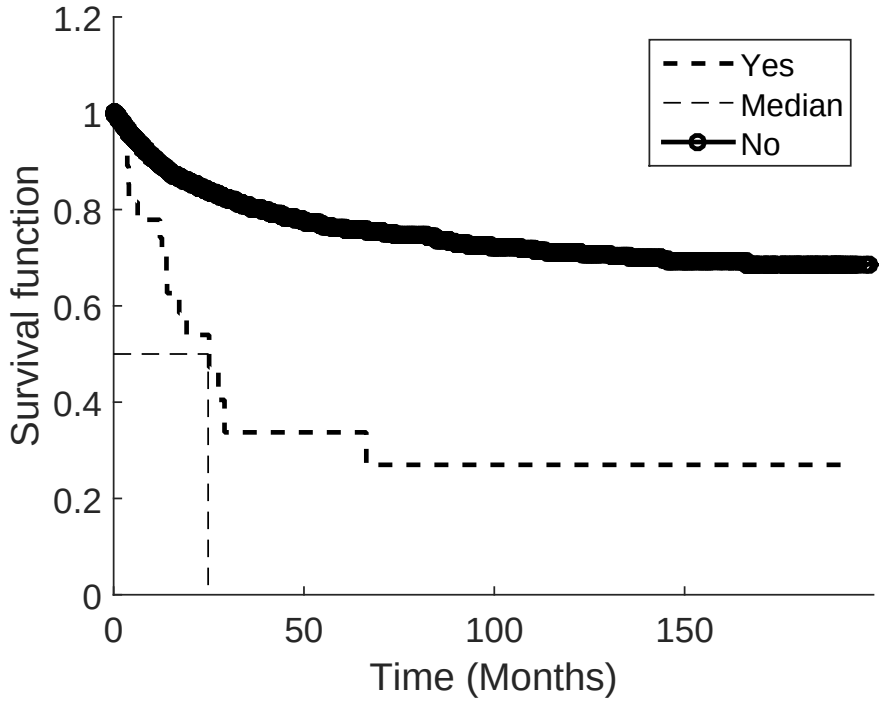


Figure 5.11: Kaplan-Meier Analysis on cystectomy treatment.

the observed values in the case of the artificial data. It is worth noting that it would be erroneous to compare the predicted median survival times with the observed times for individuals whose failure times are censored. A better approach would be to compare times for only those individuals whose failure times were observed exactly (Non-censored). This thesis makes use of both the defined c-index in equation (5.8), the proposed modified *ROC* analysis and the *RMSE* of predicted median times and observed times for non-censored observations to compare results of proposed fuzzy modelling framework and the

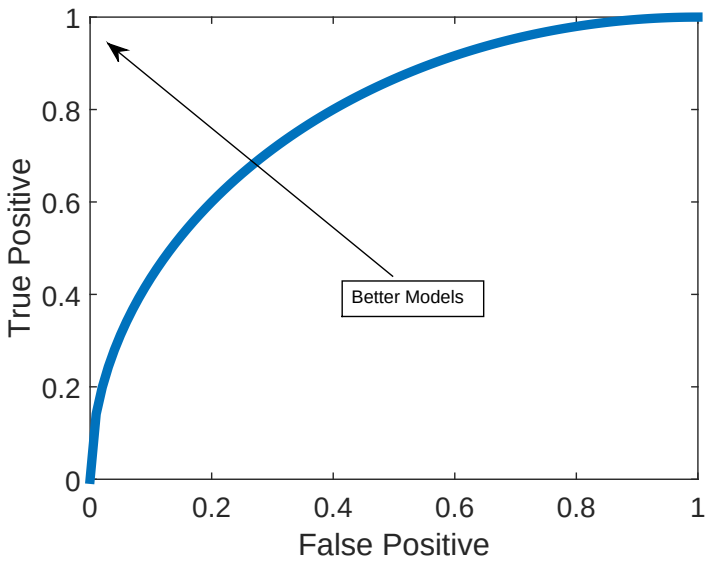


Figure 5.12: Graphical representation of the *ROC* curve.

Cox-model.

Several methods exist to define the risk groups of subjects from survival times, for example see [159]. Here, the simple approach, usually common among clinicians, is taken in this research whereby a time frame is chosen (typically 5 years in cancer studies), and subjects with a failure time greater than this threshold are assumed to be low risk individuals and vice versa. Only censored observations that has survival times greater than this threshold may be included in the study. Nothing can be concluded on individuals with censored times less than this time threshold. The *ROC* curve is constructed as the true positive (*TP*) against the false positive (*FP*) rate as already described.

This thesis investigates the ability of using fuzzy survival modelling results to find risk groups on two the data sets and thereafter provide recommendations for treatment therapy (one artificial data set and one real data set). These two data sets are examined in more details in the next subsections.

5.4.2.1 Artificial Data

The purpose of this artificial data generation is to demonstrate the case where non-linear modelling technique is needed to generate good modelling results. The artificial data generation mechanism is similar to the one in [160]. The data is highly non-linear and using a linear Cox model will lead to sub-optimal modelling results as will be seen. Additionally, using the artificial data, one is able to investigate the hypothetical scenario that one has the event times but were censored by some mechanism. It would then be of interest to compare the elicited model performances with the real times. The artificial data set is generated according to the following expression:

$$t = (x_0+x_1+x_2-15)^2+(x_3+x_4-10)^2+x_5+(x_6+x_7-15)^2+(x_8+x_9-10)^2+\epsilon \tag{5.10}$$

The x's are the covariates drawn from uniform continuous distributions so that they have values between 0 and 10. ϵ is a random value (noise) added to the times and $p(\epsilon) = \frac{1}{2\beta}\exp\left(-\frac{|\epsilon|}{\beta}\right)$, $\beta = 0.4$ for input variable noise and 2.0

for output noise. Independent censoring was achieved by randomly choosing 50% of the data set and randomly drawing a uniform number from 0 to the event times of each the chosen data points. Two thousand (2000) data points were generated via MATLAB® 2015a software for training and 1000 were used for testing. The artificial data set allows for a hypothetical scenario where one knows the exact failure times of the censored observations (an information one is not privy to in real life situations). The model is trained as discussed using the censored times and the event times of the uncensored observations. One can then investigate how the predicted median times (a function of the prognostic index and baseline hazard) compare with what would have been the event times had they not been censored. The *RMSEs* between the predicted median values and event times were used for this comparison in this study. As already stated, the k-fold cross validation was used to select optimal number of fuzzy rules which was found to be 17. Fig. 5.13 shows the plot of the artificial data for two input dimensions (x_0 and x_1) against the output dimension which is time t. The other input variables are set equal to 5 to show the non-linearity of the data at these values. Fig. 5.14 shows the data distribution of the times and one input variable (x_o).

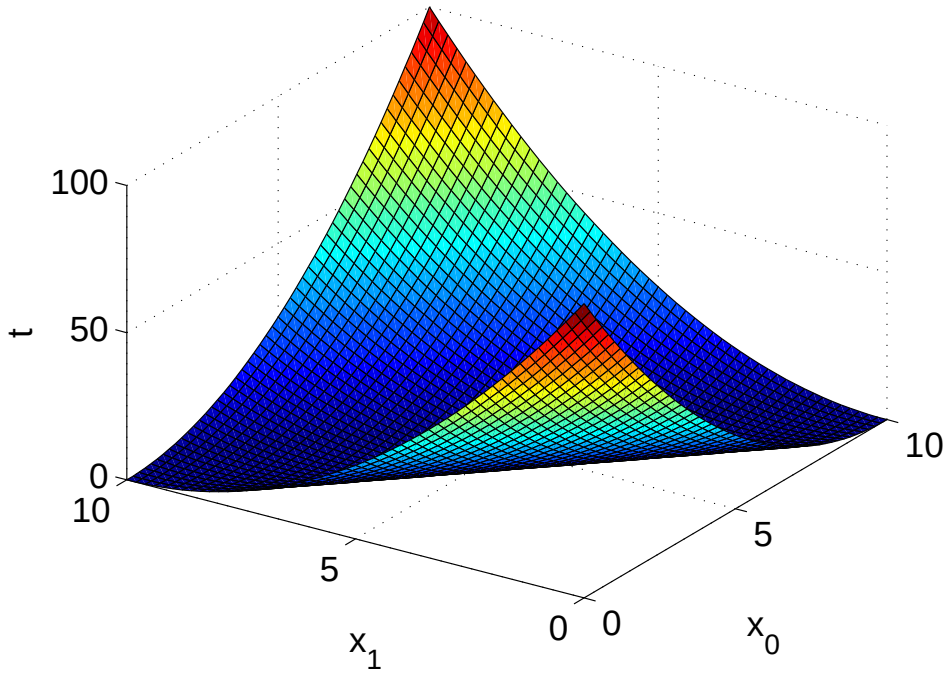


Figure 5.13: Plot of two input variables (x_0 and x_1) for the artificial data to display non-linearity. The other variables were set equal to be equal 5.

5.5 Proposed New Method for risk Prognostics

5.5.1 Modelling Framework

It is proposed to **integrate** the traditional Cox model with fuzzy systems modelling such that the linear part of Cox model is replaced with a flexible fuzzy model. The elicited model is thus able to infer risk groups (low risk or high risk) in a non-linear intepretable fashion from a survival data. Using the hazard for this subdivision is thus natural; A high risk/hazard indicates lower

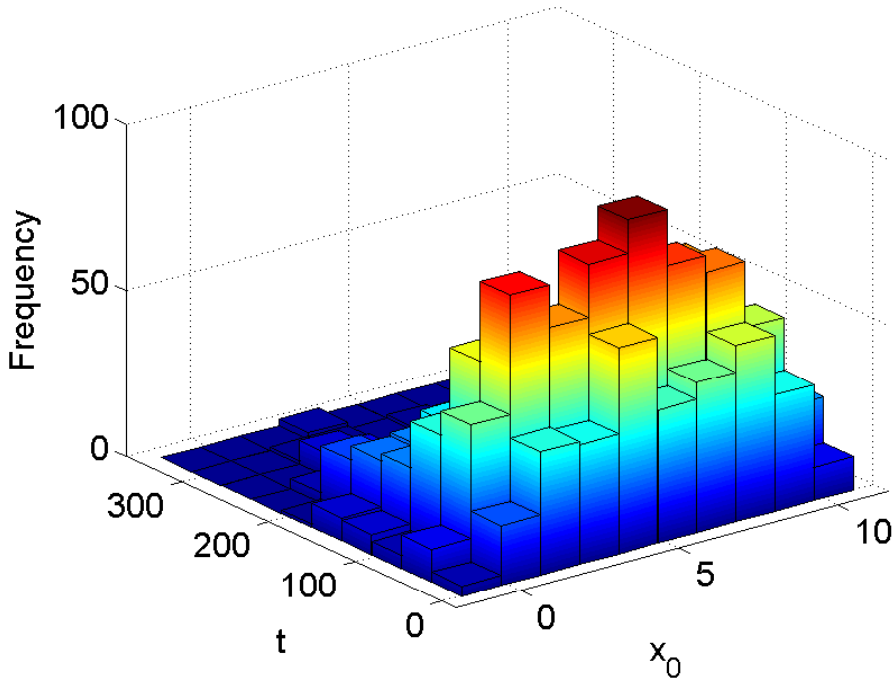


Figure 5.14: Histogram of simulated times and one input variable (x_0). The times in this case are the real times and not the censored times.

failure time[4]. From equation (5.3), the link function, λ, which is also taken to be the prognostic index p, is taken to be the exponent of the output of the fuzzy logic model as defined by the following equation:

$$\lambda_j = \exp(f^j_{FLS}(\mathbf{x}_j, \boldsymbol{\alpha})) \tag{5.11}$$

So that,

$$h_j = h_o(t)\lambda_j = h_o(t)\exp(f^j_{FLS}(\mathbf{x}_i, \boldsymbol{\alpha})) \tag{5.12}$$

[4] $S_i(t) = S_o(t)^{\lambda}$. λ is the link function, $S_i(t)$ is the survival time of individual i and $S_0(t)$ is the baseline survival function. Since $S_0(t) < 1 \; \forall \; t$, higher value for λ i.e. higher risk/hazard and lower survival time.

where $f^j_{FLS}(\mathbf{x}_j, \boldsymbol{\alpha})$ is the output of the *FLS* with parameters $\boldsymbol{\alpha}$ for individual with covariate values $\mathbf{x}_j$. λ_j is the prognostic index of the individual. It can easily be shown that since λ_j is mapped from the output space of the fuzzy model using a monotonic function ($\exp$), then according to Zadeh's extension principle [127], the the membership function is maintained as a result of the one-to-one mapping caused by the monotonic transformation. This will ensure that interpretability is maintained since a linguistic value of the output space maintains the same linguistic value (e.g high, low) in the transformed space.

Lemma. *Consider two universes of discourse X and Y, and a monotonic function $y = f(x)$, then a fuzzy set A in X has same interpretation with image of B in Y, such that $B = f(A)$.*

Proof. Proof follows from Zadeh's extension principle which says that the MF of B:

$$\mu_B(y) = \max_{y=f(x)} \mu_A(x) \tag{5.13}$$

So that the same *MF* (f is a one to one mapping) is retained in B and consequently has the same interpretation. ■

Therefore, it can easily be seen that interpretability is not lost as long as the function is monotonic (an exponential in this study). Fig. 5.15 shows a

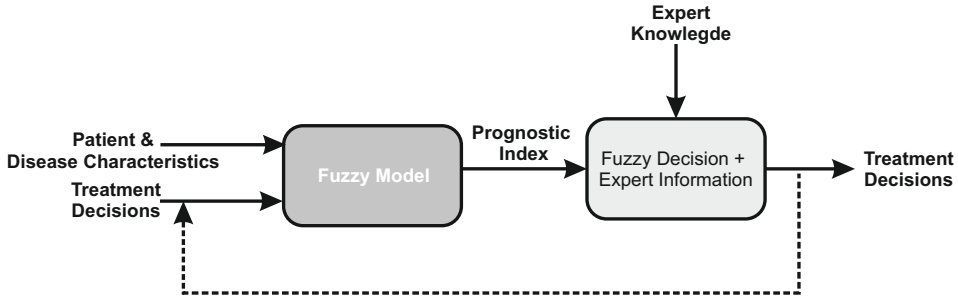


Figure 5.15: Schematic diagram of the proposed modelling framework. The first stage consists of building an interpretable model that classifies a patient as a low or high risk. The elicited models have implicitly added the treatment decisions (if any) which can help proffer the type of treatment in the second stage.

schematic diagram of the proposed modelling framework. The output of the fuzzy model, as previously discussed, is a prognostic index that is indicative of the degree of risk a patient has. This prognostic index can be combined with prior expert information so that a treatment decision can be made to lower the prognostic index and thus recommend a risk management decision.

5.5.1.1 Training the Model

In survival data modelling with right censored observations, one observes the data of size N and a triple $(T_j, \delta_j, \mathbf{x}_j)$, $j = 1, \cdots, N$, where T_j is the time observed for individual j, $\delta_j \in \{0, 1\}$ is an event indicator ($\delta_j = 1$ means event times T_j are observed directly, $\delta_j = 0$ represents a situation where event times T_j are censored.) for individual j. Estimation of parameters is usually

carried-out by maximizing an objective function such as the c-index [160], area under the survival curve [50], partial likelihood [161], etc.

The fuzzy model is trained according to the partial likelihood methodology given in equation (5.14) which is based on the same premise of the orignal Cox model. The idea of the partial likelihood methodology is to find those parameters that ensure that a patient with a lower event time is ranked higher (higher prognostic index) than one with a higher event time. With the unique event times (censored times are excluded) ordered, the Cox's partial likelihood methodology is thus a rank-based objective function defined as follows:

$$L(\boldsymbol{\alpha}) = \prod_{j=1}^{R} \frac{\exp(f(\mathbf{x}_j, \boldsymbol{\alpha}))}{\sum_{m=1}^{n_j} \exp(f(\mathbf{x}_m, \boldsymbol{\alpha}))} \tag{5.14}$$

where R is the number of unique event times and $f(\mathbf{x}_j, \boldsymbol{\alpha})$ is the output of a *FLS* with parameters $\boldsymbol{\alpha}$ for individual j with covariate values $\mathbf{x}_j$. n_j represents the number of individuals at risk at event time t_j. This number includes the censored individuals that have not failed at this time. The parameters of the *FLS* are found by maximising equation (5.14). Optimisation of this equation was performed using a genetic algorithm (*GA*). The negative log-likelihood (*NLL*) of equation (5.14) was used to change the objective function to a minimisation problem so as to be able to use our prior designed *GA* software. The

NLL can be shown to be as follows:

$$NLL(\boldsymbol{\alpha}) = \sum_{j=1}^{R} f(\mathbf{x}_j, \boldsymbol{\alpha}) - \sum_{j=1}^{R} \log\{\sum_{m=1}^{n_j} f(\mathbf{x}_m, \boldsymbol{\alpha})\} \tag{5.15}$$

Details of the derivations of partial likelihood formula for the Cox models from data likelihood may be found in [161, 120].

5.5.1.2 Optimisation and Validation Details

GA is an evolutionary algorithm which imitates survival of the fittest [162] phenomenon. The procedure for optimising the survival fuzzy model is outlined as follows:

1. Initialise the population N_p randomly. This represents N_p fuzzy models. The fitness of the population is evaluated according to the objective function defined in (5.15).

2. Select two parents according to the tournament method [163] where the two parents are selected in turn from the higher fitness value of two randomly chosen chromosomes.

3. *Crossover* along a randomly chosen point in the genome vector is performed with probability P_c in two selected parents to form two offspring. With probability $1 - Pc$, no crossover takes place, so the two parents are cloned to become the two off-springs.

4. Every parameter of the *FLS* is subject to random *mutation* with probability P_m such that $\alpha_h = \alpha_h + \zeta$. α_h is a specific *FLS* parameter and ζ is random Gaussian number of mean zero and standard deviation one.

5. Both sets of parents and offsprings are reinserted into the population pool. N_p is kept constant by discarding the two chromosomes with the worst fitness value.

6. The optimisation run is stopped when N_g generations have passed or if change in average fitness value after each generation is less than $\epsilon = 10^{-6}$.

Using a non-linear model, such as *FLS* together with *GA*, can quickly lead to over-fitting of the training data set. To circumvent this problem, each data set was divided into two. $2/3$ for training and remaining the $1/3$ for testing. k-fold cross validation (Fig. 5.16) was performed on the training data sets to select model with the best generalisation ability $k = 5$). The maximum allowable number of rules was set to be 20 to manage computational time. In the artificial data set, the fuzzy model with 17 rules was found to have the best result based on the k-fold cross validation while the fuzzy model with 18 rules was found for the *BCa* data set. The testing data set was used to show

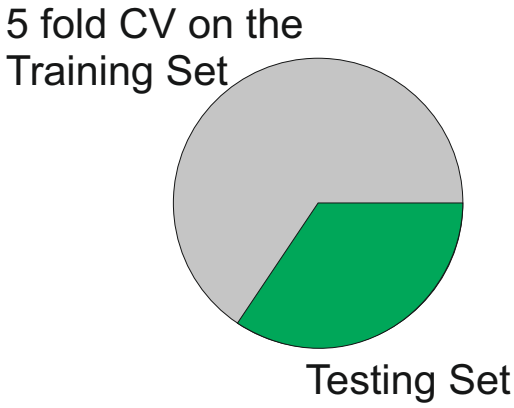


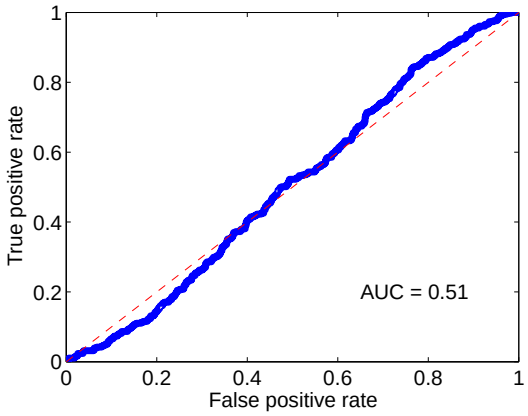
Figure 5.16: Data was partitioned with 33% in the test set and 66% in the training set. In the pilot study the best performance on the training set was found using 5-fold cross-validation.

the generalisation ability on a data set not used in the training procedure.

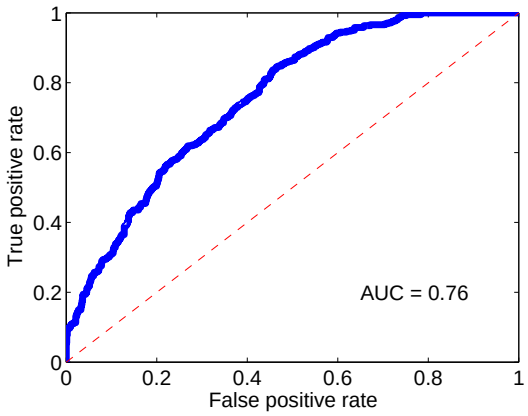
5.6 Results & Discussions

5.6.1 Artificial Data

Fig. 5.17 shows a comparison of the *ROC* performances using the Cox model and the proposed fuzzy modelling framework on the test data set of the artificial data. With an *AUC* of 0.51, the Cox model is no better than a random guess of the risk groups of patients. This is because a linear Cox model has been used on a highly non-linear data set. The fuzzy model has an *AUC* of 0.76 which represents almost 50% improvement in modelling accuracy. The



(a) Cox model.



(b) Fuzzy modelling results.

Figure 5.17: Comparison of *ROC* performances on the artificial test data set. Results show that the Linear model is not better than random guessing of the risk groups while using fuzzy modelling shows promising results. Positive means high risk (low survival time) and negative means low risk (high survival time). The red dotted line is the $y = x$ line. *ROC* curves that coincide with this line are no better than random guesses.

degraded modelling performance on the Cox model is not unexpected since the generated data set is highly non-linear. The concordance index shown in Table 5.4 gives similar result. The fuzzy model expectedly (being a non-linear model) outperforms the Cox model on artificial data. A closer inspection of the performance of the proposed fuzzy modelling framework on the artificial data shown in Fig. 5.18 reveals a model that is capable of inferring correctly both the risk groups as well as the predicted median survival times. It can also be observed that the model (Fig. 5.18D) is also able to handle the fact that some times were not observed directly and is able to infer correctly what would have been the observed survival times had they not been censored. The proposed modelling framework has been able to handle this missing data problem.

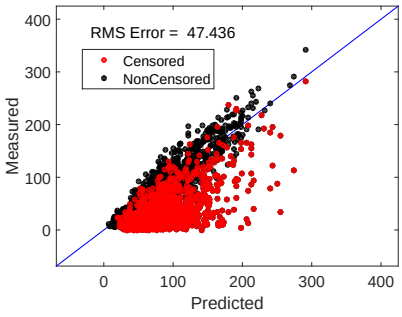
In contrast, as shown in Fig. 5.19, the Cox model performs very badly using this data set and is not able to infer risk groups correctly as well as to predict the median times.

5.6.2 Bladder Cancer Data

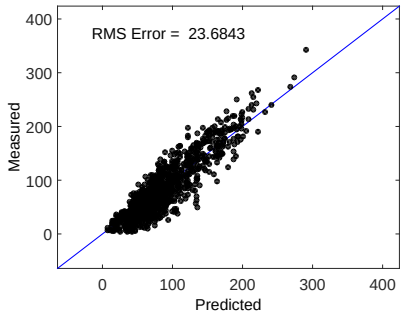
Fig. 5.20 shows the *ROC* curves on the testing data set of the *BCa* data using the proposed fuzzy model and Cox model. The *AUC* for the Cox model is 0.83 while that for the fuzzy model is 0.91 which is an improvement on the Cox model. The concordance index, as shown in Table 5.4, reveals that the proposed fuzzy modelling framework provides about 13% improvement in performance compared with the Cox model.

5.6.2.1 Therapy Recommendation

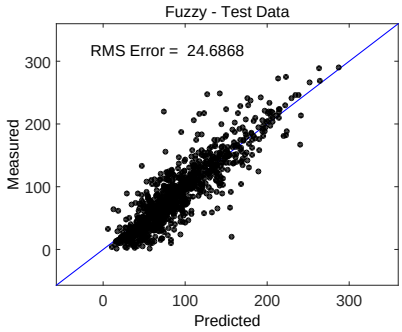
The *ROC* curves of Fig. 5.20 show the performance of the classfier at different selected thresholds for the prognostic index/risk score. In practice, however, only one value of the prognostic index is desired and values greater than this threshold are taken to be high risk patients and vice versa. In this thesis, the 'optimum' point is selected to represent a trade-off between *FP* and *TP* rates using the isocost lines method. Fig. 5.21 shows the distribution of the log of predicted prognostics indices for the testing data set of the *BCa* data. The



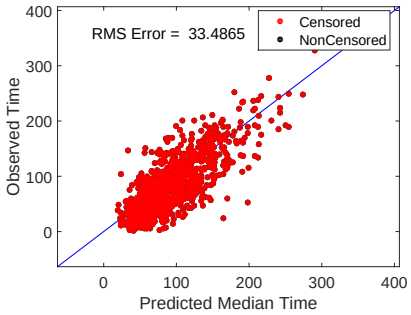
(a) All Training Data (censored and uncensored). Censored observations are indicated by red data points. It can be seen that the fuzzy model tries to predict what the failure time would be if the observation had not been censored.



(b) Training Data (Uncensored Only). The fact that the data contains censored variables does not affect the results of the uncensored observations.

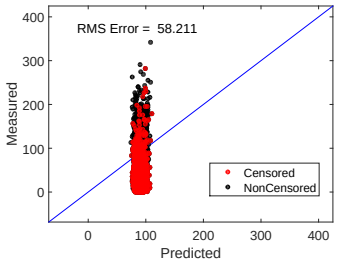


(c) Testing Data. The proposed modelling framework is able to generalise to unseen data. The Testing RMSE is not so much different from the training RMSE, even though the data is highly complex and non-linear.

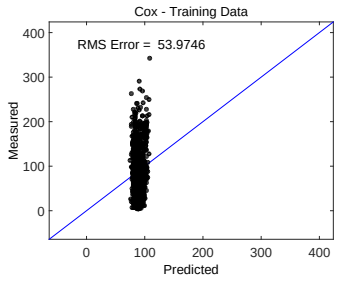


(d) Predicted median times vs. would have been failure time of censored observations. The fuzzy model does well in anticipating what the failure time would have been had the observations not been censored. It should be noted that the "would have been failure time" is not observed in real life data. For censored observations, the censoring times were used in training the fuzzy model as explained.

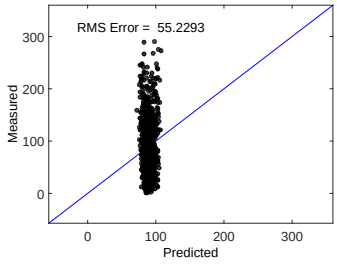
Figure 5.18: Fuzzy Model Prediction Results on Artificial Data set. The Fuzzy model outperforms the traditional Cox model both in terms of predicting performances as well as not allowing censored observations to bias the elicited models.



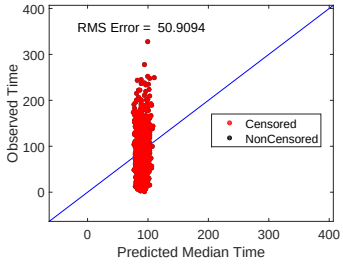
(a) All Training Data (censored and uncensored)



(b) Training Data (Uncensored Only)



(c) Testing Data



(d) Predicted median times vs. would have been failure time of censored observations.

Figure 5.19: Cox Model Prediction Results on Artificial Data set.

optimum operating point (marked o in Fig. 5.20) was found to be 0.4. Patients with prognostics indices greater than this threshold are high risks patients and vice versa.

Table 5.2 shows the confusion matrix at the selected optimum point ($FP = 0.13$ and $TP = 0.74$). It can be observed from this table that 167 patients are predicted to being high risk patients and 513 patients to being low risk. From the risk groups, one patient each is selected at random for analysis. The cystectomy and radiotherapy values as well as their prognostic

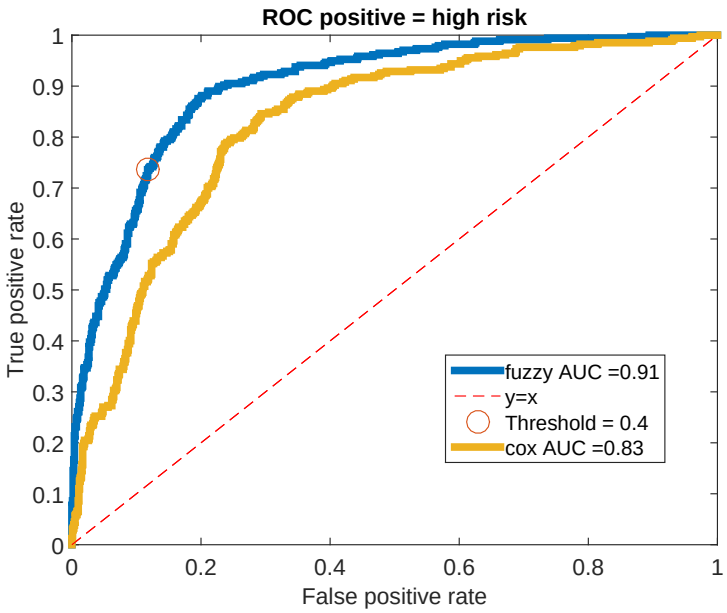


Figure 5.20: Comparison of *ROC* performances on the *BCa* test data set. Point marked o is the optimum point based on the isocost line method.

Table 5.2: Confusion matrix at the selected optimum point.

		True Class	
		Low Risk	High Risk
Hypothesized Class	Low Risk	471	44
	High Risk	64	103

indices of these two selected patients are shown in Table 5.3.

On the one hand it can be observed that where patient 1 (high risk patient with prognostic index of 20.09) is seen to neither having undergone radiotherapy nor having had cystectomy performed. On the other hand patient 2 (low-risk patient with prognostic index of 0.01) had cystectomy performed.

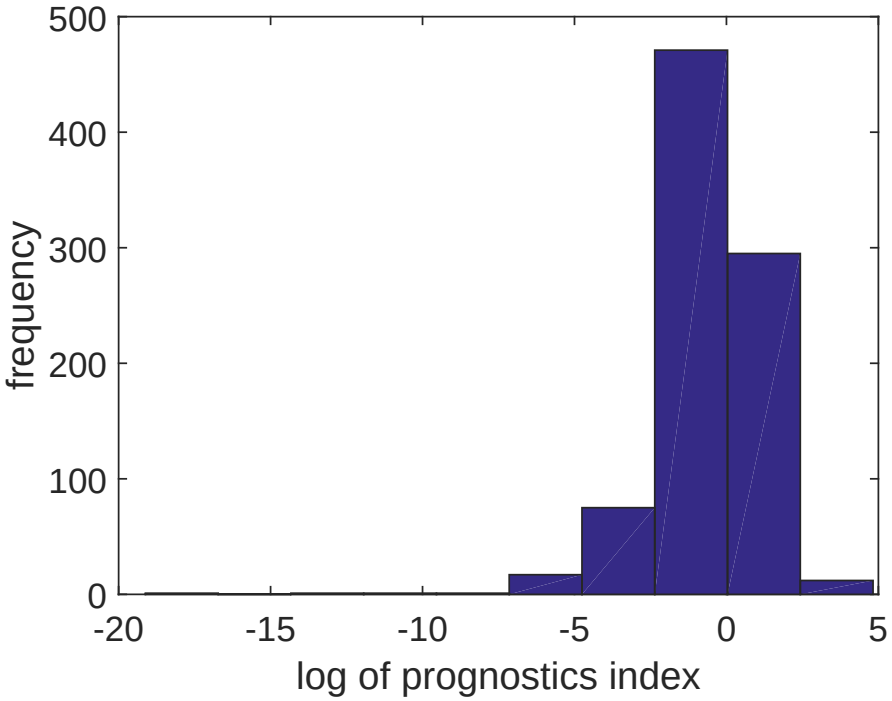


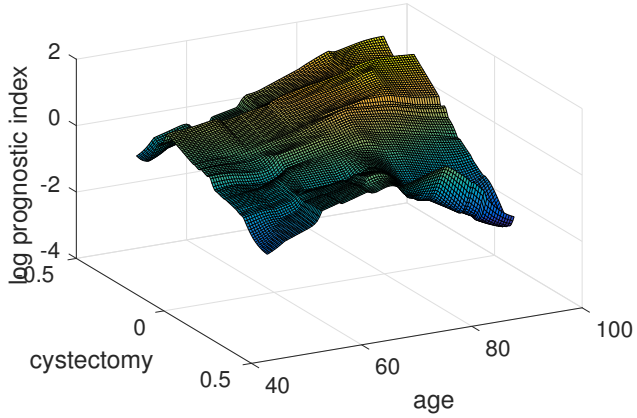
Figure 5.21: Distribution of log of the predicted prognostics indices using the fuzzy model (testing data).

Table 5.3: Selected patients characteristics

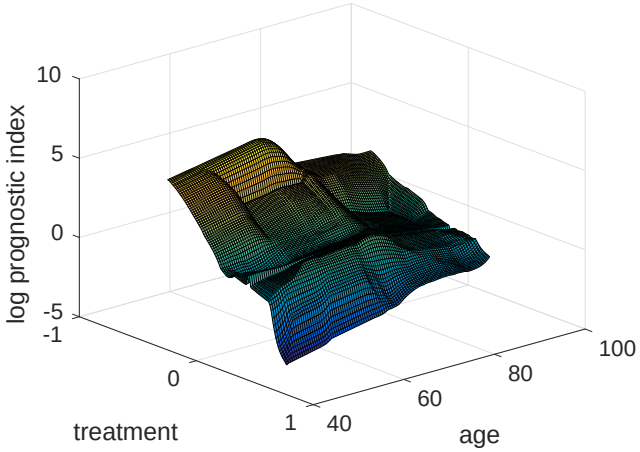
	Cystectomy	**Radiotherapy**	**Prognostic Index**	**Observed Event Times**
Patient 1	No	No	20.090 (High Risk)	10.8 months
Patient 2	Yes	No	0.013 (Low Risk)	133.4 months

Further investigation of patient 1 reveals that the prognostic index reduced to 10.090 when the cystectomy variable was changed to 'yes'. However, the patient still remains high-risk since the prognostics index is still above the threshold of 0.4.

Overall, of all the patients that received radiotherapy or cysctectomy and are in the low-risk group, 45% would have been in the high-risk group had either therapies not been performed. Also, had either of the therapies been performed on the high-risk group, 24% would have moved to the low-risk group had radiotherapy been performed. Additionally, as can be observed in the surface plots of Fig. 5.22, the fuzzy model has inferred a risk index that is a highly non-linear function of the treatments and age if other variables are set to the baseline (zero). A patient who undergoes radiotherapy (positive values = treatment administered, negative values = no treatment) tends to have a lower risk index and is typically below the threshold ($\log(0.4) = -0;916$), hence in the low-risk group. Radiotherapy seems to be more effective in younger patients. Having cystectomy performed seems to represent a more effective treatment for older patients.

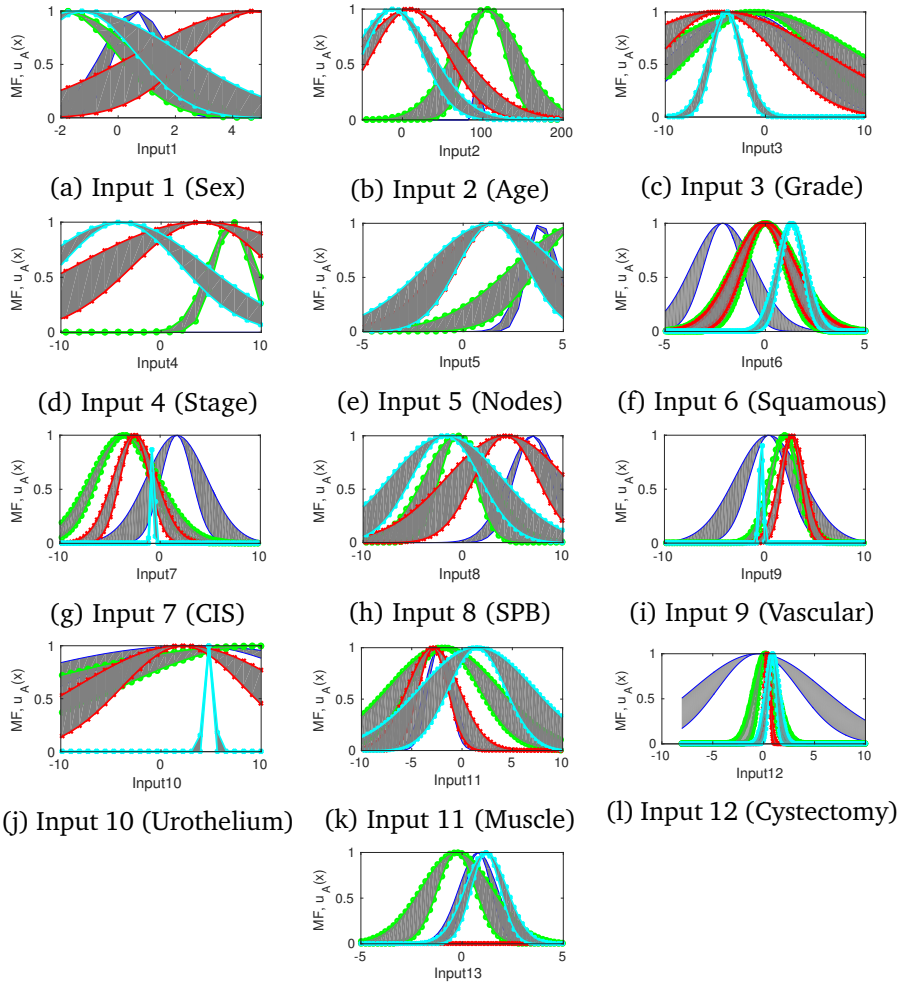


(a) cystectomy and age vs log prognostic index



(b) Radiotherapy and age vs log prognostic index

Figure 5.22: The figure shows how age and treatment decisions affect the prognostic index. Negative values of treatments mean no treatment and positive values mean treatment was performed. It can be observed that radiotherapy is more effective for younger patients. For cystectomy, the treatment seems more effective for older patients. The threshold value for the prognostic index, as discussed, is 0.4 ($\log(0.4) = -0.916$).



(a) Input 1 (Sex)

(b) Input 2 (Age)

(c) Input 3 (Grade)

(d) Input 4 (Stage)

(e) Input 5 (Nodes)

(f) Input 6 (Squamous)

(g) Input 7 (CIS)

(h) Input 8 (SPB)

(i) Input 9 (Vascular)

(j) Input 10 (Urothelium)

(k) Input 11 (Muscle)

(l) Input 12 (Cystectomy)

(m) Input 13 (Radiotherapy)

Figure 5.23: The figure shows a sample fuzzy rule-base. Only the first four rules are displayed for each input dimension.

Table 5.4: Concordance index values on the Training and Testing data sets using Cox modelling and Fuzzy modelling on the artificial data and *BCa* data.

		Training	**Testing**
Artificial Data	*Linear*	0.52	0.51
	Fuzzy	0.74	0.74
***BCa* Data**	*Linear*	0.71	0.70
	Fuzzy	0.82	0.81

5.7 Summary

This chapter has extensively analysed the bladder cancer data used in this research. A framework for risk elicitation based on integrating the Cox and fuzzy modelling frameworks proposed in earlier chapters has been discussed. This framework, which includes the treatments as inputs, allows clinicians to investigate beforehand how different therapies affect the risk of death of a patient diagnosed with bladder cancer. However, dynamic information (such as relapse from bladder cancer) was not considered. The way this dynamic information will be incorporated into the proposed modelling framework is the subject of the next chapter.

Chapter 6

Dynamic Risk Modelling

The previous chapter discussed how survival analysis and fuzzy models may be combined to elicit interpretable and flexible risk mapping from patients diagnosed with *BCa*. Survival analysis allows for inference in a database when censored responses are present. However, the assumption that the censored responses should be independent must be made. In this chapter, this non-parametric assumption and how tenable it is in the context of the *BCa* data used in this research project are investigated. The role of competing risk (the possibility of a patient dying from other causes) is also investigated in detail in this chapter. Finally, a dynamic risk model (using the principles of multi-state modelling and the fuzzy modelling frameworks proposed in earlier chapters) is elicited in this chapter. The model allows to handle and incorpo-

rate dynamic information, such as a patient relapsing from *BCa* in the course of follow-up. The chapter concludes with testing the proposed framework on the *BCa* data.

6.1 Independent Censoring

It is worth investigating how the competing risk methodology could be used to improve elicited models. The elicited model in chapter 5 assumes the independence between censored times and event times. This is simply means that in the hypothetical situation where both censored and event times are observed, both times must be independent in the statistical sense i.e. the knowledge of one does not determine the other. However as discussed in [164], when there is a possibility of patients dying from other causes, then the independence assumption is likely not valid. The independence assumption simply states that if one removes other causes, then the risk of dying from the other remains the same. This assumption may not be justifiable in *BCa* risk prognosis since it is possible that the a patient may have died from a disease aggravated by *BCa*. Competing risk analysis, as discussed in the next section, allows one to handle such scenarios.

6.2 Competing Risks of Bladder Cancer Data

The competing risk[1] approach to modelling makes fewer assumptions than standard survival techniques which could potentially result in increased modelling performances. Competing risks help one handle cases where more than one cause of failure is possible (Fig. 6.1). For example, in this research project, a patient diagnosed with *BCa* or may die from the disease or die from other causes[2]. As is common in the literature, patients dying from the *BCa* have been grouped in Diseases Specific Mortality (***DSM***) and those dying from other causes have been grouped in **O**ther **C**ause **M**ortality (*OCM* (Fig. 6.2). In the *BCa* data, there are 1420 (48.66%) censored observations (lost due to follow up), 885 (30.33%) *OCM* and 613 deaths from *BCa* (*DSM*).

6.2.1 Kaplan-Meier Methods on the Bladder Cancer Data (Revisited)

In Section 5.3, the Kaplan-Meier analysis was performed on the *BCa* data to find the risk of death from the *BCa* without accounting for covariates as well as on different strata of the covariates. It was concluded in the same section that this method is not fit for multidimensional data analysis such as the case

[1] Competing risk analysis dates back to the 18th century when Bernoulli studied the effect on mortality rates if small pox was eradicated [165].

[2] All other possible causes of death have been grouped and are referred to throughout as **other causes**.

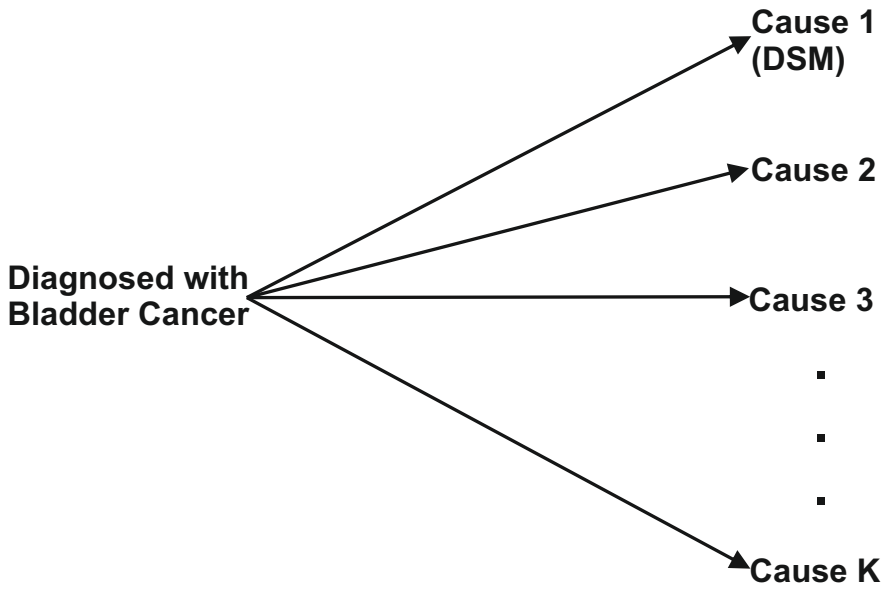


Figure 6.1: Competing risk allows one to analyse scenarios where there is more than one possible cause of failure which precludes the other from happening. K is the number of failures.

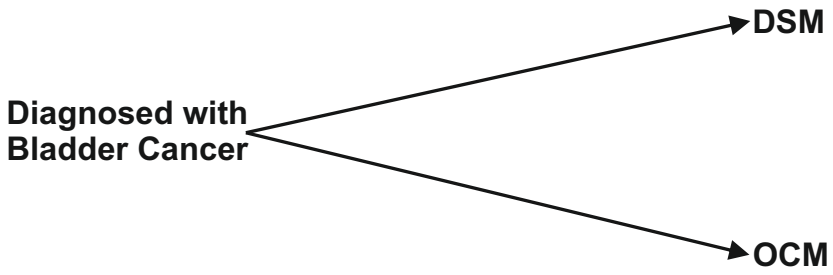


Figure 6.2: Only one of cause (*DSM*) is most times of interest with the other $K-1$ possible causes lumped into one (other causes mortality (*OCM*)).

of the *BCa* data in this research. In this section, another problem encountered when using the standard survival techniques (such as the Kaplan-Meier and the Cox) is discussed. The problem is called the violation of the independent censoring assumption.

Using Kaplan-Meier method to find *OCM* and *DSM* separately provides a simple and intuitive way to understand this problem. The probabilities of death from each of the causes (*DSM* and *OCM*) estimated using the Kaplan-Meier method was shown in chapter 5 (Fig. 5.11) but presented again (in more details) in Fig. 6.3. It should be noted that when one event is of interest, other causes are taken as censored. For example, when finding the survival curve for patients dying from other causes using the Kaplan-Meier method, patients who die from *BCa* are taken as censored.

It can be observed from Fig. 6.3 that at 190 months, the survival function value was 0.2466 (0.7534 probability of death) for *OCM* and the survival function value for *DSM* was 0.6828 (0.3172 probability of death). The ***central criticism*** of the Kaplan-Meier method and indeed all other methods, which assume independent censoring for competing risk problems, is that probabilities of death from from both causes which was $(0.7534 + 0.3172 = 1.0706)$ is greater than 1. This does not make sense as a person who has died from one cause will certainly NOT die from another cause. To remedy this prob-

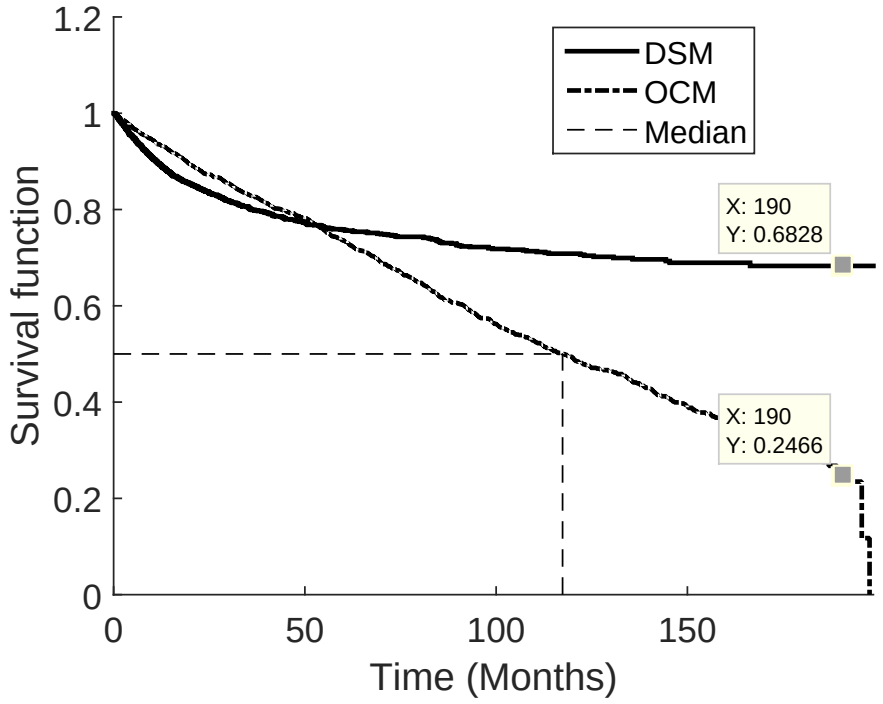


Figure 6.3: The Kaplan Meier Analysis for OCM and DSM.

lem, there have been some approaches introduced in the literature. These approaches, including limitations and strengths, are presented and discussed in the next section.

6.2.2 Approaches to Competing Risk Analysis

A number of approaches have been identified in the literature (see [166] for a review) for handling the competing risk problem. By far the most common is finding the cumulative incidence function (*CIF*) from the so-called cause spe-

cific hazard (*CSH*) [83]. The *CIF* of cause k is defined as the ($Pr(T \leqslant t, D = k)$) of dying from cause k before time i.e $1-$ *CIF* is different from the survival/distribution function in that it specifies a joint distribution of the cause of event and time of event. The *CSH* estimates the risk of dying from a specific cause (*OCM* and *DSM*) at an instant of time as is defined mathematically as follows:

$$\lambda_k(t) = \lim_{\delta t \to 0} \frac{Pr(t \leq T \leq t + \delta t, D = k | T \geq t)}{\delta t} \tag{6.1}$$

$$D = \begin{cases} 1 & \text{if } T_i \text{ is due to DSM} \\ 2 & \text{if } T_i \text{ is due to OCM} \end{cases} \tag{6.2}$$

where $k \in 1, 2$ which represents the different causes of death. $k = 1$ represents death from *BCa* (DSM) and $k = 2$ represents death from other causes (OCM).

It becomes apparent, as shown in [83], that estimating the *CSH* follows exactly the same procedure as estimating the hazard in equation (5.14) where one treats as censored events due to other causes. Therefore, estimating the *CSH* for each cause simply proceeds by following the same procedure outlined for estimating the hazard in Chapter 5. The hazards (*CSH*) is still relevant and may be used as the risk of death as described in Chapter 5. Causes not of interest are simply treated as censored variables. However, calculating the *CSH* allows to glean out more useful information from the data e.g. by calculating

the *CIF*.

Calculating the *CIF* ($I_k(t)$) is given as follows:

$$I_k(t) = \int_0^t \lambda_k(s)S(s)ds \tag{6.3}$$

if $\Lambda_k(t)$ is the cumulative *CSH* defined as follows:

$$\Lambda_k(t) = \int_0^t \lambda_k(s)ds \tag{6.4}$$

then $S(s)$ is the probability of not failing from any cause defined as follows:

$$S(t) = \exp(-\sum_{k=1}^{K} \Lambda_k(t)) \tag{6.5}$$

$S(t)$ is the overall probability no failure from ANY cause at time t.

As in chapter 5, the non-parametric approach to estimating these quantities is usually favoured since they make less assumptions than parametric methods about the distribution of the data.

6.2.2.1 Using the Cumulative Incidence Function as risk estimates

The *CIF* makes use of both the risk of death from *BCa* as well as the risk of death from other causes which could potentially lead to a more accurate modelling. The *CIF* risk framework is shown in Fig. 6.4 with the prognostic indices (derived in chapter 5) serving as inputs to the framework. The reason for this can be seen in equations (6.3) and (6.5) where it can be seen that

Figure 6.4: *CIF* risk framework.

the *CIF* of a particular cause of death is not only dependent on the *CSH* of the particular cause but also on the the *CSHs* of other causes. The output is the *CIF* for the *BCa* patient with the baseline shown in Fig. 6.5. The baseline represents the scenario where all the covariates have values of zero.

In contrast to Fig. 6.3, Fig. 6.5 has probabilities whose sum is less than 1 which is more sensible.

For a particular patient with covariate values given by the vector $\mathbf{x}$, the prognostic index is calculated discussed in chapter 5 so that the cause specific prognostic index is given as follows:

$$\lambda_k(t|\mathbf{x}) = \lambda_{k0}(t)\exp(f^k_{FLS}(\boldsymbol{\alpha}_k, \mathbf{x})) \tag{6.6}$$

where f^k_{FLS} is the fuzzy model for cause k having parameters $\boldsymbol{\alpha}_k$ for a patient with covariate values $\mathbf{x}$.

The *CIF* can now be calculated as given by equation 6.3.

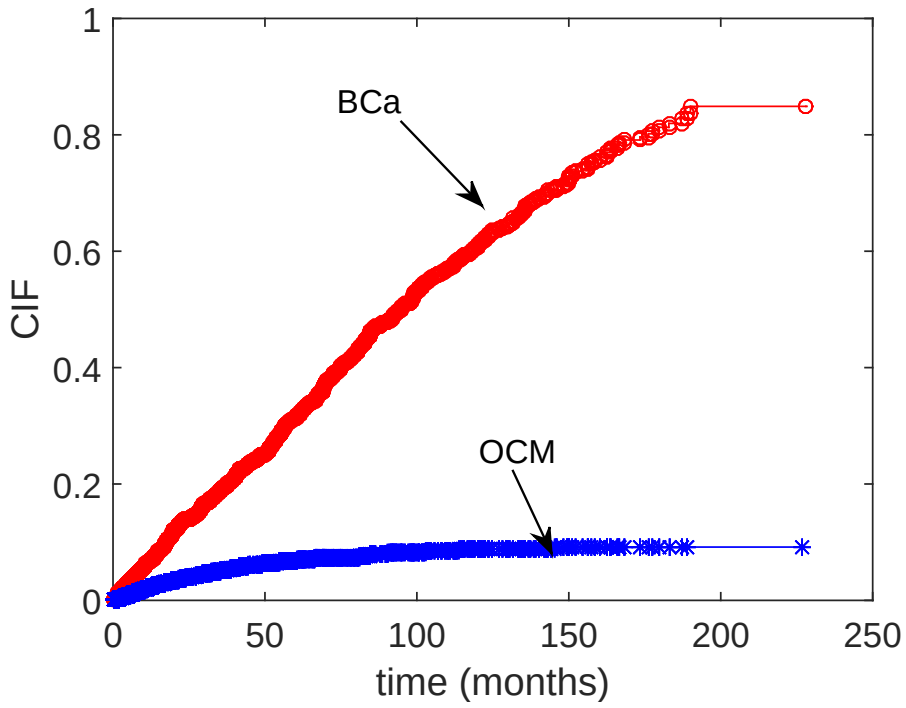


Figure 6.5: *CIF* function baseline for both *BCa* and other causes. The baseline indicates that patient with zero values for covariates (for example, no cystectomy or radiotherapy, zero years) has about 30% chance of dying from *BCa* at 60 months.

As in Chapter 5, the risk of patient may be found by taking a time point (60 months in this study) and patients with *CIF* values greater than a threshold considered high risk while those with lower *CIF* values than the threshold considered low risk. By varying this point, the *ROC* curve may be plotted as discussed in chapter 5. The same training and testing data division used in chapter 5 were used in this section and the same training procedure followed. The *ROC* curve of using the *CIF* as the risk determinant is shown in Fig. 6.6. It

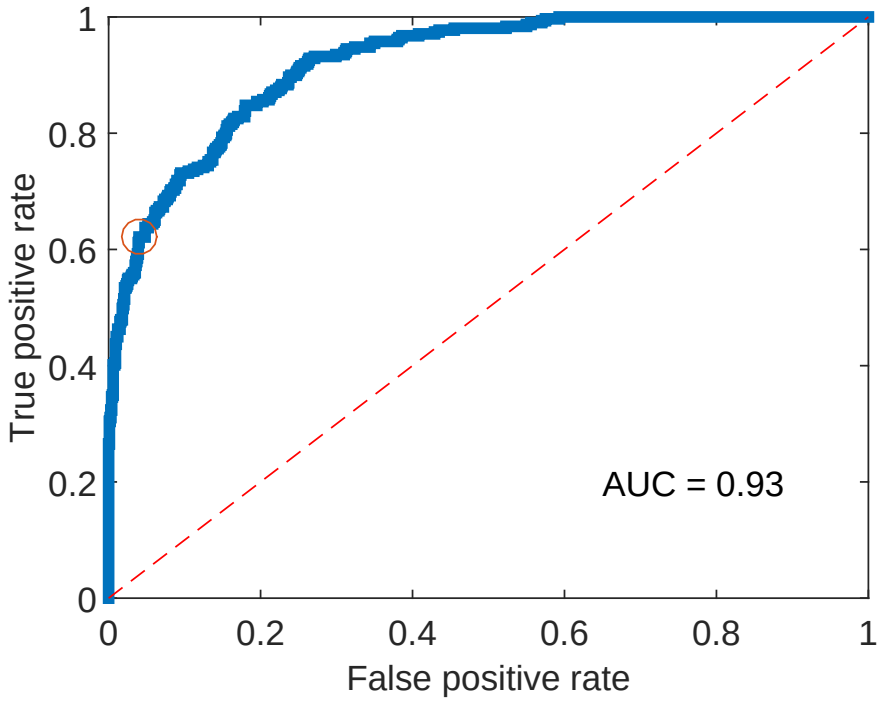


Figure 6.6: *ROC* on the test data using the *CIF* method.

can be seen from the figure that by adequately handling competing risk, there has been an increase in modelling accuracy of about 2%.

6.3 Multistate-Fuzzy Modelling Approach

The *BCa* data contains some vital information which were recorded during follow up of the patients. For example, 1477 (50%) patients were found to relapse from the disease during the course of managing the disease. This new information may significantly affect the risks (prognostic indices) calculated

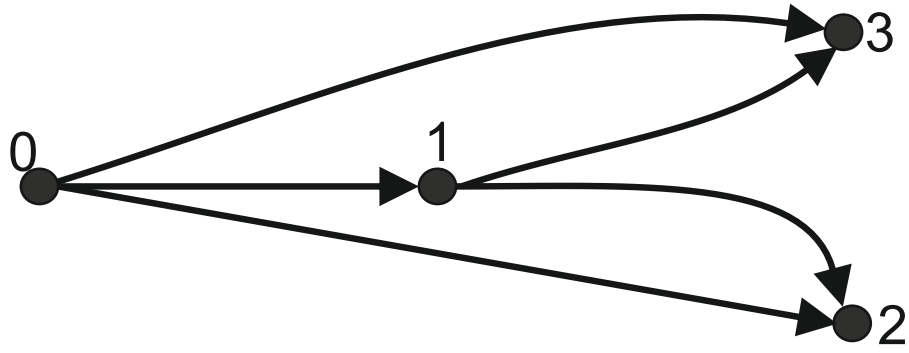


Figure 6.7: *BCa* multistate model. The *BCa* patient could be be seen as potentially moving from states to states. State 0 is living with the *BCa* disease, state 1 is having a relapse from *BCa* after undergoing initial treatment, state 2 is the *OCM* and state 3 is dying from *BCa* (*DSM*). The arrows indicate possible transitions between states. States 2 and 3 are usually called absorbing state since transition out of the two states is not possible i.e. a dead person can no longer be diagnosed with *BCa*.

in Chapter 5. To handle this, it is often desirable to consider the patient as moving within states as Fig. 6.7 illustrates.

A patient who has just been diagnosed with Bladder Cancer and has undergone therapy moves into state 0. The therapy recommendation could be as obtained in chapter 5 or via some other mechanism. This patient has three possible course which are: relapse from the disease (move into state 1), die from Bladder cancer (state 3) or die from causes unrelated to *BCa* (move into state 2), this is illustrated in Fig. 6.8.

A patient whose *BCa* is back is known to have significantly worse prognosis and death almost inevitably results even with aggressive therapy. This

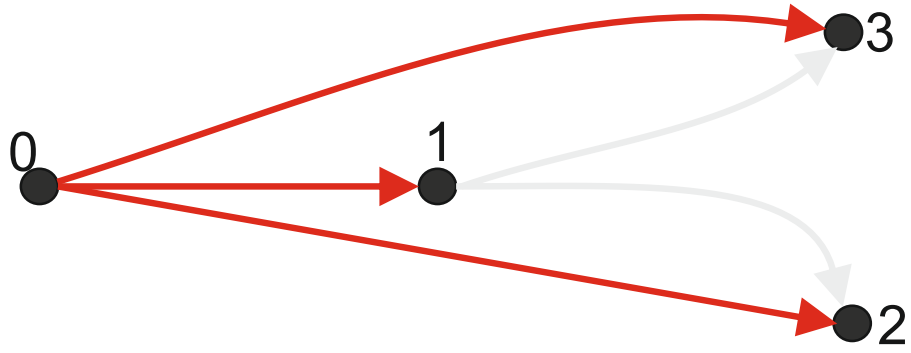


Figure 6.8: Transitions from State 0.

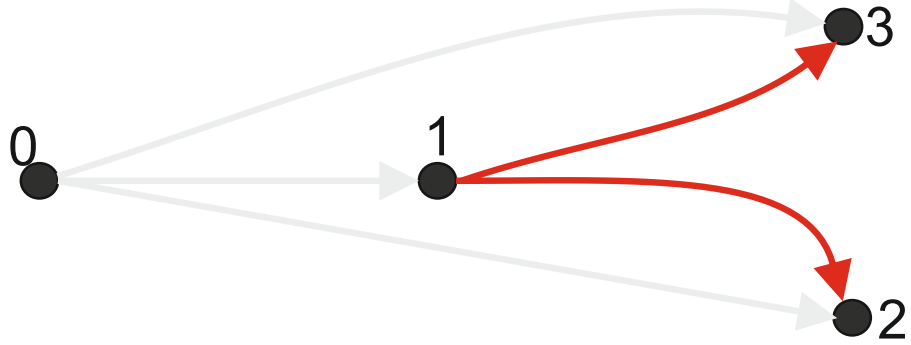


Figure 6.9: Transitions from State 1

modelling framework allows one to model such scenarios. At state 1, the patient could transition to state 2 (die from other causes) or state 3 (die from *BCa*), this is illustrated in Fig. 6.9.

Transitions out of state 2 and 3 are not possible since they are considered to be absorbing states. It turns out that this approach to dynamic risk modelling is similar to the competing risk scenario introduced in the last section (see

[164] and [83] for extensive details). Consider a patient who has just been diagnosed with *BCa* and consequently in state 0. Because this is a competing risk modelling problem, the associated analysis may proceed as described in section 6.2. Therefore, one would have three *CSH*s: $\lambda_{01}, \lambda_{02}$ and λ_{03}.

A patient that enters state 1 (the relapse state) may move into either state 1 or 2 as already stated. This is also a competing risk modelling problem but the patient is only considered after having entered state 2. Two *CSH*s result: λ_{13} and λ_{12} will be estimated. All other *CSH*s are zero as transitions to and from these states are not possible[3]. The analysis proceeds, again, as described in section 6.2.

In summary, five different *CSHs* are calculated each providing a prognostic risk index of moving from one state to another. Distribution of the prognostic indices are shown in Figs. 6.10 - 6.14. A higher prognostic index indicating a high risk of transition between the states in question.

6.3.1 Discussion and Analysis

This section analyses how the risk prognostic indices are affected by age. Figs. 6.15 - 6.19 show how the risk of moving states are affected by age. The risk of moving from initial state) to the relapse state seem to decrease as one

[3]Which is logical since the risk of moving from dying from *BCa* state (State 1) to say relapse state (state 3) is impossible.

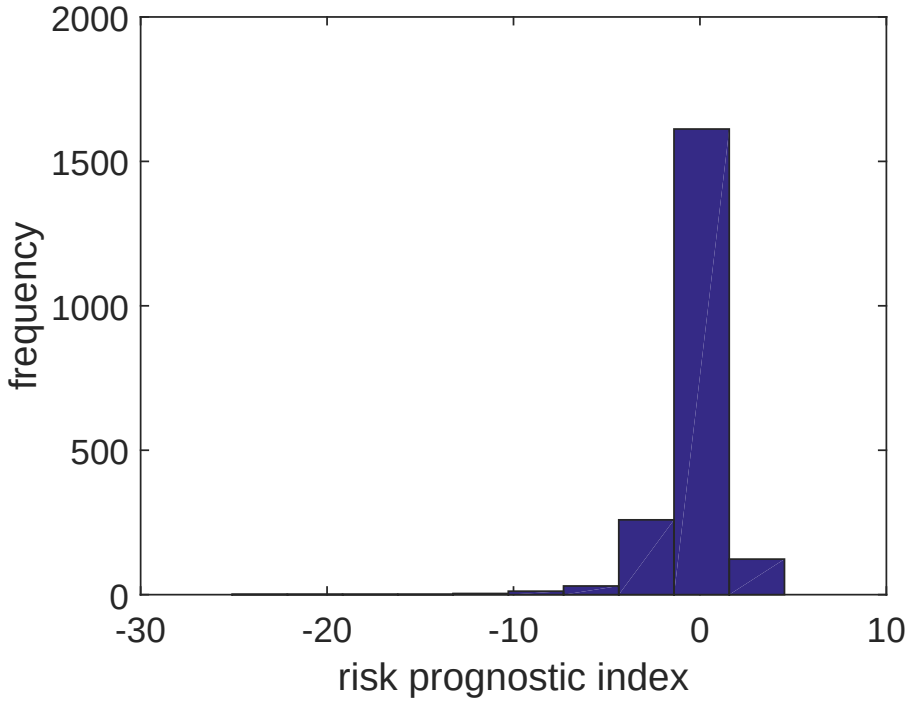


Figure 6.10: Transition from state 0 to 1 risk prognostic index (log).

progresses with age. This may be due to the fact that there are older patients who have a higher risk of death than just relapsing from the disease as shown in Figs. 6.16 (moving from initial state to dying from other causes) and 6.17 (moving from initial state to dying from *BCa*). Therefore, younger patients are more likely to relapse than die from the disease.

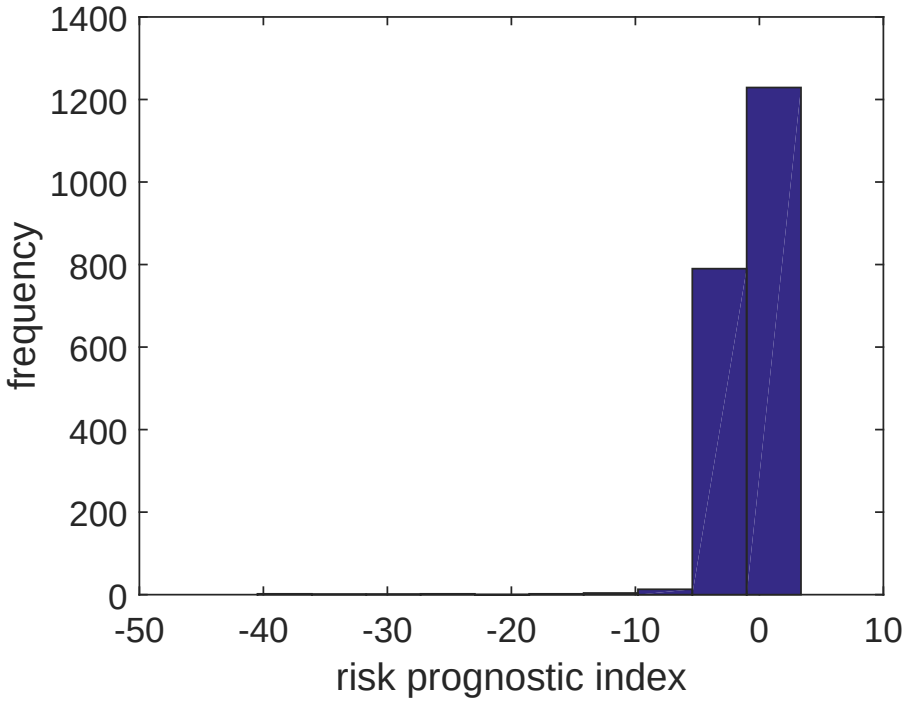


Figure 6.11: Transition from state 0 to 2 risk prognostic index (log).

6.4 Summary

This section has provided a mechanism for incorporating dynamic information during bladder cancer patient's follow-ups. Specifically, the proposed method uses a multistate modelling approach where patients who relapse are assumed to have entered into a new state. This approach allows the static model proposed in Chapter 5 to be incorporated into this dynamic modelling framework.

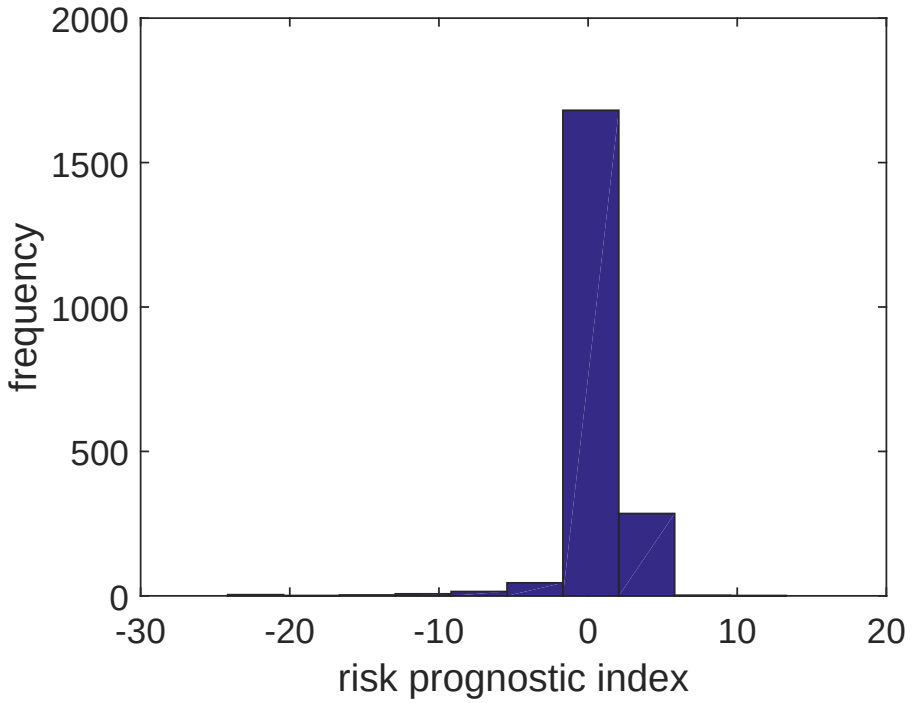


Figure 6.12: Transition from state 0 to 3 risk prognostic index (log).

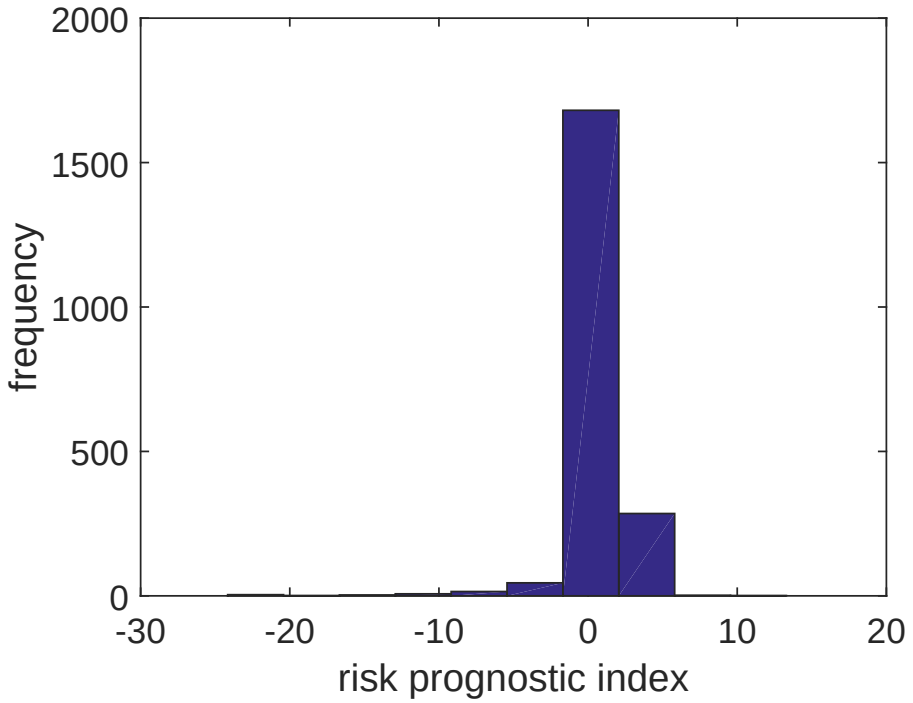


Figure 6.13: Transition from state 0 to 2 risk prognostic index (log).

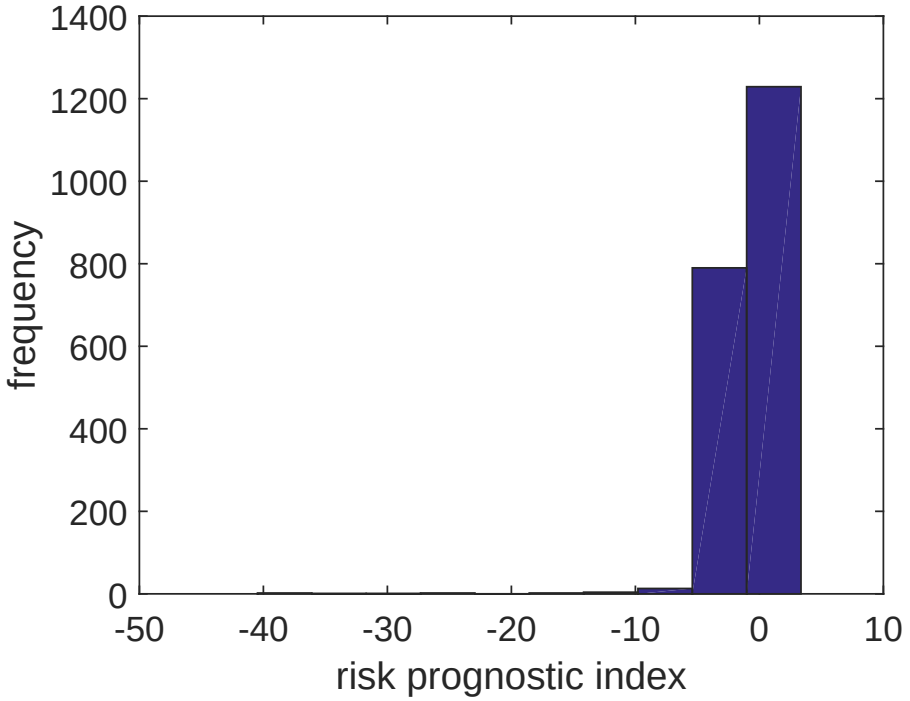


Figure 6.14: Transition from state 0 to 1 risk prognostic index (log).

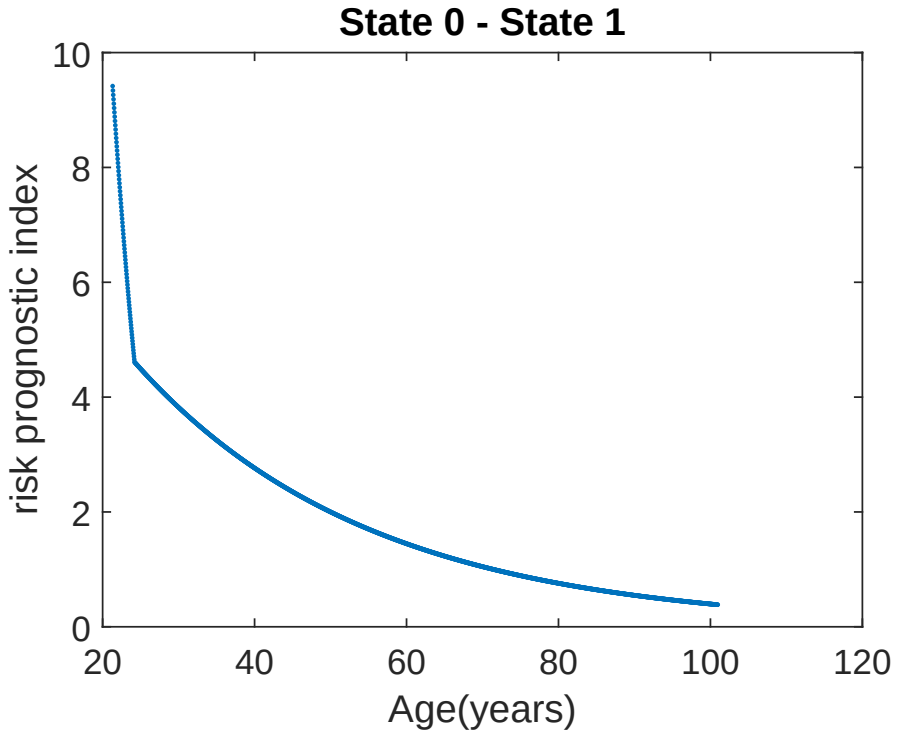


Figure 6.15: Transition from state 0 to 1 risk prognostic index as a function of age in years.

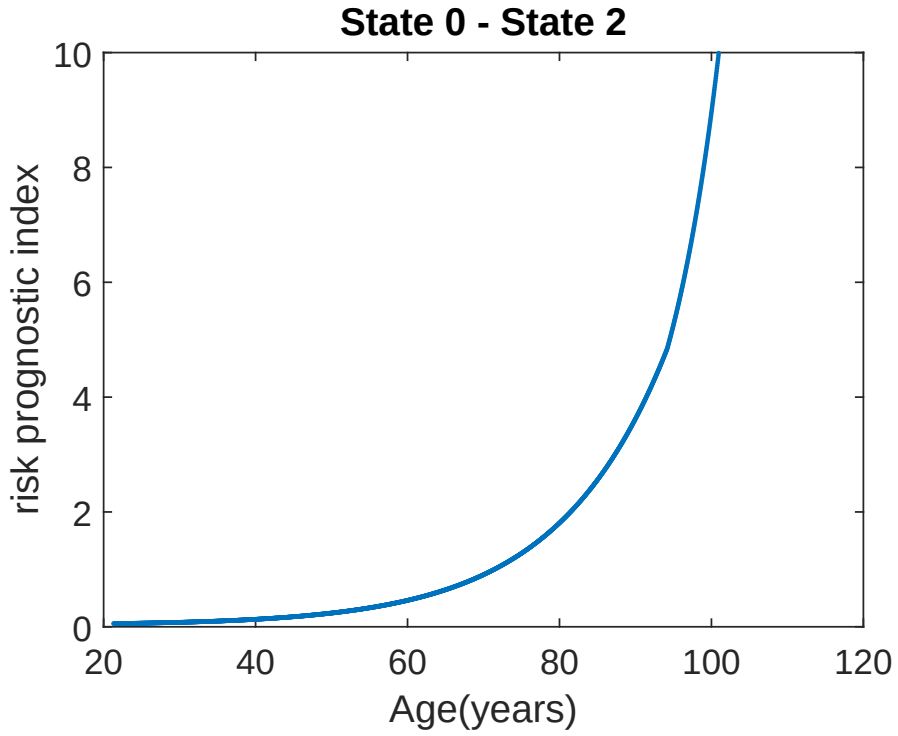


Figure 6.16: Transition from state 0 to 2 risk prognostic index as a function of age in years.

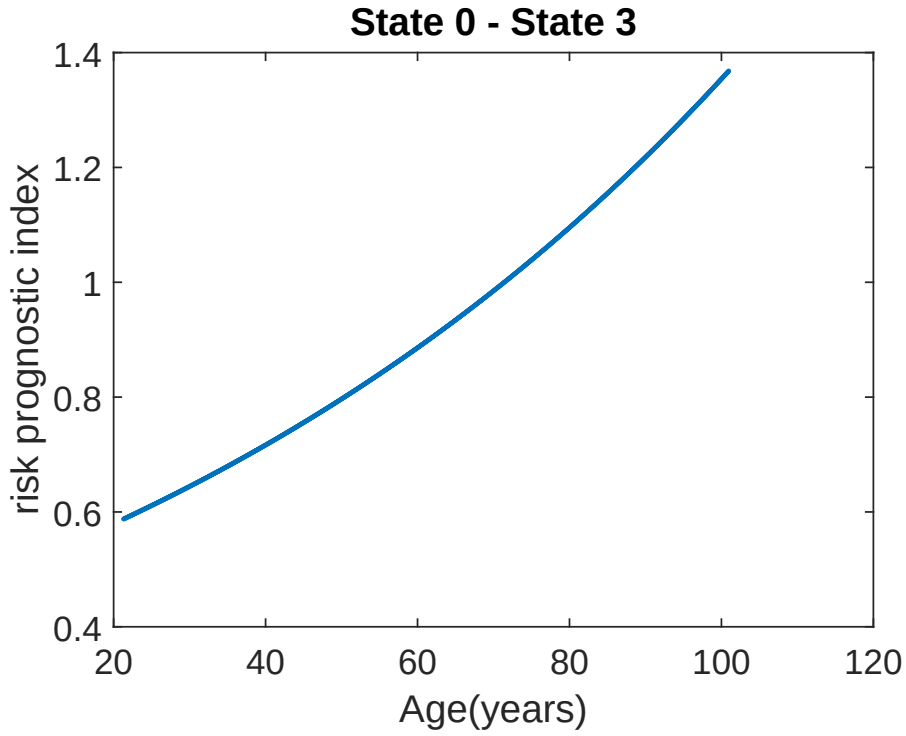


Figure 6.17: Transition from state 0 to 3 risk prognostic index as a function of age in years.

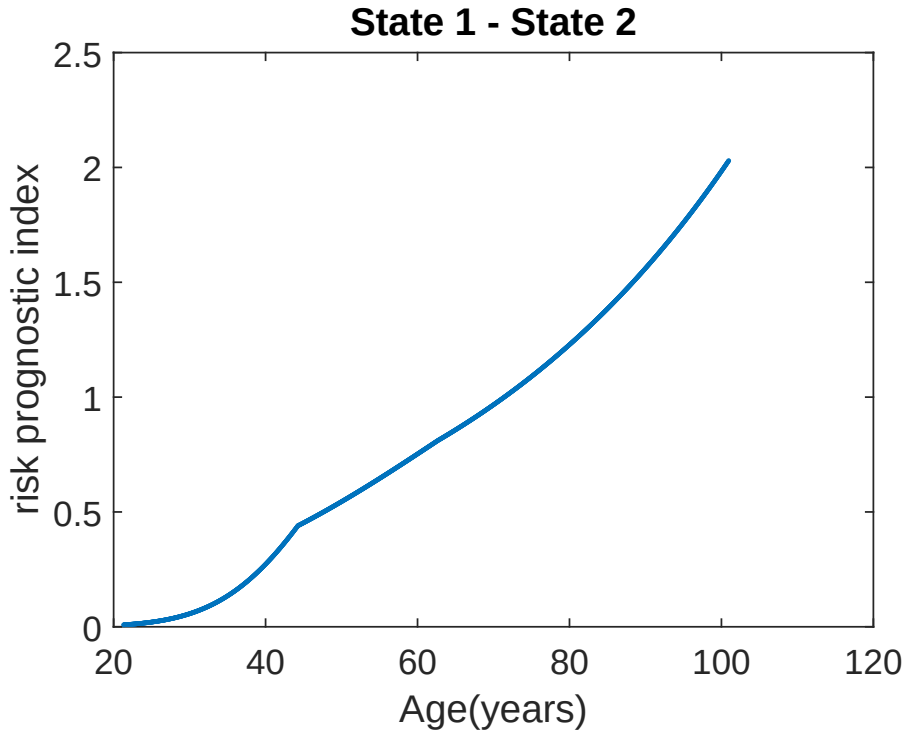


Figure 6.18: Transition from state 0 to 2 risk prognostic index as a function of age in years.

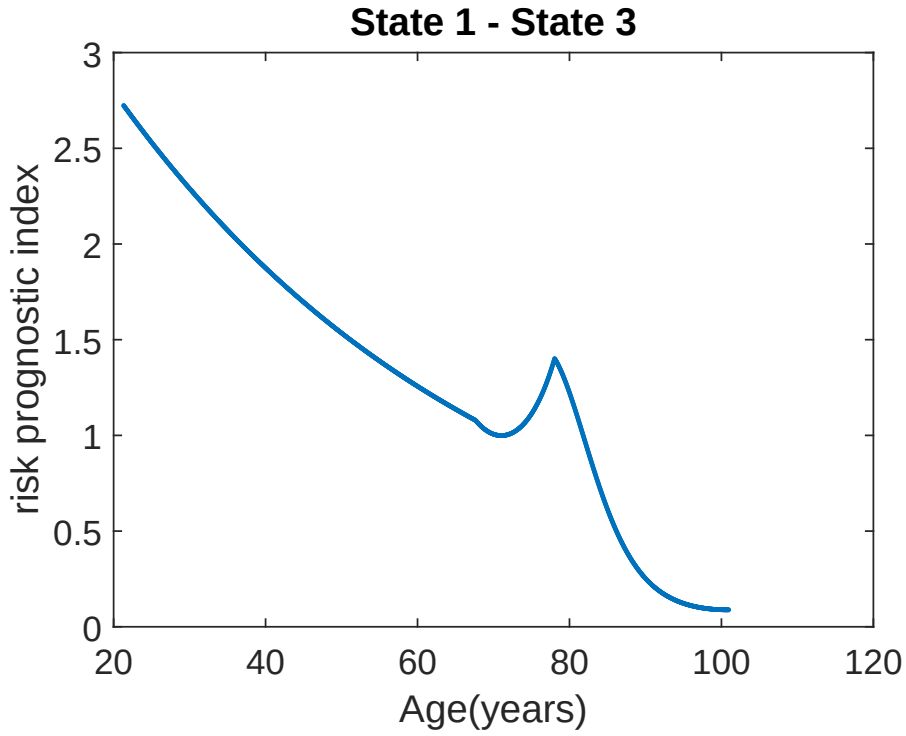


Figure 6.19: Transition from state 0 to 1 risk prognostic index as a function of age in years.

Chapter 7

Conclusions

In this research, new frameworks for understanding the risk of death after a patient has been diagnosed with bladder cancer (*BCa*) through the use of interval type-2 fuzzy logic systems (*IT2 FLSs*) has been proposed.

The thesis starts by proposing a new method to elicit interval type-2 fuzzy models from data. The approach presented is based on the interval type-2 fuzzy clustering algorithm. The proposed method allows one to systematically determine the antecedent and consequent parameters of the *IT2 FLS* before a gradient-based optimisation algorithm is derived to further tune these parameters in a bid to elicit more accurate fuzzy models. This is a shift in conventional *IT2 FLS* parameters initialisation where they are in most cases randomly initialised. This systematic and elegant initialisation leads to better

accuracy and improved performance of subsequent optimisation stages.

Tested extensively on synthetic and artificial datasets, the proposed framework outperforms existing neuro-fuzzy modelling frameworks of a similar type in both prediction and generalization capabilities.

To overcome the limitations of the existing fuzzy modelling technologies to handle random uncertainties in the data modelling process, the research proposes the incorporation of Bayesian modelling with fuzzy modelling. This approach was shown to be able to handle random uncertainties and provide accurate predictions with confidence bands.

The *BCa* database has been analysed extensively in Chapters 5 and 6. A new framework which involves the integration of the Cox modelling approach with the interval type-2 fuzzy logic system approach has been proposed. The aim of this integration strategy is to take advantage of the simplicity and mathematical convenience of the traditional Cox modelling framework and the interpretability and non-linear modelling capabilities of interval type-2 fuzzy models. This synergistic modelling architecture has been used to predict the prognostics index of an individual diagnosed with *BCa* which is indicative of the level of risk. The higher the prognostic index, the higher the risk and vice versa. The different treatments administered have been used as input variables to the predictive model in order to allow the clinicians determine

how such therapies have affected the prognostic index so as to manage risk. With an Area under receiver operating characteristic curve of 0.93, the proposed modelling framework has outperformed the traditional Cox model and artificial intelligence methods such as the neural network.

In the concluding parts of the thesis, the proposed modelling framework has been extended so as to be able to handle dynamic information (such as relapse from *BCa* during follow-up). The approach generalises the fuzzy model elicited in earlier chapters by using the multistate modelling principle whereby patients are taken to move from one state to another during follow-up.

7.1 Future Recommended Work

As a result of the studies presented in this thesis, the following future research recommendations can be made:

1. The interval type-2 fuzzy modelling approach used throughout this thesis is a simplification of the more general type-2 fuzzy models. The general type-2 approach allows for more degrees of freedom in fuzzy models but are very computationally taxing as discussed in the thesis. How to reduce this computational burden is an evolving research area and it would be interesting to

see how the general type-2 approach can be used instead of the interval type-2.

2. To incorporate patient's dynamic information, a multistate modelling approach has been used in this thesis. Specifically, the Markov assumption has been made to simplify analysis. Future work should consider relaxing this assumption (such as semi-Markov or non-Markov) which may allow for a more accurate incorporation of dynamic information. Additionally, it would be interesting to apply the Bayesian approach proposed in chapter 4 to the fuzzy modelling frameworks proposed chapters 5 and 6 in a bid to quantity and manage the uncertainties in the modelling process.

Bibliography

[1] Karl H Pang and James WF Catto. Bladder cancer. *Surgery (Oxford)*, 31(10):523–529, 2013.

[2] Cancer research uk. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bladder-cancer. Accessed: 20-02-2013.

[3] Ryoichi Nakashima, Chisaki Watanabe, Eriko Maeda, Takeharu Yoshikawa, Izuru Matsuda, Soichiro Miki, and Kazuhiko Yokosawa. The effect of expert knowledge on medical search: medical experts have specialized abilities for detecting serious lesions. *Psychological Research*, 79(5):729–738, 2015.

[4] Margaret A Knowles and Carolyn D Hurst. Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. *Nature Reviews Cancer*, 15(1):25–41, 2015.

[5] AP Noon, PC Albertsen, F Thomas, DJ Rosario, and JWF Catto. Competing mortality in patients diagnosed with bladder cancer: evidence

of undertreatment in the elderly and female patients. *British Journal of Cancer*, 108(7):1534–1540, 2013.

[6] L Scrucca, A Santucci, and F Aversa. Competing risk analysis using r: an easy guide for clinicians. *Bone Marrow Transplantation*, 40(4):381–387, 2007.

[7] Núria Porta, M Luz Calle, Núria Malats, and Guadalupe Gómez. A dynamic model for the risk of bladder cancer progression. *Statistics in Medicine*, 31(3):287–300, 2012.

[8] Robin Hesketh. *Introduction to Cancer Biology*. Cambridge University Press, 2013.

[9] History of cancer. `http://www.cancer.org/acs/groups/cid/documents/webcontent/002048-pdf.pdf`. Accessed: 20-02-2016.

[10] The role of tobacco in the development of cancer. `https://sites.duke.edu/seektobacco/2-the-role-of-tobacco-in-the-development-of-cancer/the-content/`. Accessed: 20-04-2016.

[11] Stephen B Edge and Carolyn C Compton. The american joint committee on cancer: the 7th edition of the ajcc cancer staging manual and the future of tnm. *Annals of Surgical Oncology*, 17(6):1471–1474, 2010.

[12] Bladder cancer handbook". `http://www.med.umich.edu/cancer/files/bladder-cancer-handbook.pdf`. Accessed: 23-02-2016.

[13] Craig M Steinmaus, Sandra Nuñez, and Allan H Smith. Diet and bladder cancer: a meta-analysis of six dietary variables. *American Journal of Epidemiology*, 151(7):693–702, 2000.

[14] European association of urology. `http://www.uroweb.org/guidelines/online-guidelines/`. Accessed: 21-02-2016.

[15] Sherri Rose. Mortality risk score prediction in an elderly population using machine learning. *American Journal of Epidemiology*, 177(5):443–452, 2013.

[16] Margaret A Shipp, Ken N Ross, Pablo Tamayo, Andrew P Weng, Jeffery L Kutok, Ricardo CT Aguiar, Michelle Gaasenbeek, Michael Angelo, Michael Reich, Geraldine S Pinkus, et al. Diffuse large b-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nature Medicine*, 8(1):68–74, 2002.

[17] John D Isaacs and Gianfranco Ferraccioli. The need for personalised medicine for rheumatoid arthritis. *Annals of the Rheumatic Diseases*, 70(1):4–7, 2011.

[18] Adam Hedgecoe. *The politics of personalised medicine: pharmacogenetics in the clinic*. Cambridge University Press, 2004.

[19] Carlo Gambacorti-Passerini. Part i: Milestones in personalised medicine—imatinib. *The Lancet Oncology*, 9(6):600, 2008.

[20] Illhoi Yoo, Patricia Alafaireet, Miroslav Marinov, Keila Pena-Hernandez, Rajitha Gopidi, Jia-Fu Chang, and Lei Hua. Data mining in healthcare

and biomedicine: a survey of the literature. *Journal of Medical Systems*, 36(4):2431–2448, 2012.

[21] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. The kdd process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11):27–34, 1996.

[22] Jiawei Han, Micheline Kamber, and Jian Pei. *Data mining: concepts and techniques*. Elsevier, 2011.

[23] Ronald J Brachman, Tom Khabaza, Willi Kloesgen, Gregory Piatetsky-Shapiro, and Evangelos Simoudis. Mining business databases. *Communications of the ACM*, 39(11):42–48, 1996.

[24] Margaret H Dunham. *Data mining: Introductory and advanced topics*. Pearson Education India, 2006.

[25] Christopher M Bishop. *Pattern recognition and machine learning*, volume 1. Springer, 2006.

[26] Grainne Kerr, Heather J Ruskin, Martin Crane, and Padraig Doolan. Techniques for clustering gene expression data. *Computers in Biology and Medicine*, 38(3):283–293, 2008.

[27] Jin Hwan Do and D Choi. Clustering approaches to identifying gene expression patterns from dna microarray data. *Molecules and Cells*, 25(2):279, 2008.

[28] Laura J Van't Veer, Hongyue Dai, Marc J Van De Vijver, Yudong D He, Augustinus AM Hart, Mao Mao, Hans L Peterse, Karin van der Kooy,

Matthew J Marton, Anke T Witteveen, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415(6871):530–536, 2002.

[29] Sotiris B Kotsiantis, I Zaharakis, and P Pintelas. Supervised machine learning: A review of classification techniques, 2007.

[30] Melissa Romeo, F Burden, M Quinn, B Wood, and D McNaughton. Infrared microspectroscopy and artificial neural networks in the diagnosis of cervical cancer. *Cellular and Molecular Biology (Noisy-le-Grand, France)*, 44(1):179–187, 1998.

[31] G Ball, Shahid Mian, F Holding, RO Allibone, J Lowe, S Ali, G Li, S McCardle, Ian O Ellis, Colin Creaser, et al. An integrated approach utilizing artificial neural networks and seldi mass spectrometry for the classification of human tumours and rapid identification of potential biomarkers. *Bioinformatics*, 18(3):395–404, 2002.

[32] Sergey Aleynikov and Evangelia Micheli-Tzanakou. Classification of retinal damage by a neural network based system. *Journal of Medical Systems*, 22(3):129–136, 1998.

[33] Ryan Potter. Comparison of classification algorithms applied to breast cancer diagnosis and prognosis. In *Industrial Conference on Data Mining-Posters and Workshops*, pages 40–49, 2007.

[34] Igor Kononenko, Ivan Bratko, and Matjaž Kukar. Application of machine learning to medical diagnosis. *Machine Learning and Data Mining: Methods and Applications*, 389:408, 1997.

[35] Ashutosh Sharma and Rob J Roy. Design of a recognition system to predict movement during anesthesia. *Biomedical Engineering, IEEE Transactions on*, 44(6):505–511, 1997.

[36] Andrew J Einstein, Hai-Shan Wu, Miguel Sanchez, and Joan Gil. Fractal characterization of chromatin appearance for diagnosis in breast cytology. *The Journal of Pathology*, 185(4):366–381, 1998.

[37] MR Brickley, JP Shepherd, and RA Armstrong. Neural networks: a new technique for development of decision support systems in dentistry. *Journal of Dentistry*, 26(4):305–309, 1998.

[38] J Ross Quinlan. Classification procedures and their application to chess end games. *Machine Learning: An Artificial Intelligence Approach*, page 463, 2013.

[39] J Ross Quinlan. *C4. 5: programs for machine learning*. Elsevier, 2014.

[40] Vladimir Vapnik. Pattern recognition using generalized portrait method. *Automation and Remote Control*, 24:774–780, 1963.

[41] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.

[42] Vladimir Vapnik. *The nature of statistical learning theory*. Springer Science & Business Media, 2013.

[43] Eric Schiffer, Antonia Vlahou, Andreas Petrolekas, Konstantinos Stravodimos, Robert Tauber, Jürgen E Geschwend, Jochen Neuhaus, Jens-Uwe Stolzenburg, Mark R Conaway, Harald Mischak, et al. Prediction of muscle-invasive bladder cancer using urinary proteomics. *Clinical Cancer Research*, 15(15):4935–4943, 2009.

[44] Abdul Majid, Safdar Ali, Mubashar Iqbal, and Nabeela Kausar. Prediction of human breast and colon cancers from imbalanced data using nearest neighbor and support vector machines. *Computer Methods and Programs in Biomedicine*, 113(3):792–808, 2014.

[45] Norbert Bergner, Bernd FM Romeike, Rupert Reichart, Rolf Kalff, Christoph Krafft, and Jürgen Popp. Tumor margin identification and prediction of the primary tumor from brain metastases using ftir imaging and support vector machines. *Analyst*, 138(14):3983–3990, 2013.

[46] Hong Hu, Jiuyong Li, Ashley Plank, Hua Wang, and Grant Daggard. A comparative study of classification methods for microarray data analysis. In *Proceedings of the fifth Australasian conference on Data mining and analystics-Volume 61*, pages 33–37. Australian Computer Society, Inc., 2006.

[47] Paul R Harper. A review and comparison of classification algorithms for medical decision making. *Health Policy*, 71(3):315–331, 2005.

[48] A Colin Cameron and Pravin K Trivedi. *Regression analysis of count data*, volume 53. Cambridge university press, 2013.

[49] Niels Peek, Carlo Combi, Roque Marin, and Riccardo Bellazzi. Thirty years of artificial intelligence in medicine (aime) conferences: A review of research themes. *Artificial Intelligence in Medicine*, 65(1):61–73, 2015.

[50] Jonas Kalderstam, Patrik Edén, and Mattias Ohlsson. Finding risk groups by optimizing artificial neural networks on the area under the survival curve using genetic algorithms. *PloS One*, 10(9):e0137597, 2015.

[51] Theodore Anagnostou, Mesut Remzi, Michael Lykourinas, and Bob Djavan. Artificial neural networks for decision-making in urologic oncology. *European Urology*, 43(6):596–603, 2003.

[52] Paulo JG Lisboa, H Wong, P Harris, and Ric Swindell. A bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer. *Artificial Iintelligence in Medicine*, 28(1):1–25, 2003.

[53] Edwin Pun Hui, Linda KS Leung, Terence CW Poon, Frankie Mo, Vicky TC Chan, Ada TW Ma, Annette Poon, Eugenie K Hui, So-shan Mak, Maria Lai, et al. Prediction of outcome in cancer patients with febrile neutropenia: a prospective validation of the multinational association for supportive care in cancer risk index in a chinese population and comparison with the talcott model and artificial neural network. *Supportive Care in Cancer*, 19(10):1625–1635, 2011.

[54] Azzam FG Taktak, Anthony C Fisher, and Bertil E Damato. Modelling survival after treatment of intraocular melanoma using artificial neural networks and bayes theorem. *Physics in Medicine and Biology*, 49(1):87, 2003.

[55] James WF Catto, Derek A Linkens, Maysam F Abbod, Minyou Chen, Julian L Burton, Kenneth M Feeley, and Freddie C Hamdy. Artificial intelligence in predicting bladder cancer outcome a comparison of neuro-fuzzy modeling and artificial neural networks. *Clinical Cancer Research*, 9(11):4172–4177, 2003.

[56] Ashlesha Jain, Ajita Jain, Sandhya Jain, and Lakhmi Jain. *Artificial intelligence techniques in breast cancer diagnosis and prognosis*, volume 39. World Scientific, 2000.

[57] Abdelghani Bellaachia and Erhan Guven. Predicting breast cancer survivability using data mining techniques. *Age*, 58(13):10–110, 2006.

[58] Harry B Burke, Philip H Goodman, David B Rosen, Donald E Henson, John N Weinstein, Frank E Harrell, Jeffrey R Marks, David P Winchester, and David G Bostwick. Artificial neural networks improve the accuracy of cancer survival prediction. *Cancer*, 79(4):857–862, 1997.

[59] Dursun Delen, Glenn Walker, and Amit Kadam. Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial Intelligence in Medicine*, 34(2):113–127, 2005.

[60] Tar Timothy Chen. History of statistical thinking in medicine. *Advanced Medical Statistics*, 3:11–14, 2003.

[61] Pierre Simon, Marquis de Laplace, Frederick Wilson Truscott, and Frederick Lincoln Emory. *A philosophical essay on probabilities*, volume 166. Dover Publications, 1951.

[62] Isaac Todhunter. *A history of the mathematical theory of probability*. Cambridge University Press, 2014.

[63] J Rosser Matthews. Quantification and the quest for medical certainty. 1995.

[64] Philippe Pinel. *Traite medico-philosophique sur lalienation mentale*. J. Ant. Brosson, 1809.

[65] Joseph Berkson and Robert P Gage. Calculation of survival rates for cancer. In *Proceedings of the Staff Meetings. Mayo Clinic*, volume 25, pages 270–286, 1950.

[66] Sidney J Cutler and Fred Ederer. Maximum utilization of the life table method in analyzing survival. *Journal of Chronic Diseases*, 8(6):699–712, 1958.

[67] Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.

[68] Woonyoung Choi, Sima Porten, Seungchan Kim, Daniel Willis, Elizabeth R Plimack, Jean Hoffman-Censits, Beat Roth, Tiewei Cheng, Mai Tran, I-Ling Lee, et al. Identification of distinct basal and luminal sub-

types of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. *Cancer Cell*, 25(2):152–165, 2014.

[69] Jason A Efstathiou, Daphna Y Spiegel, William U Shipley, Niall M Heney, Donald S Kaufman, Andrzej Niemierko, John J Coen, Rafi Y Skowronski, Jonathan J Paly, Francis J McGovern, et al. Long-term outcomes of selective bladder preservation by combined-modality therapy for invasive bladder cancer: the mgh experience. *European Urology*, 61(4):705–711, 2012.

[70] Devon C Snow-Lisy, Steven C Campbell, Inderbir S Gill, Adrian V Hernandez, Amr Fergany, Jihad Kaouk, and Georges-Pascal Haber. Robotic and laparoscopic radical cystectomy for bladder cancer: long-term oncologic outcomes. *European Urology*, 65(1):193–200, 2014.

[71] John W Boag. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(1):15–53, 1949.

[72] David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 187–220, 1972.

[73] Ross L Prentice. Introduction to cox (1972) regression models and life-tables. In *Breakthroughs in Statistics*, pages 519–526. Springer, 1992.

[74] Nancy Reid. A conversation with sir david cox. *Statistical Science*, 9(3):439–455, 1994.

[75] William R Lane, Stephen W Looney, and James W Wansley. An application of the cox proportional hazards model to bank failure. *Journal of Banking & Finance*, 10(4):511–531, 1986.

[76] David M Mannino, D Thorn, A Swensen, and F Holguin. Prevalence and outcomes of diabetes, hypertension and cardiovascular disease in copd. *European Respiratory Journal*, 32(4):962–969, 2008.

[77] Bas Kreike, Guus Hart, Harry Bartelink, and Marc J van de Vijver. Analysis of breast cancer related gene expression using natural splines and the cox proportional hazard model to identify prognostic associations. *Breast Cancer Research and Treatment*, 122(3):711–720, 2010.

[78] David R Holmes, Vivek Y Reddy, Zoltan G Turi, Shephal K Doshi, Horst Sievert, Maurice Buchbinder, Christopher M Mullin, Peter Sick, Protect AF Investigators, et al. Percutaneous closure of the left atrial appendage versus warfarin therapy for prevention of stroke in patients with atrial fibrillation: a randomised non-inferiority trial. *The Lancet*, 374(9689):534–542, 2009.

[79] Jong Gun Lee, Sue Moon, and Kavé Salamatian. Modeling and predicting the popularity of online contents with cox proportional hazard regression model. *Neurocomputing*, 76(1):134–145, 2012.

[80] Walter R Gilks. *Markov chain monte carlo*. Wiley Online Library, 2005.

[81] Martin J Wainwright and Michael I Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2008.

[82] Joseph G Ibrahim, Ming-Hui Chen, and Debajyoti Sinha. *Bayesian survival analysis*. Wiley Online Library, 2005.

[83] Jan Beyersmann, Arthur Allignol, and Martin Schumacher. *Competing risks and multistate models with R*. Springer Science & Business Media, 2011.

[84] Thomas R Fleming and David P Harrington. *Counting processes and survival analysis*, volume 169. John Wiley & Sons, 2011.

[85] Per Kragh Andersen, Ornulf Borgan, Richard D Gill, and Niels Keiding. *Statistical models based on counting processes*. Springer Science & Business Media, 2012.

[86] Odd Aalen, Ornulf Borgan, and Hakon Gjessing. *Survival and event history analysis: a process point of view*. Springer Science & Business Media, 2008.

[87] Andries P Engelbrecht. *Computational intelligence: an introduction*. John Wiley & Sons, 2007.

[88] Lotfi A Zadeh. The concept of a linguistic variable and its application to approximate reasoning—i. *Information Sciences*, 8(3):199–249, 1975.

[89] John J Hopfield. Artificial neural networks. *Circuits and Devices Magazine, IEEE*, 4(5):3–10, 1988.

[90] Thomas Bäck. Evolutionary algorithms in theory and practice. 1996.

[91] Mahdi Mahfouf, CH Kee, Maysam F Abbod, and Derek A Linkens. Fuzzy logic-based anti-sway control design for overhead cranes. *Neural Computing & Applications*, 9(1):38–43, 2000.

[92] Maysam F Abbod, Diedrich G von Keyserlingk, Derek A Linkens, and Mahdi Mahfouf. Survey of utilisation of fuzzy technology in medicine and healthcare. *Fuzzy Sets and Systems*, 120(2):331–349, 2001.

[93] Mahdi Mahfouf, Maysam F. Abbod, and Derek A. Linkens. A survey of fuzzy logic monitoring and control utilisation in medicine. *Artificial Intelligence in Medicine*, 21(1):27–42, 2001.

[94] Shu-Heng Chen and Paul P Wang. *Computational intelligence in economics and finance*. Springer, 2004.

[95] Bob Djavan, Mesut Remzi, Alexandre Zlotta, Christian Seitz, Peter Snow, and Michael Marberger. Novel artificial neural network for early detection of prostate cancer. *Journal of Clinical Oncology*, 20(4):921–929, 2002.

[96] Ashutosh Tewari and Perinchery Narayan. Novel staging tool for localized prostate cancer: a pilot study using genetic adaptive neural networks. *The Journal of Urology*, 160(2):430–436, 1998.

[97] R Joseph Babaian, Herbert Fritsche, Alberto Ayala, Vijaya Bhadkamkar, Dennis A Johnston, William Naccarato, and Zhen Zhang. Performance of a neural network in detecting prostate cancer in the prostate-specific antigen reflex range of 2.5 to 4.0 ng/ml. *Urology*, 56(6):1000–1006, 2000.

[98] RN Naguib, MC Robinson, DE Neal, and FC Hamdy. Neural network analysis of combined conventional and experimental prognostic markers in prostate cancer: a pilot study. *British Journal of Cancer*, 78(2):246, 1998.

[99] Ashutosh Tewari, Mutta Issa, Rizk El-Galley, Hans Stricker, James Peabody, Julio Pow-Sang, Asim Shukla, Zev Wajsman, Mark Rubin, John Wei, et al. Genetic adaptive neural network to predict biochemical failure after radical prostatectomy: a multi-institutional study. *Molecular Urology*, 5(4):163–169, 2001.

[100] Javed Khan, Jun S Wei, Markus Ringner, Lao H Saal, Marc Ladanyi, Frank Westermann, Frank Berthold, Manfred Schwab, Cristina R Antonescu, Carsten Peterson, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature Medicine*, 7(6):673–679, 2001.

[101] James WF Catto, Maysam F Abbod, Derek A Linkens, and Freddie C Hamdy. Neuro-fuzzy modeling: an accurate and interpretable method for predicting bladder cancer progression. *The Journal of Urology*, 175(2):474–479, 2006.

[102] Ernest J Feleppa, Ronald D Ennis, Peter B Schiff, Cheng-Shie Wuu, Andrew Kalisz, Jeffery Ketterling, Stella Urban, Tian Liu, William R Fair, Christopher R Porter, et al. Ultrasonic spectrum-analysis and neural-network classification as a basis for ultrasonic imaging to target brachytherapy of prostate cancer. *Brachytherapy*, 1(1):48–53, 2002.

[103] Philip S Maclin, Jack Dempsey, Jay Brooks, and John Rand. Using neural networks to diagnose cancer. *Journal of Medical Systems*, 15(1):11–19, 1991.

[104] Ulrich Scheipers, Helmut Ermert, Hans-Joerg Sommerfeld, Miguel Garcia-Schürmann, Theodor Senge, and Stathis Philippou. Ultrasonic multifeature tissue characterization for prostate diagnostics. *Ultrasound in Medicine & Biology*, 29(8):1137–1149, 2003.

[105] Maysam F Abbod, James WF Catto, Derek A Linkens, and Freddie C Hamdy. Application of artificial intelligence to the management of urological cancer. *The Journal of Urology*, 178(4):1150–1156, 2007.

[106] Trevor Hastie and Robert Tibshirani. Generalized additive models. *Statistical Science*, pages 297–310, 1986.

[107] RJ Gray. Hazard estimation with covariates: algorithms for direct estimation, local scoring and backfitting. Technical report, Technical Report, 1994.

[108] R Gentleman and J Crowley. Local full likelihood estimation for the proportional hazards model. *Biometrics*, pages 1283–1296, 1991.

[109] Orna Intrator and Charles Kooperberg. Trees and splines in survival analysis. *Statistical Methods in Medical Research*, 4(3):237–261, 1995.

[110] Chong Gu. Penalized likelihood hazard estimation: a general procedure. *Statistica Sinica*, pages 861–876, 1996.

[111] Frank E Harre, Kerry L Lee, and Barbara G Pollock. Regression models in clinical studies: determining relationships between predictors and response. *Journal of the National Cancer Institute*, 80(15):1198–1202, 1988.

[112] Sylvain Durrleman and Richard Simon. Flexible regression models with cubic splines. *Statistics in Medicine*, 8(5):551–561, 1989.

[113] Trevor J Hastie and Robert J Tibshirani. Generalized additive models, volume 43 of monographs on statistics and applied probability, 1990.

[114] Lynn A Sleeper and David P Harrington. Regression splines in the cox model with application to covariate effects in liver disease. *Journal of the American Statistical Association*, 85(412):941–949, 1990.

[115] Trevor Hastie and Robert Tibshirani. Exploring the nature of covariate effects in the proportional hazards model. *Biometrics*, pages 1005–1016, 1990.

[116] David M Zucker and Alan F Karr. Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. *The Annals of Statistics*, pages 329–353, 1990.

[117] Robert J Gray. Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association*, 87(420):942–951, 1992.

[118] Michal Abrahamowicz, Todd MacKenzie, and John M Esdaile. Time-dependent hazard ratio: modeling and hypothesis testing with applica-

tion in lupus nephritis. *Journal of the American Statistical Association*, 91(436):1432–1439, 1996.

[119] Terry M Therneau and Patricia M Grambsch. *Modeling survival data: extending the Cox model*. Springer Science & Business Media, 2000.

[120] David Collett. *Modelling survival data in medical research*. CRC press, 2015.

[121] Lotfi A Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

[122] Jerry M Mendel. Uncertainty, fuzzy logic, and signal processing. *Signal Processing*, 80(6):913–933, 2000.

[123] Timothy J Ross. *Fuzzy logic with engineering applications*. John Wiley & Sons, 2009.

[124] Li-Xin Wang. *A course in fuzzy systems*. Prentice-Hall press, USA, 1999.

[125] Kevin M Passino, Stephen Yurkovich, and Michael Reinfrank. *Fuzzy control*, volume 42. Citeseer, 1998.

[126] George Klir and Mark Wierman. *Uncertainty-based information: elements of generalized information theory*, volume 15. Springer Science & Business Media, 1999.

[127] Jerry M Mendel. Uncertain rule-based fuzzy logic system: introduction and new directions. 2001.

[128] Nilesh N Karnik and Jerry M Mendel. Centroid of a type-2 fuzzy set. *Information Sciences*, 132(1):195–220, 2001.

[129] Nilesh N Karnik, Jerry M Mendel, and Qilian Liang. Type-2 fuzzy logic systems. *Fuzzy Systems, IEEE Transactions on*, 7(6):643–658, 1999.

[130] Nilesh N Karnik and Jerry M Mendel. Operations on type-2 fuzzy sets. *Fuzzy Sets and Systems*, 122(2):327–348, 2001.

[131] Jerry M Mendel. Type-2 fuzzy sets and systems: an overview. *Computational Intelligence Magazine, IEEE*, 2(1):20–29, 2007.

[132] Jerry M Mendel, Robert I John, and Feilong Liu. Interval type-2 fuzzy logic systems made simple. *Fuzzy Systems, IEEE Transactions on*, 14(6):808–821, 2006.

[133] Janos Abonyi. Introduction. In *Fuzzy Model Identification for Control*, pages 1–21. Springer, 2003.

[134] Janos Abonyi and Balazs Feil. *Cluster analysis for data mining and system identification*. Springer Science & Business Media, 2007.

[135] Shen Wang and Mahdi Mahfouf. A new computationally efficient mamdani interval type-2 fuzzy modelling framework. In *Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on*, pages 1–8. IEEE, 2012.

[136] Miguel Delgado, Antonio F Gómez-Skarmeta, and Fernando Martín. A fuzzy clustering-based rapid prototyping for fuzzy rule-based modeling. *Fuzzy Systems, IEEE Transactions on*, 5(2):223–233, 1997.

[137] Byung-In Choi and Frank Chung-Hoon Rhee. Interval type-2 fuzzy membership function generation methods for pattern recognition. *Information Sciences*, 179(13):2102–2122, 2009.

[138] Cheul Hwang and Frank Chung-Hoon Rhee. Uncertain fuzzy clustering: interval type-2 fuzzy approach to c-means. *Fuzzy Systems, IEEE Transactions on*, 15(1):107–120, 2007.

[139] Seihwan Park and H Lee-Kwang. A designing method for type-2 fuzzy logic systems using genetic algorithms. In *IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th*, pages 2567–2572. IEEE, 2001.

[140] Robert Babuška. *Fuzzy modeling for control*, volume 12. Springer Science & Business Media, 2012.

[141] Maowen Nie and Woei Wan Tan. Towards an efficient type-reduction method for interval type-2 fuzzy logic systems. In *Fuzzy Systems, 2008. FUZZ-IEEE 2008.(IEEE World Congress on Computational Intelligence). IEEE International Conference on*, pages 1425–1432. IEEE, 2008.

[142] Jerry M Mendel. Computing derivatives in interval type-2 fuzzy logic systems. *Fuzzy Systems, IEEE Transactions on*, 12(1):84–98, 2004.

[143] Shen Wang and Mahdi Mahfouf. Multi-objective optimisation for fuzzy modelling using interval type-2 fuzzy sets. In *Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on*, pages 1–8. IEEE, 2012.

[144] G Panoutsos and M Mahfouf. Granular computing and evolutionary fuzzy modelling for mechanical properties of alloy steels. *Dimension*, 500:1500, 2005.

[145] Ruth M Ripley, Adrian L Harris, and Lionel Tarassenko. Neural network models for breast cancer prognosis. *Neural Computing & Applications*, 7(4):367–375, 1998.

[146] Ralph C Smith. *Uncertainty quantification: theory, implementation, and applications*, volume 12. SIAM, 2013.

[147] Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. An introduction to mcmc for machine learning. *Machine Learning*, 50(1-2):5–43, 2003.

[148] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[149] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. the MIT Press, 2006.

[150] Zoubin Ghahramani. Probabilistic machine learning and artificial intelligence. *Nature*, 521(7553):452–459, 2015.

[151] Peter Congdon. *Applied bayesian modelling*, volume 595. John Wiley & Sons, 2014.

[152] Jean-Michel Marin, Pierre Pudlo, Christian P Robert, and Robin J Ryder. Approximate bayesian computational methods. *Statistics and Computing*, 22(6):1167–1180, 2012.

[153] Radford M Neal et al. Mcmc using hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2:113–162, 2011.

[154] Stephen P Jenkins. Survival analysis. *Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK*, 2005.

[155] Ruth M Ripley, Adrian L Harris, and Lionel Tarassenko. Non-linear survival analysis using neural networks. *Statistics in Medicine*, 23(5):825–842, 2004.

[156] Paul C Lambert and Patrick Royston. Further development of flexible parametric models for survival analysis. *Stata Journal*, 9(2):265, 2009.

[157] Patrick Royston and Mahesh KB Parmar. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*, 21(15):2175–2197, 2002.

[158] Frank E Harrell, Kerry L Lee, and Daniel B Mark. Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 15:361–387, 1996.

[159] Vanya Van Belle, Patrick Neven, Vernon Harvey, Sabine Van Huffel, Johan AK Suykens, and Stephen Boyd. Risk group detection and survival function estimation for interval coded survival methods. *Neurocomputing*, 112:200–210, 2013.

[160] Jonas Kalderstam, Patrik Edén, Pär-Ola Bendahl, Carina Strand, Mårten Fernö, and Mattias Ohlsson. Training artificial neural networks directly on the concordance index for censored data using genetic algorithms. *Artificial Intelligence in Medicine*, 58(2):125–132, 2013.

[161] John D Kalbfleisch and Ross L Prentice. *The statistical analysis of failure time data*, volume 360. John Wiley & Sons, 2011.

[162] David E Goldberg. *Genetic algorithms*. Pearson Education India, 2006.

[163] M Mahfouf, M Jamei, and DA Linkens. Optimal design of alloy steels using multiobjective genetic algorithms. *Materials and Manufacturing Processes*, 20(3):553–567, 2005.

[164] Hein Putter, M Fiocco, and RB Geskus. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*, 26(11):2389–2430, 2007.

[165] Daniel Bernoulli. Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir. *Histoire de l'Acad. Roy. Sci.(Paris) avec Mém. des Math. et Phys. and Mém*, pages 1–45, 1760.

[166] Philip Hougaard. Multi-state models: a review. *Lifetime Data Aanalysis*, 5(3):239–264, 1999.