**Article:**
Obajemu, O., Mahfouf, M. and Catto, J.W.F. (2018) A new fuzzy modeling framework for integrated risk prognosis and therapy of bladder cancer patients. IEEE Transactions on Fuzzy Systems, 26 (3). pp. 1565-1577. ISSN 1063-6706

https://doi.org/10.1109/TFUZZ.2017.2735939

# A New Fuzzy Modelling Framework for Integrated Risk Prognosis and Therapy of Bladder Cancer Patients

Olusayo Obajemu, Mahdi Mahfouf and James W. F. Catto

**Abstract**

This paper presents a new fuzzy modelling approach for analysing censored survival data and finding risk groups of patients diagnosed with bladder cancer. The proposed framework involves a new procedure for integrating the frameworks of interval type-2 fuzzy logic and Cox modelling intrinsically. The output of this synergistic framework is a score/prognostics index which is indicative of the patient's level of mortality risk. A threshold value is selected whereby patients with risk scores that are greater than this threshold are classed as high risk patients and vice versa. Unlike in the case of black-box type modelling approaches, the paper shows that interpretability and transparency are maintained using the proposed fuzzy modelling framework.

Two data sets are used to test the modelling accuracy of the elicited models. The first is an artificial data set which has similar characteristics as in a typical survival data. The second relates to real-life bladder cancer data from which one requires a model that identifies the low risk and high risk patients and then recommends risk management decisions based on, predicted risk level, patient history and characteristics, disease pathology and event times. The performance of the proposed framework is compared with the traditional Cox model, logistic regression as well as a non-linear survival data modelling technique based on neural networks.

This is the first time an attempt has been made to exploit the transparency advantages of fuzzy models and the principled statistical framework of the Cox model in order to identify risk groups and recommend risk management decisions from complex survival data sets. In both the artificial data and real data, the proposed modelling framework, although minimalistic, shows better generalisation performances than the previously reported models against which the results were compared.

O. Obajemu and M. Mahfouf are with the Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield S1 3JD, U.K. (emails: o.obajemu@sheffield.ac.uk, m.mahfouf@sheffield.ac.uk)

J. W. F. Catto is with the Department of Oncology and Metabolism, The Medical School, The University of Sheffield, Sheffield, S10 2RX, U.K. (email: j.catto@sheffield.ac.uk)

**Index Terms**

survival analysis, fuzzy modelling, flexibility, risk, bladder cancer.

## I. Introduction

WHEN a patient is diagnosed with bladder cancer (*BCa*), clinicians often require an estimate of the risk of death/relapse from the disease which may assist in answering pertinent questions such as: is this patient of a low or a high risk? What is the expected survival time if no treatment is administered? What treatment intervention could be performed to minimise this risk? Such estimates may be quantitative, possibly obtained from data modelling methods, or qualitative often obtained from knowledge experts. Obtaining quantitative estimates often requires collating and analysing the time of events (such as duration of time from being diagnosed with *BCa* to dying from *BCa* [*BCa* being the event]). More often than not, however, obtaining such estimates poses a great challenge because of the problem of censored[1] observations. Analysing censored time-to-event data is the mainstay of a branch of statistics called survival analysis of which the Cox proportional hazards model [1] (henceforth called the Cox model) is by far the most popular technique. The wide adoption of the Cox model is due to the fact that it is easy to use, it is highly interpretable and is also supported by a plethora of software packages which facilitates easy estimation and interpretation [2]. Additionally, its analysis can be carried-out without the need to specify the baseline hazard function by using the partial likelihood method thereby keeping assumptions to the minimum.

In the Cox technique, the hazard function (see Section II-B), which provides estimates of the mortality risk of an individual, is modelled directly and is assumed to be a product of a baseline hazard (a strict function of time) and a linear function of the individual's parameters. Consequently, the Cox model assumes that the effects of the input variables are obtained through a linear combination of variables and parameters ($\boldsymbol{\alpha}^\intercal \mathbf{x}$) usually called a *link function*, the exponent of which is usually called the *relative hazard* [3]. However, for many applications, the assumption of linearly related covariates may not always be tenable. For this reason, several other techniques have been proposed such as those using neural network-based models [4] [5], splines [6], Gaussian processes [7] in order to take advantage of the statistically principled Cox modelling framework and the high interpolation abilities of these non-linear models. These approaches have resulted in more accurate models as discussed in [4], however, their usage in practice

[1]Only right censored observations are investigated in this study. Right censoring is a special type of missing data problem. This occurs when the response variable (time of dying from *BCa*, for example.) is known to be greater than a particular value but is not observed exactly.

has been limited due to the loss of interpretability/transparency[2]. In this paper, it is proposed to employ a fuzzy model in this link function of the Cox model for the analysis of survival data. This integrated approach will be shown to allow for concomitantly meeting the two objectives of 1. eliciting accurate flexible models, 2. allowing for models to maintain transparency which is natural in fuzzy models [8] in the analysis of complex survival data.

Fuzzy logic was first introduced in 1965 in Zadeh's seminal paper [9]. It has found wide applicability ranging from control [10], signal processing [11] and modelling [12], [13]. Fuzzy modelling has been shown to be able to provide an intuitive representation of complex systems via if-then rules to which humans can relate and which makes them interpretable [14]. Approaches such as those proposed in [15] have only attempted to extrinsically combine fuzzy methods and the Cox model. Consequently, the fuzzy modelling stage is no different from the existing fuzzy frameworks which are inadequate for handling the censoring problem (which is discussed in Section II-B). It is surprising that up until now, the prospect of intrinsically integrating fuzzy logic with statistically principled survival data modelling methods (such as the Cox) to solve the conundrum of competing objectives of flexibility and interpretability has not been investigated in the open literature. It is presumed that this may be due to two reasons: the first is the belief that fuzzy logic and probabilistic reasoning are conflicting and cannot be reconciled. However, as discussed in [16], fuzzy logic differs from probability as both handle different types of imprecision and both can in fact work complementarily to provide a synergistic powerful framework. Therefore, there exists, we believe, ample room to exploit the advantages of fuzzy and statistical modelling. The second reason may be that the fuzzy community may not yet be aware that some fuzzy modelling problems may be posed in a survival analysis framework to improve modelling performance. Therefore, the purpose of this paper is of two fold: the first is to show how using fuzzy models can improve survival data models in terms of both accuracy and interpretability, and the second to prove emphatically the fact that fuzzy logic and statistical modelling are indeed complementary by successfully applying fuzzy models in the context of a survival data modelling problem such as predicting risk groups in patients diagnosed with *BCa*. Consequently, the overarching original contributions of this new study include:

- the development of a framework where type-2 fuzzy modelling is for the first time intrinsically integrated into the Cox model for risk management of patients diagnosed with *BCa*.
- the development of a newly proposed efficient algorithm for calculating the receiver operator char-

[2]they are typically called black-box models because it is difficult to interpret the parameters of the model

acteristics (*ROC*) when some of the data points are censored.

- the validation result to test the generalising properties of the proposed integrated model using our own real patient clinical data drawn from a cohort of 2918 patients.

Two data sets are used in this research. The first is an artificial non-linear data set similar to the one reported in [17] which is used to investigate the complexities of survival data. This data set allows one to understand the survival data generation process and for checking whether the proposed framework may correctly predict the time of death in the presence of censored observations. The second data set is a real-life data containing history of patients diagnosed with *BCa* collected at the Sheffield Royal Hallamshire Hospital (*RHH*), United Kingdom (UK) between 1995 and 2009. It is of interest in this data set to predict the risk group to which a new patient (with specific variables) diagnosed with *BCa* belongs. The approach hence adopted involves an implicit model in which the type of treatment administered to a patient (if any) is taken as an input to the modelling framework. One may then investigate how different therapies affect the predicted prognostics indices (the output of the fuzzy model) which may lead to risk management decisions. Hence, the resulting model is not only able to predict risk but should also be capable of providing recommendations for risk management. For simplicity, the binary risk classification is used in this research such that a patient may either have low or high risk of mortality although other risk grouping mechanisms are possible using the proposed framework.

The remainder of the paper is organised as follows: Section II provides, very briefly, the theories of survival analysis, fuzzy logic and fuzzy modelling. Section III details the proposed modelling framework and provides analysis of data used in the paper. Section IV provides a comparative study between the original Cox model, a neural network-based non-linear model which has been applied in the literature and the proposed fuzzy modelling framework. Section V concludes the paper by providing recommendations for practical model design and directions for future work.

## II. Fuzzy Logic and Survival Analysis

### A. Fuzzy Systems Modelling

This section briefly introduces fuzzy logic systems (*FLS*) and readers are referred to [18] and [19] for more in-depth analyses. A Fuzzy logic system (*FLS*) is capable of representing subjective knowledge [20]. A conventional *FLS* block diagram has four elements as shown in Fig. 1.

It can be seen that a *FLS* takes in an input and returns an output and consequently may be taken as a mapping from the input space $\mathcal{X}$ to an output space $\mathcal{Y}$. The fuzzifier block takes a crisp input and provides

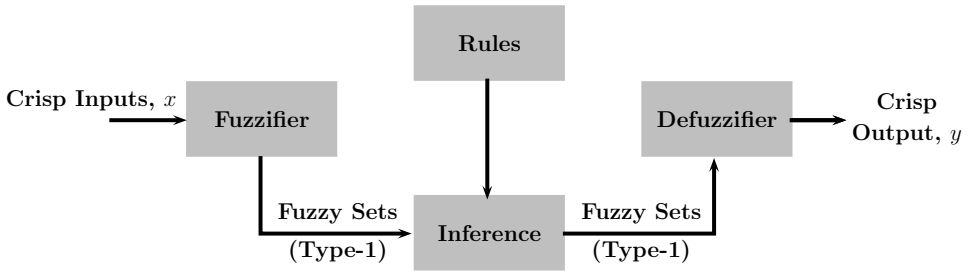



Figure 1: Block diagram of the Type-1 Fuzzy Logic System.

a fuzzy mapping of this crisp input. This is done to allow for scenarios of uncertain input measurements. The singleton fuzzification (as opposed to non-singleton) is used throughout in this research since the analysis of the *FLS* becomes simplified without any loss of generality [18]. The rules provide a linguistic representation of the system. The fuzzy inference engine (*FIS*) is the heart of the *FLS* and it is where the fuzzified inputs are combined with rules which results in a fuzzified output. Because crisp outputs are required to operate engineering systems, the output of the *FIS* is mapped into a crisp value through a process known as defuzzification (the defuzzifier block). Fuzzy systems interpretability stems from the rules block. Given a *FLS* with $n$ inputs ($\mathbf{x} \in \mathbb{R}^n$), $x_1 \in X_1, x_2 \in X_2, \cdots, x_n \in X_n$ , and one output $y \in Y$, with a rule-base consisting of $c$ rules, the $i$th rule of a typical IF-THEN fuzzy rule base may be expressed in the following form:

$$\begin{aligned} R^i \text{ :IF} \quad & x_1 \text{ is } \mathbf{A}_1^i \text{ and } x_2 \text{ is } \mathbf{A}_2^i \ \cdots \text{ and} \\ & x_n \text{ is } \mathbf{A}_n^i, \quad \text{THEN } y^i \ is \ \mathbf{B}_n^i \end{aligned} \tag{1}$$

where $\mathbf{A}_j^i$ represents the $j$th antecedent *MF* of the $i$th rule and $\mathbf{B}_n^i$ the consequent *MF*. This paper uses the Takagi-Sugeno-Kang (*TSK*) type of *FLS* such that $\mathbf{B}^i$ is represented by a linear function $g^i(x)$ of the inputs i.e. $g^i(x) = \boldsymbol{\beta}_i^T \mathbf{x}$ [18], where $\boldsymbol{\beta}_i$ is a vector of consequent parameters for $i = 1, 2, \cdots c$. Usually, these *MFs* are used to represent words such as high, low etc. The *MFs* are in turn represented by a fuzzy set (*FS*) which represents the subjective information that an individual has about the previous words.

This human-like representation of rules gives *FLSs* their subjectivity and also their interpretability. However, this subjectivity can also be a source of great conundrum. Because 'words mean different things to different people' [18], the definition of the $MFs$ of the *FSs* can vary amongst individuals/experts. For

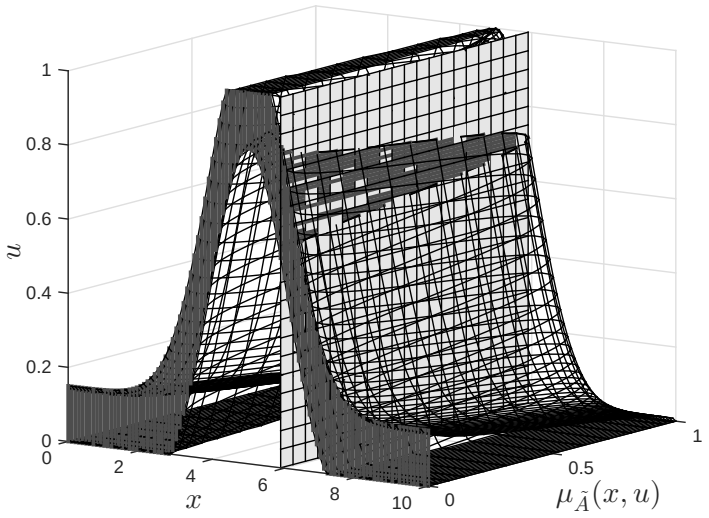



Figure 2: An example of a Type-2 Fuzzy Set ($\tilde{A}$) employing Gaussian primary *MF* $u_{\tilde{A}}(x)$ and triangular secondary *MF* ($\mu_{\tilde{A}}(u,x)$). The primary *MF* defines the range of the degree of membership that element $x$ belongs to the *FS*, $\tilde{A}$. The secondary *MF* defines to what extent the *PMF* values belong to $\tilde{A}$ as represented by the vertical slice in the figure.

this reason, type-2 fuzzy sets *T2 FS* (Fig. 2) were introduced more than four decades ago [21].

Using *T2 FSs* instead of the conventional *FS* (henceforth called type-1 fuzzy set - *T1 FS*) can address these *linguistic uncertainties*. Additionally and especially in modelling, using type-2 counterparts provides an extra degree of freedom which may, more often than not, improve the accuracy and the generalisation capabilities of models [11]. However, the disadvantage of using *T2 FS* is the inclusion of a new block in the *T2 FLS* block diagram (Fig. 3) which results from the need to type-reduce the *T2 FSs* into a *T1 FS* before any defuzzification operation is performed. This type reduction stage involves reducing the resulting *T2 FS* from the *FIS* output to a *T1 FS*. This process is discussed in detail in [22]. As shown in [18], this is a computational demanding stage because the type-reduction process involves enumerating the many (possibly infinite) embedded *T1 FS* in the *T2 FS* and then defuzzifying them in turn. There exist many ways to reduce computational burden of this process [23]. A popular approach consists of using the interval type-2 fuzzy sets (*IT2 FS*) (Fig. 4) in the *FLS* instead of the general *T2 FLS*. The Karnik-Mendel algorithms and its variants [22]–[24] provide a fast iterative mechanism for type reduction. This work makes use of the *IT2* approach since it represents a trade-off between exploiting the degree of freedom and the reasonable computational speed especially for high dimensional modelling problems. Additionally, the *IT2* approach, as discussed in [18], may also help in handling uncertainties in data modelling. However, it is worth noting that the proposed techniques are capable of extending to both the *T1* and the *GT2*

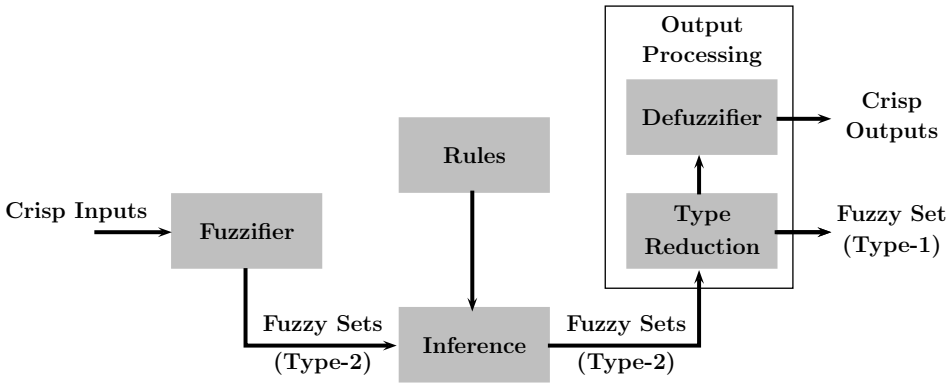



Figure 3: Block diagram of the Type-2 Fuzzy Logic System.

approaches.

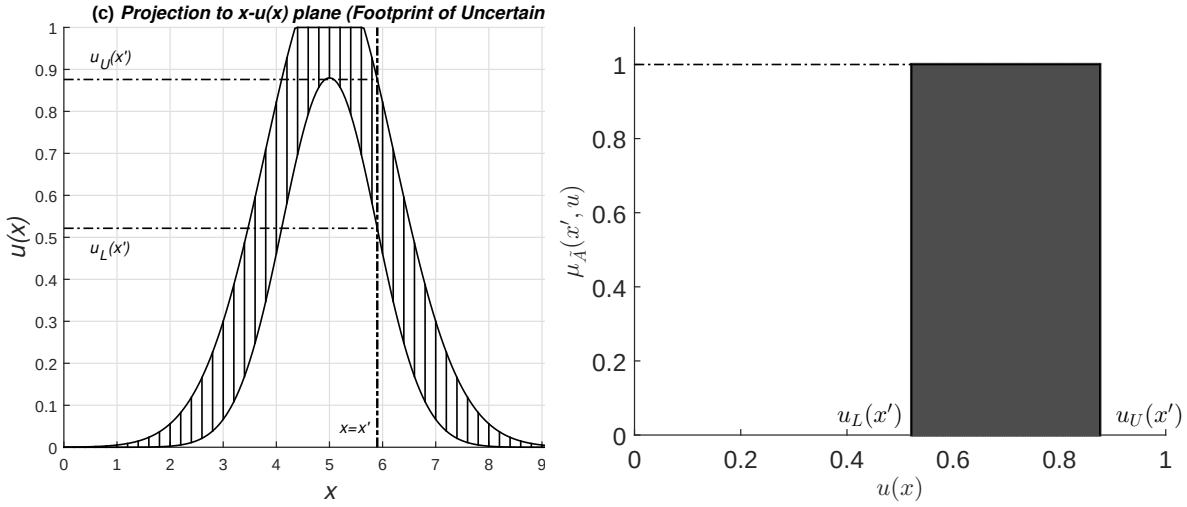
The general *FLS* is consequently a mapping from the input space $\mathcal{X}$ to the output space $\mathcal{Y}$, which may be represented as follows:

$$\hat{y} = \sum_{i}^{c} \phi_i(\mathbf{x})\zeta_i \tag{2}$$

where $\hat{y}$ is the estimated output returned by *FLS* after type-reduction, $\phi_i(\mathbf{x})$ represents the validity function for $i$th rule for a total of $c$ rules for input defined as $\mathbf{x}$. For the *TSK FLS* method employed in this research, $\zeta_i = \mathbf{a}_i^\intercal \mathbf{x}_i + b_i$. For the *IT2 FLS* used here, $\phi_i(\mathbf{x})$ has two components ($\overline{\phi}_i(\mathbf{x})$ and $\underline{\phi}_i(\mathbf{x})$) which represent the left and right end points of the validity functions as derived from the Karnik-Mendel algorithms. $\hat{y}$ is the average of the two calculated values for after plugging the values of $\overline{\phi}_i(\mathbf{x})$ and $\underline{\phi}_i(\mathbf{x})$) into (2). It is worth noting that the Gaussian primary *MF* with fixed mean and uncertain width as derived in [18] and [25] is employed in this research with the t-norm operator being taken as the product.

### *B. Survival Analysis*

A typical *BCa* database usually contains information relating to the patient characteristics (such as age), disease pathology (stage and grade of the disease), disease history (relapse and administered therapies) and time of death from *BCa* (if any). Clinicians often require information on how these varying characteristics affect the risk of an individual on being diagnosed with *BCa*. Naturally, high risk patients would have a lower survival time $T$ and vice versa. The major problem when analysing such databases is that not all patients in the database have died from *BCa*. For example, some patients are lost to follow-up, possibly due to moving to other locations and had never contacted the hospital again, and some have died from




(a) (b)

Figure 4: The Interval Type-2 Fuzzy Set (*IT2 FS*) is a simplification of the *T2 FS*. The primary *MFs* remain unchanged (as shown in the left panel) but the secondary membership values are all taken to be unity (as shown in the right panel). Thus, the *IT2 FS* is completely defined by the so-called footprint of uncertainty (*FOU*) [18].

causes completely unrelated to *BCa*. A natural and easy approach to analysing such survival data is to completely exclude these patients from the analysis. However, this approach may well result in the significant loss of useful information, especially in cases where the proportion of patients with incomplete information is significant. It is possible, however, that one is privy to the time of the last follow-up or time of death from other causes ($C$) so that $T$ is known to be greater than $C$. This problem is known as censoring and is illustrated in detail in Fig. 5.

Survival analysis provides an excellent framework for analysing such databases even in the presence of such incomplete information (censoring). Survival analysis relates to the analysis of time until an event occurs. Mathematically, $T$ is taken as a random variable which represents the time of death from *BCa* and survival analysis is concerned with identifying $\mathrm{P}(T > t)$ from data. Two broad frameworks for survival analysis (when covariates are involved) are the Cox models and accelerated failure time (*AFT*) models. The Cox model is due to [1] in which, as already mentioned, the so-called hazard is modelled directly. The hazard is defined as the instantaneous risk of an event (the event is death from *BCa* in this study) and is given in the Cox model [1] by the following equation:

$$h_j(t, \mathbf{x}_j) = h_o(t) \exp(\boldsymbol{\alpha}^\intercal \mathbf{x}_j) \tag{3}$$

where $\boldsymbol{\alpha}$ is a vector of parameters and $\mathbf{x}_j$ represents the values of the input variables (covariates) for the

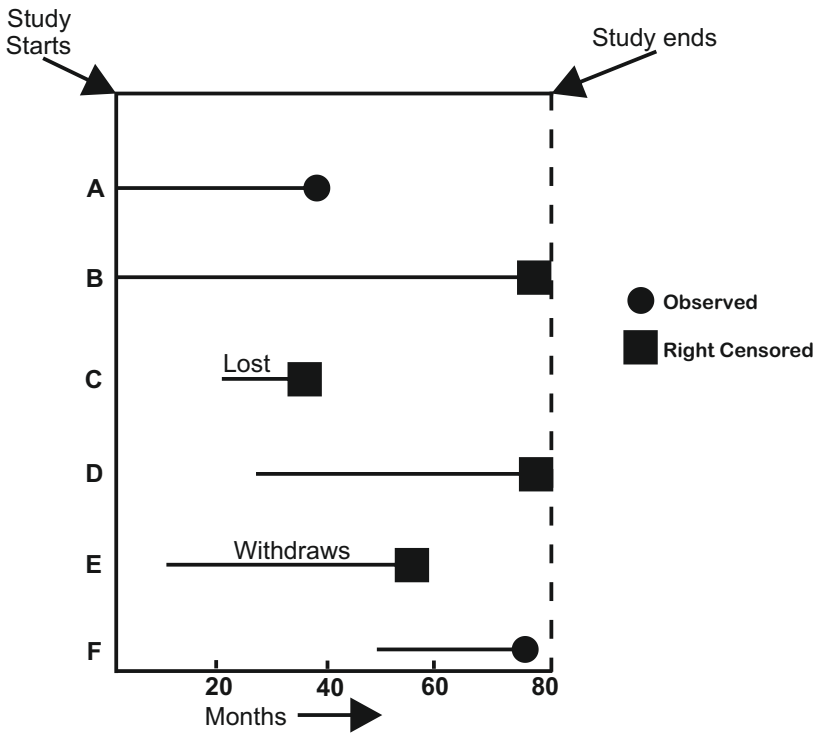



Figure 5: Illustration of right censoring. The event times of patients B and D are not known since the study ended before they could be observed. Only their censoring times are known. Patients A and F have their event times fully observed as they died from the event of interest at approximately 40 and 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 [3]. Patient E withdrew from the study at 60 months (censoring time is 50 months) while patient C was lost to follow-up at 40 months (censoring time is 20 months). The duration of the study is 80 months.

$j$th subject with $h_j(t, \mathbf{x}_j)$ representing the corresponding hazard for $j = 1, 2, \cdots N$. $N$ is the number of subjects under study.

The survival function ($S(t) = \mathrm{P}(T > t)$), is given as follows:

$$S(t) = S_0(t)^{\exp(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_j)} \tag{4}$$

where $S_0(t)$ is called the baseline survivor function and represents an hypothetical survivor function where all the covariates have zero values.

The relationship between the hazard function and the survival function for a specific covariate $\mathbf{x}_j$ is given as follows:

$$S(t) = \exp\left\{-\int_0^t h(t)dt\right\} \tag{5}$$

which easily shows, as a result of (3), that a high value of $\exp(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_j)$ for patient $j$ would result in a

narrower survival function (and consequently lower median time). Eq. 3 simply means that the hazard (or risk) of failure at any point in time $t$ for a subject $j$ with the covariates defined by the vector $\mathbf{x}_j$ is a product of two functions. The first function is time-dependent and is called a baseline hazard function $(h_o(t))$ 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 $(\exp(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_j))$. 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. This time-independent function completely specifies the hazard of an individual and can be estimated without the need to find the baseline hazard function. It is for this reason that the relative hazard, which is defined as the exponent of the link function $(\exp(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_j))$, represents useful information.

The *AFT* approach models the time of event directly by assuming the times and the covariates act multiplicatively on a time scale [3]. This approach is very similar to an ordinary linear regression. The *AFT* approach is not considered in this research since the assumption that the survival time follow a particular distribution must be made which is not usually justifiable in complex modelling problems such as eliciting risk models from a *BCa* database. Other types of survival data modelling interpretations exist e.g. proportional odds [26].

Using linear models in the link function of the Cox model has peculiar intuitive appeals, especially to clinicians. However, even though this linear assumption may not be tenable, there is a reluctance to using black-box non-linear models because of the loss of this intuitiveness and interpretability. It is argued in this research that the use of fuzzy systems will provide a better solution to helping one to understand survival data both accurately and in a more transparent way.

*Performance Indices in Survival Modelling:* 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*) and the *ROC* used in classification, may be challenging [27]. For this reason new performance indices have been defined in the literature [4] but the concordance-index (c-index), which is one of the considered performance indices in this paper, appears to have the best intuitive appeal. The concordance index makes use of the fact that a subject with a high risk or prognostics index ($p_k$), is likely to have a lower survival time than the one with a lower prognostics index ($p_l$). Hence, if $p_k > p_l, t_k < t_l$, then the subject pair $(p_k, p_l)$ is said to be in concordance. The c-index is useful when one is strictly not interested

in the times of failure but rather in grouping the observations into risk groups according to their risk ranks as is the case in many medical studies and in this paper. The concordance index is defined as follows:

$$\text{c-index} = \frac{1}{N_p} \sum_{k,l} I(p_k, p_l) \tag{6}$$

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

$$I(p_k, p_l) = \begin{cases} 1 & \text{if } p_k > p_l, \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

The interpretation of Eqs. (6) and (7) is that a higher risk/prognostics index indicates a tendency for a lower survival time. When the prognostics indices for the population have been found, existing methods of finding the *ROC* cannot be used because some observed times are censored. This is because 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, an efficient *ROC* algorithm based on that discussed in [28] is modified to make it suitable for eliciting an *ROC* for censored survival data.

The algorithm which is used to find the true positive ($TP$) rate $\left(\approx \frac{\text{positives correctly classified}}{\text{Total positives}}\right)$ and the false positive ($FP$) rate $\left(\approx \frac{\text{Negatives incorrectly classified}}{\text{Total negatives}}\right)$ is shown in Algorithm 1.

Additionally, the Breslow estimator [3] allows for calculating the baseline hazards and baseline survival functions which will consequently facilitate calculating the 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 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. This paper makes use of both the defined c-index in (6), the proposed modified *ROC* analysis and the *RMSE* of predicted median times and observed times for non-censored observations to compare the results of the proposed fuzzy modelling framework, with those of the Cox-model and the neural networks model.

Several methods exist to define the risk groups of subjects from survival times, for example see [29].

**Algorithm 1** Efficient method for generating the *ROC* curve in the presence of censored observations.
**Inputs:** Calculated prognostics indices $p(k)$, observed times $T(k)$, censoring indicator $\Delta(k)$, and threshold $(TH)$. For $k = 1 \cdots N$, $N$ being the number of subjects.
**Outputs:** The *ROC* points (stack) with increasing $FP$ rate.

$P_{sorted} \leftarrow P$, sorted by decreasing $p$ scores
$FP \leftarrow TP \leftarrow 0$
$k \leftarrow 1$
$Ne \leftarrow |T > Th|$
$Po \leftarrow |T < Th \quad \& \quad \Delta = 1|$
**while** $k \leqslant |P_{sorted}|$ **do**
  **if** $p(k) \neq p_{prev}$ **then**
    push $\left(\frac{FP}{Ne}, \frac{TP}{Po}\right)$ into $ROC$
    $p_{prev} \leftarrow p(k)$
  **end if**
  **if** $P_{sorted}[k]$ is a high risk example i.e. $T[k] < Th \quad \& \quad \Delta[k] = 1$ **then**
    $TP \leftarrow TP + 1$
  **else if** $T[k] < Th \quad \& \quad \Delta[k] = 0$ **then**
    $TP \leftarrow TP$
  **else**
    $FP \leftarrow FP + 1$
  **end if**
  $k \leftarrow k + 1$
**end while**

Here, we take a simple approach usually common among clinicians. A time frame is chosen (typically 5 years in cancer studies), and subjects with failure time greater than this threshold are assumed to be low risk individuals and vice versa. Only censored observations that have survival times greater than this threshold may be included in the study. No conclusion can be drawn for individuals whose censored times were less than this time threshold, as shown in Algorithm 1 when calculating the *ROC* points. For more details on how to calculate the *AUC*, readers are referred to [28].

## III. Modelling Framework and Data Generation Mechanism

### *A. Modelling Framework*

As already mentioned, it is proposed in this study to **integrate** for the first time the traditional Cox model with fuzzy systems modelling such that the linear part (link function) of Cox model is replaced with a flexible fuzzy model. The elicited model, as will be shown, is able to infer risk groups (low risk or high risk) using a *BCa* database in a non-linear manner while maintaining the transparency that is inherent in fuzzy systems modelling. From the relationships in (3) and (5), a high value for the hazard (prognostics index) means that the patient has a lower median survival time and would tend to die sooner

than a patient with lower prognostics index. It thus intuitive and natural to consider patients with higher prognostics indices as high risk patients and vice versa. Using the hazard/prognostics index for this risk subdivision is thus a natural and intuitive approach[3]. From (3), the link function defined as $\lambda$ is the prognostics index $p$ and is taken to be the exponent of the output of the fuzzy logic model so that for subject $j$, the prognostics index is defined by the following equation:

$$\lambda_j = \exp(f^j_{FLS}(\mathbf{x}_j, \boldsymbol{\alpha})) \tag{8}$$

so that

$$h_j = h_o(t)\lambda_j = h_o(t)\exp(f^j_{FLS}(\mathbf{x}_i, \boldsymbol{\alpha})) \tag{9}$$

where $f^j_{FLS}(\mathbf{x}_j, \boldsymbol{\alpha})$ is the output of the *FLS* defined in (2) with parameters $\boldsymbol{\alpha}$ for individual with covariate values $\mathbf{x}_j$. $\lambda_j$ is the prognostics index of the individual. The reason for taking the exponent is to ensure that the hazard function is positive and defined. It is worth noting that (8) and (9) are similar to the traditional Cox model (see (3)), the only difference being that the linear part has been replaced by a *FLS*. Estimating the parameters of the *FLS* consequently includes similar steps to those followed in the case of the Cox model, thereby exploiting its mathematical convenience and simplicity. It can easily be shown that since $\lambda_j$ is mapped from the output space of the fuzzy model using a monotonic function $(\exp)$, then according to Zadeh's extension principle [18], the *MF* is maintained as a result of the one-to-one mapping caused by the monotonic transformation. Consequently, this will ensure that the transparency inherent in fuzzy models 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 $B$:

$$\mu_B(y) = \max_{y=f(x)} \mu_A(x) \tag{10}$$

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

Therefore, it can easily be seen that interpretability is not compromised as long the function is monotonic

[3] $S_i(t) = S_o(t)^{\lambda}$. $\lambda$ 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 $\lambda$ i.e. higher risk/hazard, lower survival time.

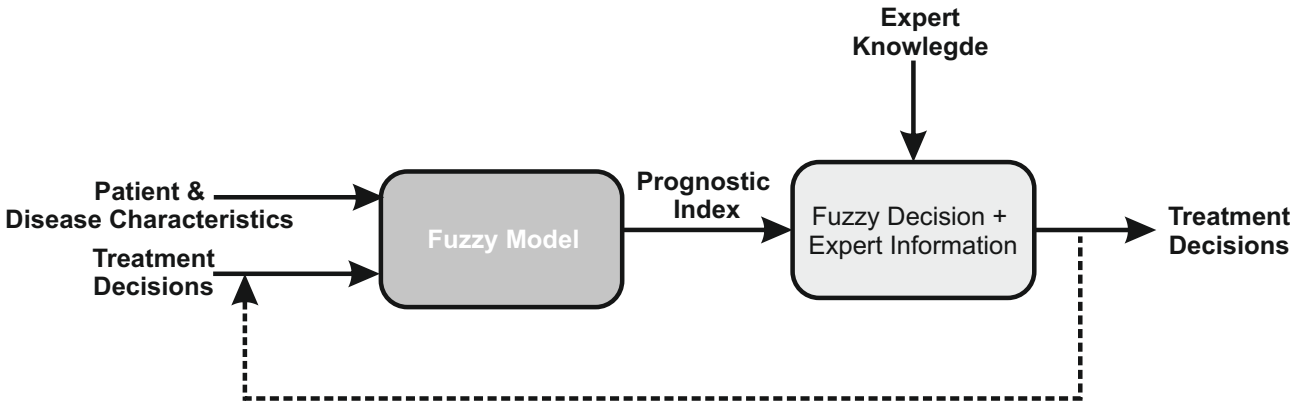



Figure 6: Schematic diagram of the proposed modelling framework. The first stage consists of building an interpretable model that classifies a patient as a low or high risk. The elicited models have implicitly added the treatment decisions (if any) which can help proffer the type of treatment in the second stage.

(an exponential in this study). The optimisation stage of the fuzzy model elicitation operation may often lead to overlapping and indistinguishable *MFs* thereby degrading the intrepretability of the resulting fuzzy model. Interested readers may refer to [30] and [31] for more techniques on interpretability enhancement. Fig. 6 shows the schematic diagram of the proposed modelling framework. The output of the fuzzy model, as discussed, is a prognostics index that is indicative of the degree of risk a patient has. The prognostics index is the output of the fuzzy model which it being transparent may easily be integrated with expert knowledge. For example, a clinician may know from experience if a patient is of a low or a high risk of mortality. Fuzzy modelling provides a natural framework for this clinician to investigate this assertion.

*1) Training the Model:* In survival data modelling with the right-censored observations (see Fig. 5), 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 the observed times $T_j$ is the censored time). Estimation of parameters is usually carried-out by maximizing an objective function such as the c-index [17], area under the survival curve [4] and partial likelihood [32].

The fuzzy model is trained according to the partial likelihood methodology given in (11) which is based on the same premise of the original 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 prognostics index) than the 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:

$$L(\boldsymbol{\alpha}) = \prod_{j=1}^{R} \frac{\exp(f(\mathbf{x}_j, \boldsymbol{\alpha}))}{\sum_{m=1}^{n_j} \exp(f(\mathbf{x}_m, \boldsymbol{\alpha}))} \tag{11}$$

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 $t_j$ and this number includes the censored individuals that have not failed at this time. The parameters of the *FLS* are found by maximising (11). In this research, a Genetic Algorithm (*GA*) is used to perform the optimisation of this function. The negative log-likelihood (*NLL*) of (11) 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:

$$NLL(\boldsymbol{\alpha}) = \sum_{j=1}^{R} f(\mathbf{x}_j, \boldsymbol{\alpha}) - \sum_{j=1}^{R} \log \left\{ \sum_{m=1}^{n_j} f(\mathbf{x}_m, \boldsymbol{\alpha}) \right\} \tag{12}$$

Details of the derivations of partial likelihood formula for the Cox models from data likelihood may be found in [32] and [3].

*2) Optimisation and Validation Details:* *GA* is an evolutionary algorithm which simulates natural survival of the fittest. This is the optimisation method of choice in this work because it is a tested and trusted in many applications and has the capability of returning global optima solutions [33].

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 parts. About $2/3$ for training and remaining the $1/3$ for testing and the $k$-fold cross validation was performed on the training data sets to select the model with the best generalisation ability $k = 5$. The maximum allowable number of rules was set at $20$ to manage the computational time. In the artificial data set, the fuzzy model with $17$ rules was found to lead to the best results 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 the generalisation ability on a data set that was not used in the training procedure.

### *B. Data*

As already stated and especially in real systems, the assumption that the hazards are affected by an exponentiated linear function of the covariates as given by (3) may not be tenable. It has already been shown in the last section how interpretability of a fuzzy model output is maintained and that this output is indicative of the risk a patient faces as far as the onset of a disease is concerned. In this section, it is shown

that using *FLS* may be an ideal framework candidate for non-linear behaviour and when interpretability is important. The proposed modelling framework is now tested on two datasets, one artificial data and a real *BCa* data, and the test results are compared with standard modelling methods.

The artificial data generation mechanism is similar to the one used in [17] but has been modified to make it more non-linear. The real data is a real life data on *BCa* patients. This paper investigates the ability of using fuzzy survival modelling results to identity risk groups in the data and thereafter providing recommendations for therapy. These two data sets are examined in greater details in the next subsections.

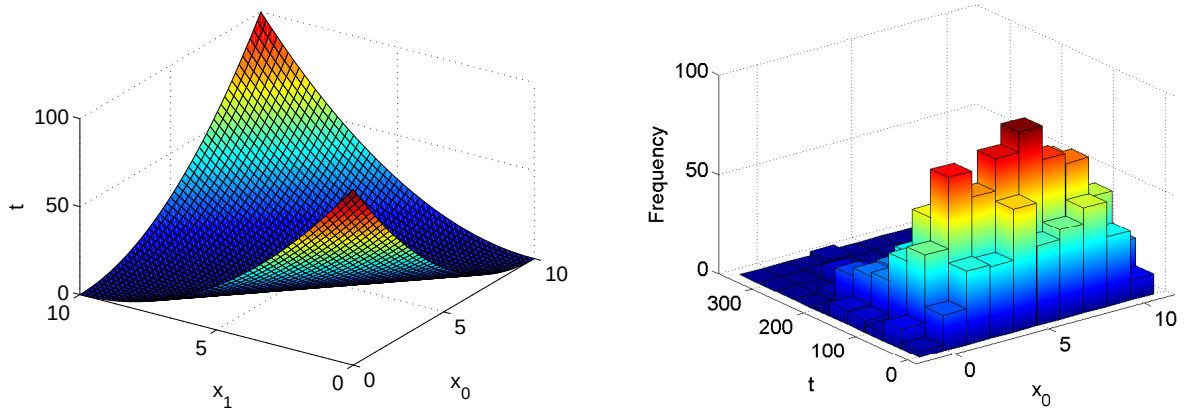
*1) Artificial Data:* The purpose of this artificial data generation is to demonstrate the case where a non-linear modelling technique is needed to generate good modelling results in a noisy data set. The artificial data set is generated according to the following equation:

$$\begin{aligned} t = (x_0 + \sqrt{x_1} + x_2^2 - 15)^2 + (x_3 + x_4 - 10)^2 + \\ x_5 + (x_6^2 + \sqrt{x_7} - 15)^2 + (x_8 + x_9 - 10)^2 + \epsilon \end{aligned} \tag{13}$$

The $x$'s are the covariates drawn from uniform continuous distributions so that they have values between $0$ and $10$. $\epsilon$ is a random noise added to the times and $p(\epsilon) \sim \mathcal{N}(0, \rho)$ where $\rho = 5.0$. Independent censoring was achieved by randomly choosing a proportion of the data set and randomly drawing a uniform number from between 0 to the event times of the chosen data points.

The advantage of testing the framework on the artificial data is that one is able to compare the predicted median times against what the failure time would have been had the events not been censored which may provide an excellent indication on how the models perform. It is perhaps worth noting that in real datasets, one would not be privy to such information. $2000$ data points were generated MATLAB® 2015A software for training and $1000$ was used for testing.

The model is trained following the procedures already discussed in Section III-A. Investigations on how the predicted median times (a function of the prognostics index and baseline hazard) compare with what '*would have been the event time*', had they not been censored, are carried-out using the *RMSE* as the performance index. As already stated, the k-fold cross validation was used to select optimal number of fuzzy rules which was found to be 17 for a maximum number of 20 rules to manage computational time. Fig. 7a 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 a constant value (mean value) to show the non-linear distributions of the data. Fig. 7b shows the data distribution of the times and one




(a) Plot of two input variables for the artificial data for two input variables only ($x_0$ and $x_1$) to display the non-linearity. The other variables were set equal to their mean values.

(b) Histogram of simulated times and one input variable ($x_0$). The times in this case are the real times (with noise added) and not the censored times.

Figure 7: Artificial data characteristics.

input variable ($x_o$).

*2) Bladder Cancer Data:* The *BCa* data were obtained from a study of *BCa* patients at the Royal Hallamshire Hospital in Sheffield, United Kingdom (*UK*)[4]. To understand the long term outcomes of patients diagnosed with *BCa*, the hospital created a database of all patients diagnosed with the disease for the 16-year period between 1 January, 1994 and 31 December, 2009. Tumours were given histological grading and staging following the World Health Organisation (*WHO*) convention of 1973 as well as the Tumour-Node-Metastasis (*TNM*) classification. A detailed statistical-based study was previously conducted by the third author of this paper (Professor W. F. Catto) as published in his earlier work of [34], whereby the Kaplan-Meier method was applied in order to identify the most significant features via the Kruskal-Wallis and $\chi^2$ tests. Patients diagnosed with the disease, but who are still alive after the study period, were automatically censored. These expert and statistical-based analyses resulted in identifying thirteen (13) explanatory variables (input variables) for each patient which included details of disease pathology, patient specific characteristics and treatment interventions (if any). Details of each variable are shown in Table I. There were $3634$ patients with primary *BCa* but those with insufficient follow up ($< 6$ months) were excluded from the analysis. Additionally, patients with missing covariates were excluded from the analysis leaving 2918 patients in the database.

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

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

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

| **Continuous Variables** | **Median** | **Mean** | **Range** |
|---|---|---|---|
| Age (years) | 72.7 | 71.6 | 21.3 - 101.0 |
| Stage | 4.03 | 4.02 | 0.00 - 9.00 |
| Urothelium | 2.00 | 3.42 | 0.00 - 6.00 |
| Nodes Detail | 4.00 | 3.94 | 0.00 - 4.00 |
| **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% |
| CIS Present | No | 2548 | 87.32% |
| | Yes | 370 | 12.68% |
| SPB | 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% |

study, loss to follow up or death due to other causes. Table II includes typical non-linear and stochastic behaviours in survival data. This table includes 4 male patients having identical values of input variables showing widely different event times. In particular, patient B is the oldest of the four patients but lived longer than all but patient A. Patient A lived for more than 70 months after being diagnosed with *BCa* but patient C with about the same age as patient A lived less than a month even though exactly the same values were recorded for the other input variables. This is evident of the non-linearity and noise embedded in a typical survival data. It would be interesting to see how the newly proposed framework handles these challenges.

The distribution of the times is shown in Fig. 8. There is a higher proportion of censored observations

Table II: Evidence of noise and non-linearity in the *BCa* data. The patients are represented by A, B, C and D.

| **Variables** | **A** | **B** | **C** | **D** |
|---|---|---|---|---|
| Age | 63.6 | 70.2 | 62.4 | 69.6 |
| Stage | 2.00 | 2.00 | 2.00 | 2.00 |
| Urothelium | 2.00 | 2.00 | 2.00 | 2.0 |
| Nodes Detail | 4.00 | 4.00 | 4.00 | 4.00 |
| Sex | Male | Male | Male | Male |
| Tumour Grade | Poor | Poor | Poor | Poor |
| Squamous | No | No | No | No |
| CIS Present | No | No | No | No |
| SPB | Solid | Solid | Solid | Solid |
| Vascular Invasion | No | No | No | No |
| Muscle Invasion | Yes | Yes | Yes | Yes |
| Cystectomy | No | No | No | No |
| Radiotherapy | No | No | No | No |
| **Event Time** | **72.72** | **36.16** | **0.62** | **2.56** |

at longer follow-up times than at lower follow-up times. This is typical of survival analysis and is due to the study having a fixed duration which means that patients with highest the follow up times tend to be censored. It is worth re-emphasizing at this stage that treatment decisions (cystectomy[5] and/or radiotherapy) were implicitly modelled by including them as part of the input variables. This can provide information on how these treatments dynamically affect the risk prognostics index and how they can also help in providing clinicians with recommendations for therapy as already discussed.

## IV. Results & Discussions

### *A. Artificial Data*

Fig. 9 shows a comparison of the *ROC* performances using the newly proposed fuzzy modelling framework and other standard modelling methods on the test data of the artificial data set. With an *AUC* of $0.53$, the Cox model is just marginally better than a random guess of the risk groups of patients. This is because a Cox model has been used on a highly non-linear and noisy data set. The same conclusion

[5]Cystectomy is the surgical removal of bladder and it is usually administered in advanced stages of the bladder cancer, usually after radiotherapy has been administered.

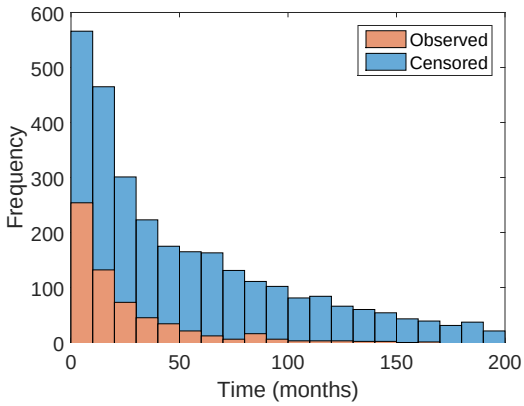


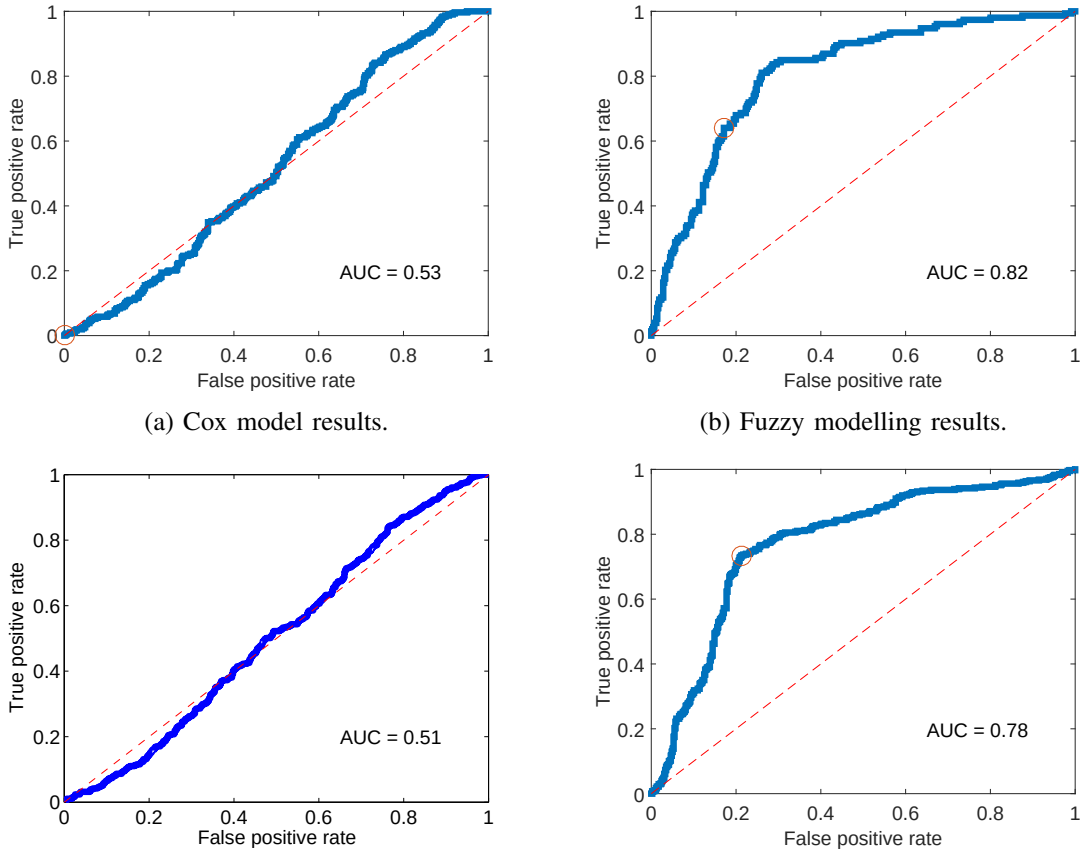

Figure 8: Histogram of the times of events for both censored and uncensored observations for the bladder cancer data set. Patients with higher follow-up times tend to be the most censored as is typical in survival data studies. This is is because the study was for a limited amount of time and patients with high event times become automatically censored at the conclusion of the test.

can be drawn when the logistic regression is used to analyse the data with censored times taken as event times which has an *AUC* of $0.51$.

The fuzzy model has a *AUC* of $0.82$ which represents approximately a 50% improvement in the modelling accuracy as compared with the Cox and logistic regression models. Compared with a non-linear model such as the one based on a neural network earlier proposed in [17], the proposed fuzzy model provides an improvement (of approximately 5%) in performance.

The concordance index shown in Table III gives similar results. The fuzzy model expectedly (being a non-linear model) outperforms the Cox and logistic regression models on the artificial data. It also shows a better generalisation capability than the neural network model result which was obtained using the test data.

A closer inspection of the performance of the proposed fuzzy modelling framework on the artificial data shown in Fig. 10 reveals that it is able to both infer correctly the risk groups as well as the predicted median survival times even in the presence of noise. It can also be observed that the model (Fig. 10D) is capable of handling the fact that certain 'times' were not observed exactly (censored) and is able to infer correctly what would have been the observed survival times had they not been censored. The proposed modelling framework has also been able to handle this missing data (information) problem. In contrast, as shown in Fig. 11, the Cox model performs very badly on this data set and is not able to infer the risk groups correctly as well as to predict the median times.




(a) Cox model results.

(b) Fuzzy modelling results.

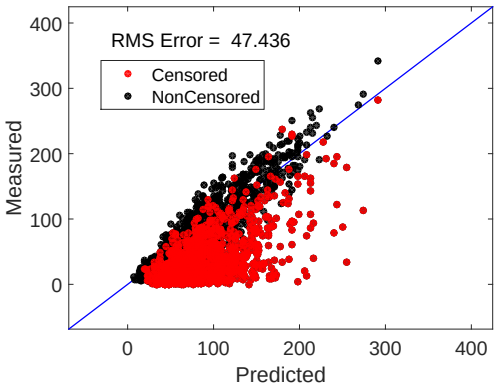
(c) Logistic Regression results.

(d) Neural networks results.

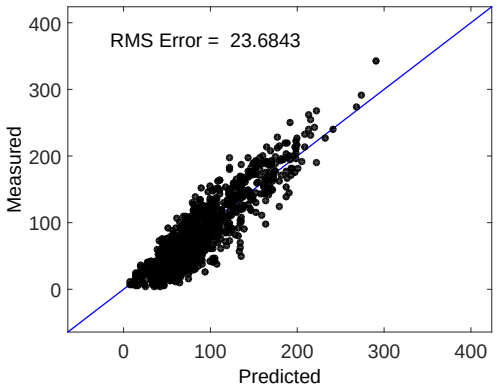
Figure 9: Comparison of *ROC* performances on the artificial test data set. Results show that the Linear model is not better than random guessing of the risk groups while using fuzzy modelling gives the best results. Positive means high risk (low survival time) and negative means low risk (high survival time). The red dotted line is the $y = x$ line. *ROC* curves that coincide with this line are no better than random guessing.

Table III: Concordance index values on the Training and Testing data sets using Cox, Logistic Regression (*LoR*), Neural Network (*NN*) and Fuzzy modelling (Type-1 and Type-2) on the artificial data and *BCa* data. The Type-1 fuzzy model was elicited following the same procedure described for the *IT2* fuzzy model.

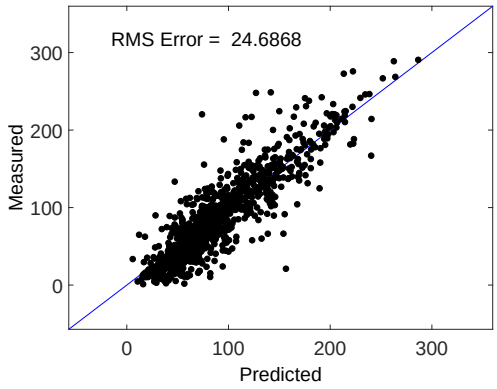
| | | **Training** | **Testing** |
|---|---|---|---|
| **Artificial Data** | *Cox* | 0.54 | 0.53 |
| | *LoR* | 0.53 | 0.51 |
| | *NN* | 0.72 | 0.68 |
| | *T1 FS* | 0.72 | 0.65 |
| | *IT2 FLS* | 0.73 | 0.72 |
| ***BCa* Data** | *Cox* | 0.71 | 0.70 |
| | *LoR* | 0.68 | 0.67 |
| | *NN* | 0.83 | 0.77 |
| | *T1 FS* | 0.80 | 0.78 |
| | *IT2 FLS* | 0.82 | 0.81 |



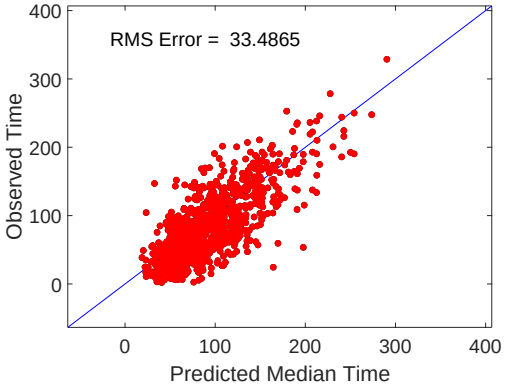

(a) Training data (censored and uncensored). Censored observations are indicated by red data points. It can be seen that the fuzzy model tries to predict what the failure time would be if the observation had not been censored.




(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 (uncensored only). 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.

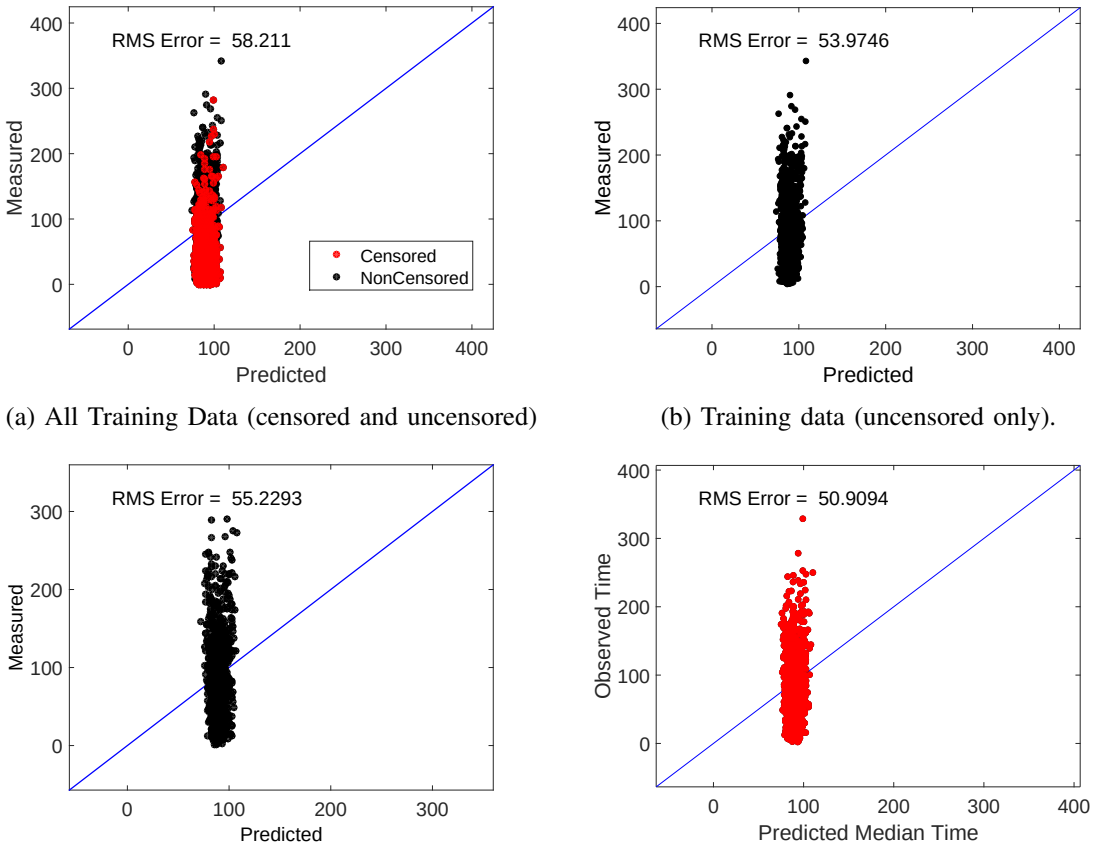



(d) Predicted median times vs. would have been failure time of censored observations. The fuzzy model does a good job in anticipating what the failure time would have been had the observations not been censored. It should be noted that the "would have been failure time" is not observed in real life data.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

Figure 10: Fuzzy model prediction results on the artificial data set. The fuzzy model outperforms the traditional Cox model both in terms of prediction performances as well as not allowing censored observations to bias the elicited models.

### *B. Bladder Cancer Data*

Figs. 12 and 13 show the *ROC* curves on the testing data set of the *BCa* data using the proposed fuzzy model and standard methods. The *AUC* for the Cox model is $0.83$ while that for the fuzzy model is $0.91$ which represents a significant improvement on the Cox model. The *AUCs* using the logistic regression and neural network based models were $0.76$ and $0.88$ respectively which shows that the proposed fuzzy modelling method led to the best generalisation performance. The *ROC* of the elicited model, as compared




(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 11: The Cox model prediction results on the artificial data set.

with other modelling techniques, are shown in Table IV. The concordance index, as shown in Table III, reveals that the proposed fuzzy modelling framework provides approximately 13%, 16% and 5% improvement in performance as compared with the Cox model, logistic regression and neural network based models respectively based on the testing data performance. On the *BCa* training data, only the neural network-based model marginal improvement in performance accuracy. The interval type-2 fuzzy model elicited in this study outperformed the type-1 counterpart in all performance measures. This is not surprising since type-2 fuzzy modelling includes one more degree of freedom and can better cope with uncertainties than its type-1 counterpart. The log-prognostics indices for patients A, B, C and D shown in Table II were found to be $-5.21$ (low risk), $0.2$ (high risk), $2.18$ (high risk) and $2.01$ (high risk) which is in line with the event times, demonstrating that the proposed model is able to discriminate the noise and censored event times to elicit a more accurate and more reliable risk model.

*1) Therapy Recommendation:* The *ROC* curves of Figs. 12 and 13 show the performance of the classifiers at different selected thresholds for the prognostics index/risk score. In practice, however, only

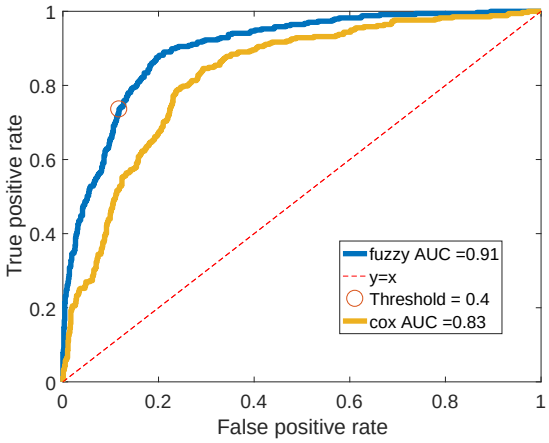
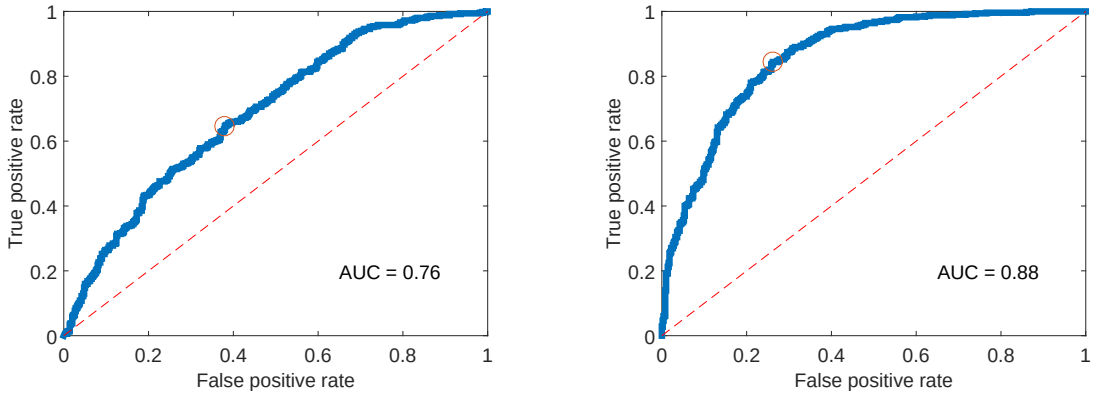



Figure 12: Comparison of *ROC* performances on the *BCa* test data set using Cox and fuzzy-based models. Point marked o is the optimum point based on the isocost.




(a) Logistic regression modelling performance on the *BCa* data.

(b) Neural network based model performance on the *BCa* data

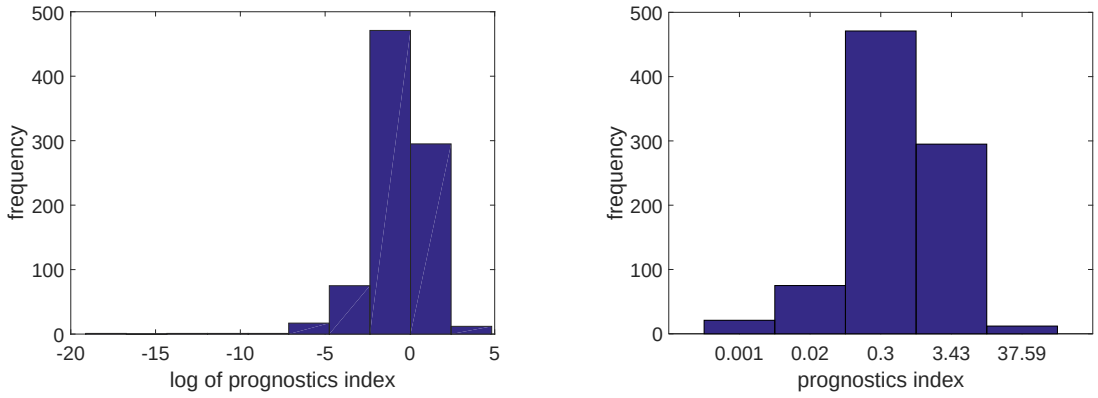
Figure 13: Comparison of *ROC* performances on the *BCa* test data set using the neural network and logistic regression based models.

one value of the prognostics index is desired and values greater than this threshold are taken to be high risk patients and vice versa. In this paper, the 'optimum' point is selected to represent a trade-off between *FP* and *TP* rates using the isocost lines method. Fig. 14 shows the distribution of the log of the predicted prognostics indices for the testing data set of the *BCa* data. It is worth recalling that the output of the fuzzy model is the logarithm of the prognostics index for a particular patient. A sample rule (in one dimension) is of the form *if age is 'big', then* log*(risk) is 'high'*. As a consequence of the Lemma in Section III.A, the prognostics index (defined in (8)) has exactly the same interpretation as the output of the fuzzy model since the relationship between this output and prognostics index is a monotonic transformation. Therefore, it would also be correct to transform the sample rule above into *if age is 'big', then risk is 'high'*.

The optimum operating threshold (marked o in Fig. 12) was found to be 0.4. Patients with prognostics

Table IV: *AUC* values on the training and testing data sets using Cox, logistic regression (*LoR*), neural network (*NN*) and fuzzy modelling (type-1 and type-2) on the artificial data and *BCa* data. The type-1 fuzzy model was elicited following the same procedure described for the *IT2* fuzzy model.

| | | **Training** | **Testing** |
|---|---|---|---|
| **Artificial Data** | *Cox* | 0.54 | 0.53 |
| | *LoR* | 0.52 | 0.51 |
| | *NN* | 0.79 | 0.78 |
| | *T1 FLS* | 0.77 | 0.75 |
| | *IT2 FLS* | 0.82 | 0.82 |
| ***BCa* Data** | *Cox* | 0.83 | 0.82 |
| | *LoR* | 0.76 | 0.74 |
| | *NN* | 0.88 | 0.84 |
| | *T1 FLS* | 0.88 | 0.83 |
| | *IT2 FLS* | 0.92 | 0.91 |




(a) Log of prognostics index.

(b) Prognostics index (not to scale.)

Figure 14: Distributions of the predicted prognostics indices using the fuzzy model (testing data). The output of the fuzzy model is the logarithm of the prognostics index (left panel). Exponentiation of the output gives the prognostic index. This is a monotonic transformation so the fuzzy rules are not changed.

Table V: Confusion matrix at the selected optimum point.

| | | **True Class** | |
|---|---|---|---|
| | | Low Risk | High Risk |
| **Hypothesized Class** | Low Risk | 471 | 44 |
| | High Risk | 64 | 103 |

indices greater than this threshold are high risks patients and vice versa.

Table V 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 515 patients to being low risk. From the risk groups, one patient each is selected at random for analysis. The cystectomy

Table VI: Selected patients characteristics

| | **Cystectomy** | **Radiotherapy** | **Prognostics 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 |

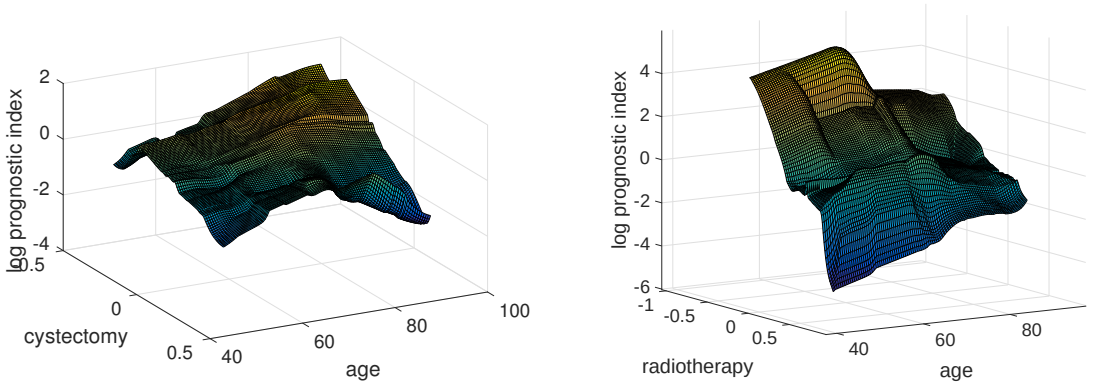
and radiotherapy values as well as their prognostics indices of these two selected patients are shown in Table VI.

On the one hand, it can be observed that where patient $1$ (high risk patient with prognostics 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 prognostics index of $0.01$) had cystectomy performed. Further investigation of patient $1$ reveals that the prognostics 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. 15, 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.

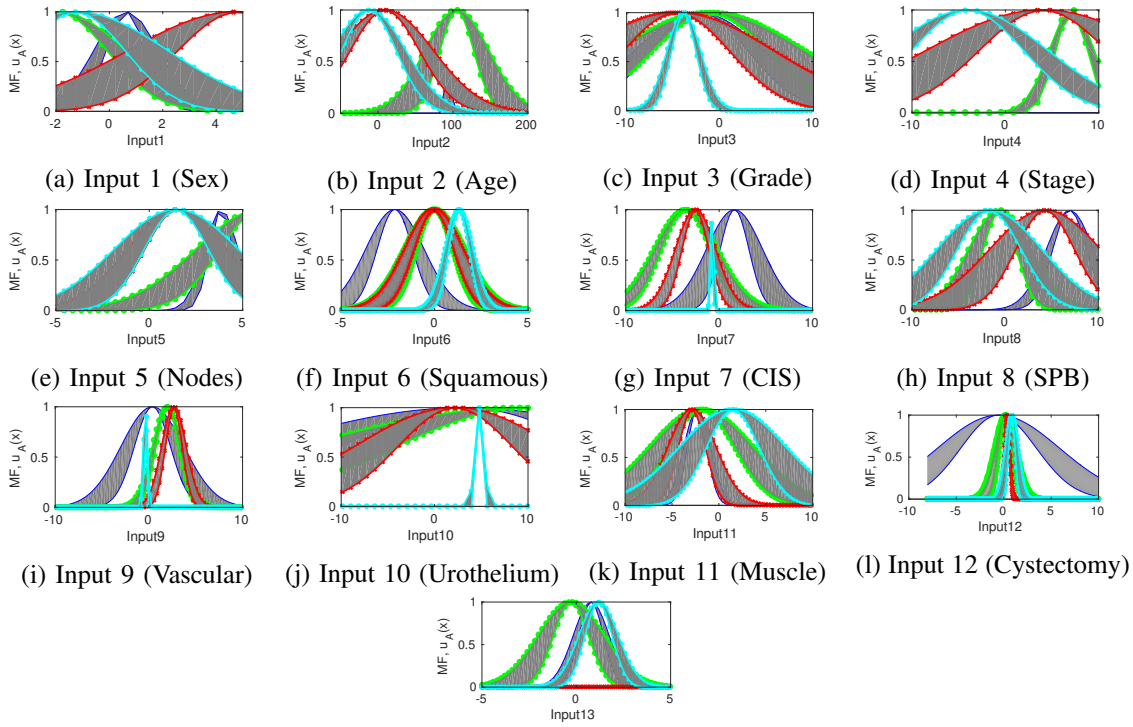
## V. Conclusion

This paper has introduced a new fuzzy modelling approach for solving the conundrum of interpretability and flexibility in survival data modelling. The approach is based on a new framework that integrates intrinsically and, for the first time, the interval type-2 fuzzy modelling approach with the statistically principled Cox model so that the risk scores/prognostics indices can be predicted accurately and in an interpretable but also flexible manner. This new framework is tested on two challenging data sets. The first is a highly non-linear artificially generated data set. The fuzzy model outperforms the Cox model and a neural network-based method, recently reported in the literature, by approximately 5% and




(a) cystectomy and age vs log prognostics index (b) Radiotherapy and age vs log prognostics index

Figure 15: The modelled relationship of age and treatment decisions affect the prognostics 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 prognostics 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 16: A sample rule base of the elicited fuzzy model for the *BCa* data. Only the first four rules are displayed for each input dimension. Similar colours represent same rule.

4% respectively. The second data set is a real life data set containing information relating to patient characteristics, administered treatment (if any) and disease characteristics of patients diagnosed with bladder cancer. The interest in this latter data set is driven by the 'rationale' that one wishes, more often than not, to infer risk groups and provide automatic treatment recommendations as well as risk management decisions for clinicians. When compared with standard modelling frameworks, such as the Cox model, logistic regression and type-1 fuzzy modelling, the proposed framework outperforms these standard modelling methods both in accuracy and generalisation abilities proving emphatically that fuzzy models are more effective alternatives when flexible but interpretable survival models are targeted. The new modelling framework is also flexible since it allows for the elicitation of a minimalistic fuzzy topology consisting of only a handful of fuzzy rules without compromising on accuracy. Future works may consider: 1. Expanding this reduced framework to allow for the possibility of including more fuzzy rules. 2. The inclusion of a probabilistic framework (Bayesian) so that uncertainties in the parameter estimates and outputs of the fuzzy model may be quantified. Additionally, it would be worth investigating how a multi-state modelling approach may be incorporated into the proposed modelling framework such that it includes patients dynamic information such as changing clinical treatments and lifestyles.

## REFERENCES


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

[2] J. Fox, "Cox proportional-hazards regression for survival data," in *An R and S-PLUS companion to applied regression*, Sage, 2002.

[3] D. Collett, *Modelling survival data in medical research*. Florida, USA: CRC press, 2015.

[4] J. Kalderstam, P. Edén, and M. Ohlsson, "Finding risk groups by optimizing artificial neural networks on the area under the survival curve using genetic algorithms," *PloS one*, vol. 10, no. 9, p. e0137597, 2015.

[5] R. M. Ripley, A. L. Harris, and L. Tarassenko, "Non-linear survival analysis using neural networks," *Statistics in medicine*, vol. 23, no. 5, pp. 825–842, 2004.

[6] P. C. Lambert and P. Royston, "Further development of flexible parametric models for survival analysis," *Stata Journal*, vol. 9, no. 2, p. 265, 2009.

[7] J. E. Barrett and A. C. Coolen, "Gaussian process regression for survival data with competing risks," *arXiv preprint arXiv:1312.1591*, 2013.

[8] M. Setnes, R. Babuska, and H. B. Verbruggen, "Rule-based modeling: Precision and transparency," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 28, no. 1, pp. 165–169, 1998.

[9] L. A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, no. 3, pp. 338–353, 1965.

[10] H. O. Wang, K. Tanaka, and M. F. Griffin, "An approach to fuzzy control of nonlinear systems: stability and design issues," *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 1, pp. 14–23, 1996.

[11] J. M. Mendel, "Uncertainty, fuzzy logic, and signal processing," *Signal Processing*, vol. 80, no. 6, pp. 913–933, 2000.

[12] O. Obajemu, M. Mahfouf, and L. A. Torres-Salomao, "A new interval type-2 fuzzy clustering algorithm for interval type-2 fuzzy modelling with application to heat treatment of steel," in *World Congress*, vol. 19, no. 1, 2014, pp. 10 658–10 663.

[13] G. Zhang, M. Mahfouf, M. Abdulkareem, S.-A. Gaffour, Y.-Y. Yang, O. Obajemu, J. Yates, S. A. Soberanis, and C. Pinna, "Hybrid-modelling of compact tension energy in high strength pipeline steel using a Gaussian mixture model based error compensation," *Applied Soft Computing*, vol. 48, pp. 1–12, 2016.

[14] X. Hu, S. E. Li, and Y. Yang, "Advanced machine learning approach for lithium-ion battery state estimation in electric vehicles," *IEEE Transactions on Transportation Electrification*, vol. 2, no. 2, pp. 140–149, 2016.

[15] M. Mohammadi, S.H. Iranmanesh, R. Tavakkoli-Moghaddam and M. Abdollahzadeh, "Hierarchical alpha-cut fuzzy c-means, fuzzy ARTMAP and Cox regression model for customer churn prediction," *International Journal of Engineering-Transactions C: Aspects*, vol. 27, no. 9, pp. 1405-1414, 2014.

[16] L. A. Zadeh, "Discussion: Probability theory and fuzzy logic are complementary rather than competitive," *Technometrics*, vol. 37, no. 3, pp. 271–276, 1995.

[17] J. Kalderstam, P. Edén, P.-O. Bendahl, C. Strand, M. Fernö, and M. Ohlsson, "Training artificial neural networks directly on the concordance index for censored data using genetic algorithms," *Artificial intelligence in medicine*, vol. 58, no. 2, pp. 125–132, 2013.

[18] J. M. Mendel, *Uncertain rule-based fuzzy logic system: Introduction and new directions*, New Jersey, USA: Prentice Hall, 2001.

[19] K. M. Passino, S. Yurkovich, and M. Reinfrank, *Fuzzy control*. California, USA: Addison Wesley Longman, Inc., 1998.

[20] L. Martínez, D. Ruan, F. Herrera, E. Herrera-Viedma, and P. P. Wang, "Linguistic decision making: Tools and applications," *Information Sciences*, vol. 179, no. 14, pp. 2297–2298, 2009.

[21] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—i," *Information sciences*, vol. 8, no. 3, pp. 199–249, 1975.

[22] N. N. Karnik and J. M. Mendel, "Centroid of a type-2 fuzzy set," *Information Sciences*, vol. 132, no. 1, pp. 195–220, 2001.

[23] D. Wu, "Approaches for reducing the computational cost of interval type-2 fuzzy logic systems: overview and comparisons," *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 1, pp. 80–99, 2013.

[24] S. Wang and M. Mahfouf, "A new computationally efficient Mamdani interval type-2 fuzzy modelling framework," in *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE),* 2012 © IEEE. doi: 10.1109/FUZZ-IEEE.2012.6251288.

[25] Q. Liang and J. M. Mendel, "Interval type-2 fuzzy logic systems: theory and design," *IEEE Transactions on Fuzzy systems*, vol. 8, no. 5, pp. 535–550, 2000.

[26] P. Royston and M. K. Parmar, "Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostics modelling and estimation of treatment effects," *Statistics in medicine*, vol. 21, no. 15, pp. 2175–2197, 2002.

[27] F. E. Harrell, K. L. Lee, and D. B. Mark, "Tutorial in biostatistics multivariable prognostics models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors," *Statistics in medicine*, vol. 15, pp. 361–387, 1996.

[28] T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[29] V. Van Belle, P. Neven, V. Harvey, S. Van Huffel, J. A. Suykens, and S. Boyd, "Risk group detection and survival function estimation for interval coded survival methods," *Neurocomputing*, vol. 112, pp. 200–210, 2013.

[30] M. J. Gacto, R. Alcalá, and F. Herrera, "Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures," *Information Sciences*, vol. 181, no. 20, pp. 4340–4360, 2011.

[31] C.-F. Juang and C.-Y. Chen, "Data-driven interval type-2 neural fuzzy system with high learning accuracy and improved model interpretability," *IEEE transactions on cybernetics*, vol. 43, no. 6, pp. 1781–1795, 2013.

[32] J. D. Kalbfleisch and R. L. Prentice, *The statistical analysis of failure time data*. New Jersey, USA: John Wiley & Sons, 2011.

[33] M. Mahfouf, M. Jamei, and D. Linkens, " ," *Materials and Manufacturing processes*, vol. 20, no. 3, pp. 553–567, 2005.

[34] A. Noon, P. Albertsen, F. Thomas, D. Rosario, and J. Catto, "Competing mortality in patients diagnosed with bladder cancer: evidence of undertreatment in the elderly and female patients," *British journal of cancer*, vol. 108, no. 7, pp. 1534–1540, 2013.

**Olusayo Obajemu** received the M.Sc. and Ph.D. degrees from the University of Sheffield, Sheffield, U.K. in 2012 and 2016, respectively. He is currently a Research Associate with the Department of Automatic Control & Systems Engineering at the University of Sheffield, Sheffield, U.K.

His research interests include statistical machine learning, optimal control and computational intelligence methods with applications to health, aerospace and industrial systems.

**Mahdi Mahfouf** received the M.Phil. and Ph.D. Degrees from the University of Sheffield, Sheffield, U.K., in 1987 and 1991, respectively. He has held a Personal Chair in Intelligent Systems Engineering at the University of Sheffield since 2005. He is the author of more than 300 papers and has authored/co-authored one book and five book chapters. He has worked in the areas of intelligent systems, with a strong emphasis on Fuzzy Systems, modelling, control and optimization in medicine, as well as industrial systems, for more than 30 years.

**James W.F. Catto** is a Professor in Urological Surgery at the University of Sheffield and an Honorary Senior Clinical Research fellow at the University of Oxford. He was appointed an Honorary Consultant Urological Surgeon in 2007 and specialises in Uro-Oncology. He obtained a PhD in 2004 and now runs a Research Group investigating the biology of urological cancers. He was awarded the European Association of Urology thesis prize and the Bard medal from BAUS for his PhD studies. In 2010, he was awarded the Crystal Matula by the EAU and in 2011 the Astellas Education Foundation award. He is currently the Editor-In-Chief of the European Urology Journal (Impact factor 10.4). He has published more than 100 primary research manuscripts.