

Incorporation of Expert Variability into Breast Cancer Treatment Recommendation in Designing Clinical Protocol Guided Fuzzy Rule System Models

Jonathan M. Garibaldi¹, Shang-Ming Zhou^{2*}, Xiao-Ying Wang¹, Robert I. John³, Ian O. Ellis⁴

¹Intelligent Modelling and Analysis Research Group, School of Computer Science, University of
Nottingham, NG8 1BB, United Kingdom.

²Health Information Research Unit, College of Medicine, Swansea University, Swansea, SA2 8PP, United
Kingdom

³Centre for Computational Intelligence, Department of Informatics, De Montfort University, Leicester, LE1
9BH, United Kingdom.

⁴Breast Cancer Pathology Research Group, School of Molecular Medical Sciences, University of
Nottingham, NG8 1BB, United Kingdom.

* Corresponding author, email addresses: smzhou@ieee.org (Shang-Ming Zhou).

Abstract:

It has been often demonstrated that clinicians exhibit both *inter-expert* and *intra-expert* variability when making difficult decisions. In contrast, the vast majority of computerized models that aim to provide automated support for such decisions do not explicitly recognize or replicate this variability. Furthermore, the perfect consistency of computerized models is often presented as a *de facto* benefit. In this paper, we describe a novel approach to incorporate variability within a fuzzy inference system using *non-stationary fuzzy sets* in order to replicate human variability. We apply our approach to a decision problem concerning the recommendation of post-operative breast cancer treatment; specifically, whether or not to administer chemotherapy based on assessment of five clinical variables: NPI (the Nottingham Prognostic Index), estrogen receptor status, vascular invasion, age and lymph node status. In doing so, we explore whether such explicit modeling of variability provides any performance advantage over a more conventional fuzzy approach, when tested on a set of 1310 unselected cases collected over a fourteen year period at the Nottingham University Hospitals NHS Trust, UK. The experimental results show that the standard fuzzy inference system (that does not model variability) achieves overall agreement to clinical practice around 84.6% (95% CI: 84.1-84.9%), while the non-stationary fuzzy model can significantly increase performance to around 88.1% (95% CI: 88.0 -88.2%), $p < 0.001$. We conclude that non-stationary fuzzy models provide a valuable new approach that may be applied to clinical decision support systems in any application domain.

Keywords: breast cancer, decision support, expert systems, fuzzy logic, variability

1. Introduction

It is known that medical and health reasoning is inherently uncertain [1]. Doctors and nurses often find it difficult to reach an unequivocal decision for a number of reasons. The data may (will) be

inaccurate and imprecise, knowledge may be uncertain and the linguistic terms used to express the decision may be vague. Traditional modeling techniques are not always suitable for dealing with human knowledge and reasoning involving uncertainty. In contrast, fuzzy sets and fuzzy logic have proven to be able to transfer human knowledge into computer manageable models using linguistic terms. Fuzzy inference systems (also known as fuzzy rule based system, fuzzy expert systems, fuzzy models, fuzzy logic systems, etc. [2]) use a number of ‘if-then’ rules to perform non-linear mappings from input space to output space. As they are based on explicit ‘if-then’ rules featuring linguistic terms, fuzzy systems usually have good interpretability [3]. These ‘if-then’ rules are the heart of a fuzzy inference system. Fuzzy rules are normally obtained from knowledge extracted from ‘experts’ [4][5] or extracted from real-world datasets [7]. However, generating a fuzzy model with optimal (or even just acceptable) performance can often be quite difficult, especially in safety-critical contexts such as biomedical domains [8][9].

On the other hand, from the perspective of health informatics, an important way in which computer scientists assist healthcare providers in making decisions is to develop mathematical models with good predictive performance [9]. Even modest improvements in the predictive performance of mathematical models can greatly benefit patients and healthcare providers. For example, in 1994, it was estimated that just one-percent accuracy improvement in the accuracy of predicting pneumonia outcomes and hospital admissions of pneumonia patients could save about 90 million dollars in healthcare costs per year in the United States [10]. Hence improving model prediction performance is of real interest in medical and health care research.

Many techniques have been developed to tune fuzzy models in order to obtain optimal prediction performance. For example, Garibaldi used simulated annealing to optimise a rule-based expert system to validate and interpret umbilical cord blood acid-base status [11], such that the correlation between the tuned fuzzy model and clinicians was higher than between the clinicians themselves. Lee et al. [12] proposed a fuzzy neural network model which is able to tune the shape of antecedent linguistic terms, rule importance degrees and linguistic term importance. A nonlinear

function was used to evaluate the proposed method, which demonstrated good results [12]. The Adaptive-Network-based Fuzzy Inference System (ANFIS) developed by Jang [1] is a well known fuzzy model tuning method. In general, there are two key steps in building an ANFIS model. The first step is to construct fuzzy sets used to describe different fuzzy rules. The second step is to optimise rule parameters by using a hybrid method which combines gradient descent and a least squares approach. The ANFIS method has been widely used in various applications, such as [13][14][15][16]. However, these conventional fuzzy model tuning methods focus only on improving global model accuracy without consideration of preserving model interpretability, two conflicting modeling objectives [3][17].

Breast cancer is the most common cancer found in women in the United Kingdom and accounts for almost one in three of all cancers. In 2005, over 45,000 women were diagnosed with breast cancer. More than 12,000 women die from breast cancer every year [18]. In light of this, it is vital to select the proper course of treatment once a diagnosis has been made. However, this is a very difficult and complex procedure. In many countries, in order to achieve optimal treatment, a multidisciplinary team (MDT) consisting of oncologists, radiologists, pathologists and surgeons meets to make treatment decisions [19][20]. During a meeting, all members of the MDT discuss each patient case individually. Normally there is a meeting coordinator to collect case-notes, radiographs and pathology reports to facilitate the meeting [21]. Inevitably there are many uncertain and imprecise terms used during MDT discussions and decision making. Therefore a fuzzy model is an obvious choice to simulate human decision making in breast cancer.

Pena-Reyes and Sipper developed a fuzzy-genetic method applied to the Wisconsin breast cancer diagnosis data which generated a complete fuzzy inference system [8]. A neuro-fuzzy decision model was proposed for prognosis of breast cancer relapse in 2003 [22], in which the pathological data was provided by the oncology service of the University Hospital of Malaga. This model firstly used a fuzzy system as a filter to eliminate patients whose clinical parameters did not comply with specified criteria, then a back-propagation artificial neural network was applied to the

output of fuzzy system. Finally a threshold unit was used to group the output from the network into 'relapse' and 'non relapse' groups. Recently, a hybrid method of genetic algorithm and artificial immune system was developed to generate various fuzzy inference system parameters, based on a database of training cases, and this was tested on the Wisconsin breast cancer diagnosis problem [23]. The results showed that the fuzzy inference system produced from the hybrid algorithm reached the maximum classification before either the genetic algorithm or artificial immune system individually.

However, in clinical and health decision making, it has been often demonstrated that variation may occur among a panel of human experts (*inter-expert variability*) and within an individual expert over time (*intra-expert variability*) when facing with the same problem (e.g. the same patient). The reasons behind this variation might be that doctors and nurses gain more knowledge and experience whilst serving on the multidisciplinary panel or simply attributable to changes in the expert's mood or physical state (e.g. tiredness or stress) at the time of making the decision. Variability in clinical treatment of breast cancer has been well documented. For example, Gort et al reported large variations in breast conserving surgery between hospitals, with rates ranging from 25.5% to 79.3% for small tumors (<2 cm) [34], while Bueno-de-Mesquita et al reported that 14% of patients would have been given different clinical risk and treatment advice based on the Nottingham prognostic index value derived from discordant original pathological evaluations [35]. However, such variability is often ignored in most modeling techniques, including most fuzzy system design.

The concept of a "*non-stationary fuzzy set*" has recently been proposed to explicitly model variation in uncertainty [24][25][26][27]. The membership function of a *non-stationary fuzzy set* is allowed to be altered over time. A fuzzy inference system built upon such non-stationary sets will then exhibit variability in the decisions obtained. The various decisions obtained may be viewed as a set of potentially acceptable alternative decisions, such as those obtained from a panel of experts. Subsequently, ensemble and consensus methods may be employed to obtain or recommend the

best option from the set of acceptable alternatives, or may be used to explore the acceptable decision boundaries for a particular case. While the variation in non-stationary fuzzy sets may be generated randomly (as in this paper), this is by no means necessary. Further, the use of random variations in the fuzzy sets to model expert variability does *not* imply that we assume that the expert variation is itself random.

In this paper, we create a fuzzy inference system with twelve *if-then* fuzzy rules to implement the clinical post-operative treatment recommendation protocol for breast cancer used by the doctors in the Nottingham University Hospitals NHS Trust. A non-stationary fuzzy inference system with explicit representation of variability is then implemented to simulate human decision making in this context and to explore whether the incorporation of variability is of benefit. A set of 1310 breast cancer cases are used to evaluate the performance of various configurations of the non-stationary fuzzy inference system, in comparison with the system without variability and a conventional ANFIS approach. The layout of the paper is as follows. Section 2 provides background on non-stationary fuzzy sets and illustrates their use in a simple example. Section 3 provides a detailed description of the general clinical problem, the specific clinical guidelines being modeled and the data. Section 4 describes the methods used including the design of the original fuzzy inference system, the variability models and the alternative ANFIS approach. The results of the evaluation experiments are then presented, followed by discussion of our findings. Finally, the conclusions and some areas of potential future work are provided.

2. Background

In this section, we introduce the concept of non-stationary fuzzy sets with an example.

2.1. Non-Stationary Fuzzy Sets

Non-stationary fuzzy sets have membership functions that change with time. \dot{A} is a non-stationary fuzzy set over a universe of discourse X with membership function $\mu_{\dot{A}}(x, t)$, where $x \in X$, $\mu_{\dot{A}}(x, t) \in [0, 1]$, and t refers to the time at which the fuzzy set is instantiated. We let A be the underlying standard fuzzy set with $\mu_A(x)$ as its membership function. A is an *instantiation* of \dot{A} for each time t . The non-stationary fuzzy set \dot{A} can be denoted by

$$\dot{A} = \int_{t \in T} \int_{x \in X} \mu_{\dot{A}}(x, t) / x / t$$

Now we introduce a perturbation function $f(t)$ to adjust the standard membership function $\mu_A(x)$. In theory, $f(t)$ could be a true random function, following (for example) a normal distribution, but in general, any function that is a function of time is eligible, for instance, a sinusoidal function $f(t) = \sin(\omega t)$. So we now introduce the parameters of $\mu_A(x)$ as p_1, p_2, \dots, p_m , so if we express

$$\mu_{\dot{A}}(x, t) = \mu_A(x, p_1(t), \dots, p_m(t)),$$

then $p_i(t) = p_i + k f_i(t)$, where $i = 1, \dots, m$. This means that each time each parameter is varied by a perturbation function multiplied by a constant.

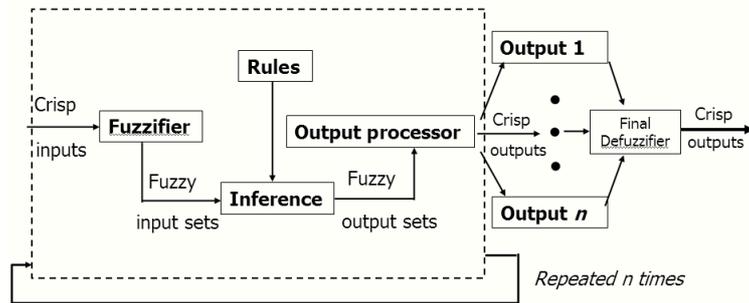


Figure 1: An outline of a non-stationary fuzzy inference system

A non-stationary fuzzy inference system (FIS) is shown in Figure 1. A non-stationary FIS is a

normal FIS that is run n times (i.e. n instantiations of an ordinary FIS).

2.2. Illustration of Non-Stationary Fuzzy Sets

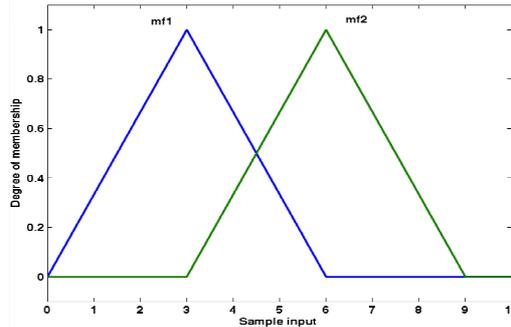
Figure 2 shows an example of a non-stationary fuzzy set. Figure 2(a) illustrates a linguistic variable containing two standard membership functions $mf1$ and $mf2$. The parameters for $mf1$ are the left-point, centre-point and right-point (0, 3, and 6, respectively). This can be expressed as $\mu_{mf1}(x, p_1, p_2, p_3) = \mu_{mf1}(x, 0, 3, 6)$. The corresponding non-stationary membership of $mf1$ is $\mu_{mf1}(x, t) = \mu_A(x, p_1(t), p_2(t), p_3(t))$, with parameters

$$p_1(t) = 0 + k_1 f_1(t),$$

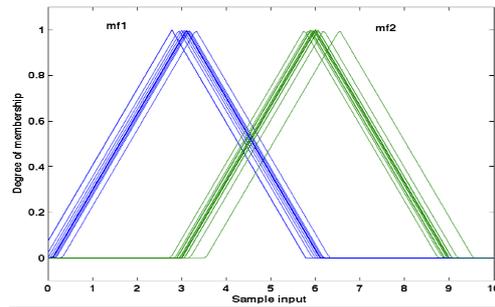
$$p_2(t) = 3 + k_2 f_2(t),$$

$$p_3(t) = 6 + k_3 f_3(t),$$

where $[0, 10]$ is the universe of discourse. The perturbation functions $f_1(t) = f_2(t) = f_3(t) = f(t)$ and $k_1 = k_2 = k_3 = k$ are used. In Figure 2(b), $f(t)$ is a normal distribution with mean of zero and standard deviation $\sigma = 0.02$, $k = 1$ and n , the number of iterations, is 10.



(a) underlying membership functions



(b) multiple instantiations

Figure 2: An example of non-stationary fuzzy sets

3. The Clinical Problems

3.1 Description of the Clinical Problem

The subsequent treatment for breast cancer following primary treatment (normally surgery) is called adjuvant therapy, and is used to reduce the chance of cancer reoccurrence. The adjuvant therapy treatment may include hormone (anti-estrogen) therapy, radiotherapy, chemotherapy, biological therapy, further operation and clinical follow-up, or any combination of these [28].

The decisions regarding the recommended course of adjuvant treatment are normally made during multi-disciplinary team (MDT) meetings. The clinical procedure employed for recording the data can be summarised by the following steps:

- the attribute information and additional comments related to each patient's treatment are recorded on a form;
- the forms are discussed by the various clinicians present during the multi-disciplinary meeting and a further course of action is agreed; and
- after the meeting, the forms are collected and sent to a data analyst for entry into a computer database.

The central aim of this paper is to design a computerised rule-based model capable of simulating

this process and reproducing the specific recommendation for the administration of chemotherapy (only). As there is no ‘gold-standard’ for the selection of adjuvant therapy and many factors that complicate the assessment of whether any particular decision was appropriate for the given patient (such as whether the patient subsequently complies with any treatment recommendation), in this study we compared the output of the automated system against the *initial recommendation of treatment* agreed by the MDT. Thus, we are assessing whether we can model the decision making process of the MDT, and are *not* assessing the *effect* of the chemotherapy recommendation and/or the actual adjuvant therapy administered in practice.

3.2 Clinical Guidelines

At Nottingham, the clinicians have developed a written protocol for breast cancer treatment decisions that is routinely used for patients, as reproduced *verbatim* in Table 1. It can be seen from the guidelines that, while the recommendation for hormone therapy is straight-forward, the recommendation for or against chemotherapy is a complex decision based on several factors, which are combined in an unspecified manner. It can also be seen that several of these factors have clear (crisp) decision boundaries, such as ‘Age < 40’, but in real world applications these boundaries are likely to be vague. As examples, firstly, although the guidelines for patients with NPI > 4.4 and ER positive status feature crisp boundaries for ages < 40 and > 60, it is unclear what the recommendation should be between these ages; secondly, the NPI boundary at 3.4 is clearly ambiguous in the ranges 3.1 – 3.4 and 3.4 – 4.4. Fuzzy sets can be a very good choice for dealing with such vagueness. Consequently, it was decided to attempt to design a fuzzy rule-based system to model the recommendation of chemotherapy (model output) based on the factors (model inputs) featuring in the protocol.

Table 1: The Nottingham University Hospitals NHS Trust clinical guidelines for adjuvant therapy following surgery (reproduced verbatim)

NPI < 3.0	None
NPI 3.1 – 3.4	
ER +ve	Recommend Hormone therapy
ER -ve	Recommend Chemotherapy if VI
NPI 3.4 – 4.4	
ER +ve	Recommend Hormone therapy
ER -ve	Recommend Chemotherapy
NPI > 4.4	
ER +ve	Discuss Chemotherapy
	Consider:
	Recommending Chemotherapy:
	Age < 40
	VI
	HER-2 +ve
	Weak ER (< 100/300)
	Recommending Against Chemotherapy:
	Age > 60
	Only 1 LN positive
	Special type cancer
ER -ve	Recommend Chemotherapy

ER +ve: ER is positive

ER -ve: ER is negative

Age: in years

HER-2: Human Epidermal growth factor Receptor 2

VI (Vascular invasion): presence of unequivocal tumor in vascular spaces

It is a challenging task to induce trustworthy if-then rules in designing a fuzzy recommendation system. Currently there are two existing approaches for this task. One is to generate fuzzy rules according to domain experts' knowledge [4][5], the other is to induce fuzzy rules from real-world datasets [6][7]. The fuzzy rules generated from experts' knowledge possess good interpretability, but the system predication ability can not always be guaranteed. The advantage of data driven rule induction approaches lies in that more specific knowledge characterised by trained fuzzy sets can be generated with high model prediction accuracy. Given that machine learning algorithms used to derive fuzzy models from data are usually accuracy-oriented, the fuzzy rules which are automatically induced from datasets may lose interpretability [3][17], as model accuracy and model interpretability are usually two conflicting modelling objectives. In this paper, we employ the first approach to generate fuzzy rules according to the written clinical protocol.

3.3 Description of Raw and Processed Study Data

The data involved in this study is a set of real world clinical data concerning post-operative breast cancer treatment that was kindly provided by The Nottingham Breast Institute (within the Nottingham University Hospitals NHS Trust). This is a set of post-operative data collected from a total of 8704 patients who had all undergone some form of breast cancer operation (e.g. wide local excision, auxiliary node clearance or sample) over a twenty-seven year period from September 1982 to August 2009. The data is comprised of a set of attributes examined on each patient's post-operative visit and adjuvant therapy treatment decision. The attributes present in the data include:

- *Encrypted patient identification code, date of birth, visit and diagnosis date.*
- *Invasive carcinoma size, grade, type and margins.*

Invasive carcinoma describes the cancer that has already broken through the layer of cells where it started (as opposed to carcinoma in situ). Invasive carcinomas are potentially life-threatening, and even with appropriate treatment not everyone will be cured.

- *Ductal carcinoma in situ (DCIS) size, grade, type and margins.*

DCIS means that the cancer cells remain confined inside the ducts but have not grown into (invaded) deeper tissues in the breast or spread to other organs in the body at the early stage of cancer. Carcinoma in situ of the breast is sometimes referred to as non-invasive or pre-invasive breast cancer.

- *Whole tumor size (in mm).*
- *Nottingham prognostic index (NPI) value.*

The NPI was developed to help prognosis following surgery for breast cancer as an attempt at using some fairly objective parameters to determine the odds that a newly diagnosed case of invasive ductal adenocarcinoma would benefit from adjuvant chemotherapy. The NPI value is calculated in terms of three pathological criteria: the size of the lesion; the number of involved lymph nodes; and the grade of the tumor.

- *Estrogen receptor (ER) test result*

Estrogen receptor is a protein molecule found inside cells that are targets for estrogen

action. The ER status tests show whether or not one or both of those hormones fuel the tumor. Hormone sensitive cancer grows slower than hormone negative cancer, so it would benefit from hormone-suppression treatment. Hormone negative cancer responds to other types of treatment rather than hormone-suppression.

- *Progesterone receptor (PR) test result.*

The progesterone receptor is an estrogen-regulated protein. The expression of PR determination indicates a responsive ER pathway, so it could predict likely response to endocrine therapy in breast cancer. For example, PR positive cancer would respond well to hormone suppression treatments.

- *Vascular invasion (VI) test result.*

Breast cancer will go into the channels, like the lymph channels, blood channels, particularly the venous system before it spreads to the rest of the body. VI tests show whether or not the tumor cells have acquired the ability to invade the walls of these channels and may spread to the lymph nodes or beyond. VI is determined according to guidelines of the Royal College of Pathologists [30] as a categorical variable.

- *Lymph node stage, the number of positive lymph nodes (LN) found from samples.*

Lymph nodes throughout the body are an important part of human immune system. The clinical significance of lymph nodes lies in that they become inflamed or enlarged in various conditions such as cancers, so it can be used for cancer staging to determine the prognosis and choose the treatment.

Several challenges arise in this study due to the clinical procedures for recording the information. For example, the data consists of the input from a number of different data analysts over the period and, because there is no standardised format for data entry, inconsistencies in data formats can often occur. In addition, the treatment decisions are not separately identified in their own data field and are, instead, hidden within a free-text comment field. The processing of the data

is further complicated by the fact that different treatment decisions may be identified through different notations or spellings. Furthermore, there can be missing values for some attributes. This affects the clarity of the data and complicates automatic classification. All patient IDs were encoded (translated from the originals using a one-way encoding algorithm), to permit correspondence analysis of patients, but not permitting actual identification. All necessary ethical permissions were obtained.

From Table 1, there are seven factors that feature in the recommendation of chemotherapy: NPI, ER, VI, age of patient (Age), lymph node (LN) status, HER-2 (Human Epidermal growth factor Receptor 2) status and 'Special type cancer'. Of these, we determined the five variables, NPI, ER, VI, Age and LN to be necessary as inputs to the fuzzy system. While lymph node involvement is incorporated into the NPI variable, the clinical protocol clearly utilises the LN feature itself separately from, and in conjunction with, NPI. While this may mean that the clinical protocol over-represents lymph node involvement, and that the protocol incorporates a subtle correlation between NPI and LN, this 'dual-involvement' of LN is maintained in the fuzzy model. By doing so, the fuzzy model has a direct correspondence to the clinical protocol (compare Tables 1 and 4). Whether this is a desirable or clinically justified feature of the protocol in use at Nottingham is outside the scope of this paper.

HER-2 status had been relatively recently introduced into the clinical guidelines (in 2007) and so most patients had no HER-2 status recorded. We decided to consider only those patients assessed without the inclusion of HER-2 status into the clinical guidelines. That is, the selected patients includes both HER-2 positive and negative patients who (historically) had been assessed by the MDT prior to HER-2 status being available and, thus, before having any effect on the decision process. Similarly, for any patient without a recorded HER-2 status, this clearly cannot have been taken into account by the MDT, and so cannot have influenced the recommendation. So, any patient with HER-2 status recorded (whether positive or negative) was eliminated from the data. The "Special type cancer" feature was neither generally available nor clearly described in

the given dataset. Further, we could identify no clearly articulated rules as to how the ‘special type’ of cancer influenced decision making. Consequently, we decided to ignore this variable within the decision model. While clearly unsatisfactory, the elimination of this feature within the model may simply lead to a reduction in the maximum overall model accuracy that may be achieved. In effect, the feature becomes a ‘hidden variable’ which may affect the decision making of the MDT in clinical practice, but which our model does not include.

The output chemotherapy variable was determined by analysis of the free-text comment field of the raw data base. Specifically, only a clear indication of chemotherapy was taken, although this manifested in several forms. For example, textual entries of ‘Chemo Yes’, ‘Chemo+’, ‘Advise Chemo’, etc., were translated to chemotherapy ‘Yes’, while entries of ‘Chemo No’, ‘Advise no Chemo’, etc., were translated as ‘No’, and entries specifically indicating ‘Discuss Chemo’ or ‘Consider Chemo’ were translated as ‘Maybe’. Any patient record with no clear indication of chemotherapy recommendation was not selected.

Only those cases having complete clinical records for these five critical variables, without recorded HER-2 status, and with a chemotherapy recommendation were retained, leading to a total sample of 1310 patients spanning a fourteen year period from July 1993 to March 2007. The patients were otherwise unselected — i.e. all patients who matched these criteria were selected. While clearly the elimination of any patient with missing information may (technically) introduce bias into the study, this limitation is not considered to be critical for the following reason. If the recommendation produced by the fuzzy model was being assessed for ‘correctness’ against a ‘gold-standard’ or against patient outcomes (survival times, etc.) in some manner, then this bias could affect such assessment. However, given that the model recommendation is compared only against the MDT recommendation, the reader should simply bear in mind that the accuracies reported in the results are all ‘when all features are present’. We make no claims for the model accuracy in the absence of any input features. The clinical guidelines for chemotherapy at Nottingham (with the exception of HER-2 status as previously described) remained the same over

this period (July 1993 to March 2007).

3.4 Input and Output Variable Description

An elementary statistical description of the input and output variables, in terms of median and range of the numeric variables and the distributions of the categorical variables are shown in Table 2 and Table 3. For more detailed description of each variable, see Section 4.1.1 below.

Table 2: Description of Continuous Variables

<i>Continuous variable</i>	<i>Median</i>	<i>Range</i>
NPI	3.56	2.04 ~ 10.00
ER	180	0 ~ 300
Age	59	26 ~ 88
LN Positive	0	0 ~ 40
LN Removed	5	0 ~ 43
LN Ratio	0.00	0.00 ~ 1.00

Table 3: Description of Categorical Variables

<i>Categorical variable</i>	<i>Number (Percentage) of Patients within each Category</i>		
	<i>No</i>	<i>Maybe</i>	<i>Yes</i>
VI	876 (66.9%)	112 (8.5%)	322 (24.6%)
Chemo	1067 (81.5%)	52 (4.0%)	191 (14.6%)

4. Methods

4.1 Standard Fuzzy System Design

A standard (type-1) fuzzy system to recommend chemotherapy treatment based on the clinical guidelines above was created. The four main interconnected components used to generate a fuzzy system are the input fuzzifier, the rule base, the inference engine, and the output processor (defuzzifier) [24]. In the following, we describe how each part of the standard fuzzy system for breast cancer treatment recommendation was constructed.

4.1.1 Input and Output Variable Determination

As described above, five input variables and one output variable were selected. By referencing the parameters provided in the guidelines given in Table 1, the terms of each of the fuzzy variables and their associated membership functions were carefully determined. In our research, either triangular or trapezoidal membership functions were used for all of the fuzzy terms, since this made it straight-forward to determine the intersection points between different membership functions. Triangular membership functions are described by the notation (l, m, r) corresponding to the left vertex, middle vertex and right vertex of the triangle, while trapezoidal membership functions are described by the notation (a, b, c, d) corresponding to the left-base, left-top, right-top and right-base vertex, respectively. The initial design of the fuzzy sets within each of the five inputs and one output is detailed as follows.

A) NPI (Nottingham Prognostic Index)

The value of NPI is calculated according to the size of the lesion; the number of involved lymph nodes; and the grade of the tumor [28]. Specifically speaking, the NPI is defined as

$$NPI = 0.2S + N + G$$

where S represents the size of the indexed lesion in centimeters, N represents the number of lymph nodes involved: $N=0$ for negative nodes, $N=2$ for 1 to 3 positive nodes, and $N=3$ for 4 or more positive nodes; and G is the grade of tumor giving a score of 1 (better) to 3 (worse).

It can be seen from the clinical guidelines that there are four ranges of NPI values, namely $NPI \leq 3.0$, $NPI \in [3.1, 3.4]$, $NPI \in [3.4, 4.4]$ and $NPI > 4.4$. Four fuzzy terms associated with these ranges are designed as *Low*, *Medium Low*, *Medium High*, and *High*, respectively. The universe of discourse of NPI was set to be $[0, 10]$. The parameters of the four membership functions of the initial NPI fuzzy sets were set to be:

- *Low*: (0, 0, 2.8, 3.2)
- *Medium Low*: (2.8, 3.2, 3.6)
- *Medium High*: (2.9, 3.9, 4.9)
- *High*: (3.9, 4.9, 10, 10)

Hence, *Low* intersects *Medium low* at 3.0, *Medium low* intersects *Medium High* at 3.4, and *Medium High* intersects *High* at 4.4. The initial NPI membership functions are shown in Figure 3(a).

B) ER (Estrogen Receptor Status)

A positive ER score gives an indication that estrogen is causing the tumor to grow, and that the cancer should respond well to hormone suppression treatments. A negative ER score means that the tumor is not driven by estrogen, so other tests, such as HER2 status, are needed to determine the most effective treatment. ER is over-expressed in around 70% of breast cancer cases, referred to as ‘ER-positive’.

The universe of discourse of the estrogen receptor status is defined clinically as [0, 300]. From the guidelines three terms for ER were apparent, namely “*negative*”, “*weak*” and “*positive*”. These terms had clearly defined, crisp clinical cut-offs of 20 (being the cut-off between “*negative*” and “*positive*”) and 100 (being the upper cut-off for “*weak*”) and were associated directly with the fuzzy terms *Negative*, *Weak* and *Positive*. The parameters of the three membership functions of the ER fuzzy sets were set to be:

- *Negative*: (0, 0, 40)
- *Weak*: (0, 40, 160)
- *Positive*: (40, 160, 300, 300)

Hence, *Negative* intersects *Positive* at 20, while *Weak* reaches a membership of 0.5 at 100. The ER membership functions are shown in Figure 3(b).

C) Age

The maximum age within our dataset was 88. From the guidelines, in which cut-offs of 40 and 60

can be seen, three fuzzy terms were readily apparent: *Young*, *Middle Aged*, and *Old*. The parameters of the three membership functions of the *Age* fuzzy sets were set to be:

- *Young*: (0, 0, 30, 50)
- *Middle age*: (30, 50, 70)
- *Old* : (50, 70, 90, 90)

Hence, *Young* intersects *Middle Aged* at 40, while *Middle Aged* intersects *Old* at 60. The *Age* membership functions are shown in Figure 3(c).

It should be noted that the three terms - *Young*, *Middle Aged*, and *Old* are just the universal linguistic terms used for describing human age. It does not mean that breast cancer for the persons aged less than 15 (for example) would be as frequent as ones aged 30s to 40s.

D) VI (Vascular Invasion)

The presence of VI is generally considered to be an adverse feature providing independent prognostic information about both local recurrence and survival. VI is determined according to guidelines of the Royal College of Pathologists [30]: the presence of unequivocal tumor in vascular spaces is recorded as VI 'Yes'; if there is doubt, but it is considered to be likely, it is recorded as 'Maybe'; and if definitely not present it is recorded as 'No'.

The VI is a categorical variable. The three categorical values of vascular invasion, *No*, *Maybe* and *Yes*, were each associated with a fuzzy term. The parameters of the three membership functions of the VI fuzzy sets were set to be:

- *No*: (1, 1, 3)
- *Maybe*: (0.75, 2, 3.25)
- *Yes*: (1, 3, 3)

The initial VI membership functions are shown in Figure 3(d).

E) LN (Positive Lymph Node Ratio)

During surgery, patients have some lymph node samples taken in order for further analysis to be

performed. For example, ten lymph nodes may be taken, for which subsequent analysis identifies that eight of the nodes are positive and two of the nodes are negative (for cancerous changes). In the guidelines provided, when NPI is greater than 4.4 and ER is positive, the recommendation will be against chemotherapy when there is only one positive lymph node found. However, in the given clinical data, the number of lymph nodes taken from patients varied from 0 to 43, and the number of positive lymph nodes found varied from 0 to 40. So, “*there is only 1 lymph node positive*” would apply equally to a patient with one positive out of one taken, and to one positive out of 40 taken.

In order to allow a fuzzy interpretation of this variable, the ratio of the number of positive lymph nodes to the number of lymph nodes taken was used as the input value for the LN variable. Note that there is no clear clinical justification for using lymph node positive ratio, as opposed to simply the number of positive lymph nodes; this is an artificial device used to permit a fuzzy interpretation of the LN variable. In the data, for the cases in which there is only one positive lymph node, the value of LN lies within a range $[1/36, 1/1]$, or roughly $[0.03, 1]$. From this, the universe of discourse was set as $[0, 1]$. Two fuzzy terms were created, corresponding to *Negative* and *Positive*. The parameters of the two membership functions of the LN fuzzy sets were set to be:

- *Negative*: $(0, 0, 0.04)$
- *Positive*: $(0, 0.2, 1, 1)$

Hence, *Negative* intersects *Positive* at approximately 0.03, corresponding to a lymph node ratio of 1 in 36. The initial LN membership functions are shown in Figure 3(e).

F) Chemotherapy (Chemo)

The output for the fuzzy system, Chemo, represents the recommendation in favor of or against chemotherapy adjuvant treatment. It should be noted that the recommendation made by the clinician in the MDT may not correspond to the treatment that was finally administered (as a patient may decide against receiving a recommendation, such as chemotherapy, even if it is

strongly recommended by the clinician). The variable contains three possible outcomes: *Yes* (corresponding to a recommendation for chemotherapy), *No* (corresponding to a recommendation against chemotherapy) and *Maybe* (corresponding to a recommendation to discuss or consider chemotherapy). If the recommendation is “*no adjuvant treatment*” or “*recommend hormone therapy*”, it was considered as a recommendation “against chemotherapy” for this study. The category of “*Maybe*” is used, because the clinical free-text comments showed that for some breast cancer cases, the MDT did not make clear recommendations like “*recommend chemotherapy*”, instead, recommendations like “*discuss chemotherapy*”, “*possible chemotherapy*” were made. In these cases, the patients might not prefer *chemotherapy* treatment, or the *chemotherapy* might help to reduce the risk of the cancer coming back but certainly was not the optimal way of treatment. So we treated these decisions as an additional treatment plan- “*Maybe*”.

The universe of discourse of the Chemo variable was set as [0, 100]. From observation of the actual recommendations in the database, it was found that many fewer cases appear in the *Maybe* category compared to either *Yes* or *No* categories. Therefore in the membership function design, *Maybe* was created to occupy a much narrower range than the other two. By a process of manual tuning, the cut-off points between these three options was found to achieve the best results when *Maybe* occupied a range around [55, 57]. Based on these findings, the three membership functions of the fuzzy sets for the Chemo output variable were set to be:

- *No*: (0, 0, 54, 56)
- *Maybe*: (55, 55, 57)
- *Yes*: (55, 57, 100, 100)

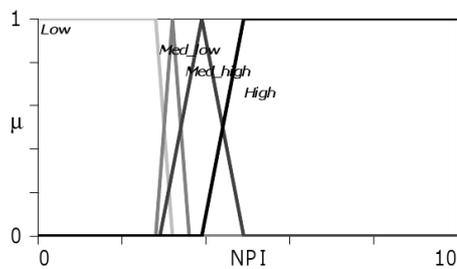
Hence, *No* intersects *Maybe* at 55, and *Maybe* intersects *Yes* at 56. The initial Chemo membership functions are shown in Figure 3(f).

4.1.3. Fuzzy Rules

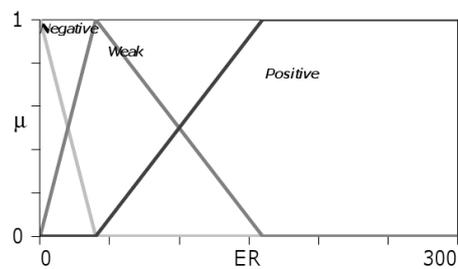
As the rules are the elements which define the relationship between the input and the output variables, the generation of a set of trustworthy rules is the key to producing a good quality fuzzy inference system. As mentioned before, fuzzy rules can be obtained from the knowledge of human experts (knowledge-driven) or extracted from an existing database (data-driven). In this paper, we used the first method to derive the rules since the clinical guidelines provided by the Nottingham University Hospitals NHS Trust provide a formal framework for clinical decision making in this context. Hence, a set of fuzzy rules were derived directly from these guidelines, as shown in Table 4.

4.1.4. Fuzzy Inference Method

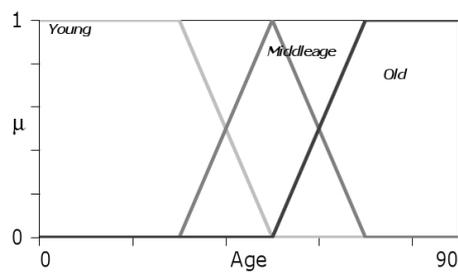
The fuzzy decision making process is implemented using a conventional Mamdani style fuzzy inference system. The usual min and max operators are used for conjunction and disjunction, respectively, and also for implication and aggregation, respectively. Centroid (centre-of-gravity) defuzzification is used to obtain the crisp output of fuzzy inference.



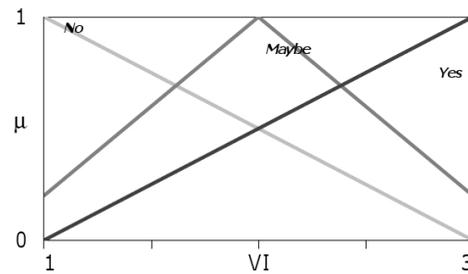
(a) NPI



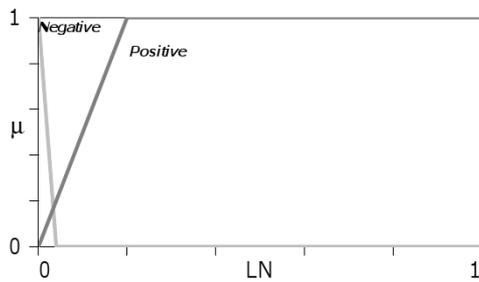
(b) ER



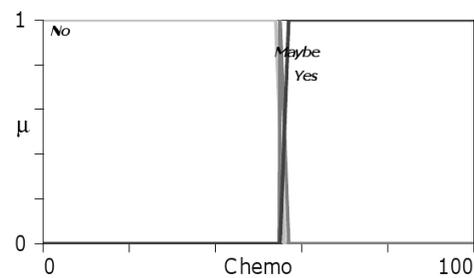
(c) Age



(d) VI



(e) LN



(f) Chemo

Figure 3: Initial membership functions for the fuzzy variables

Table 4: Fuzzy rules derived directly from the clinical guidelines

Rule	Antecedent	Consequent
1	IF (<i>NPI is Low</i>)	THEN (<i>Chemo is No</i>)
2	IF (<i>NPI is Medium low</i>) and (<i>ER is not Negative</i>)	THEN (<i>Chemo is No</i>)
3	IF (<i>NPI is Medium low</i>) and (<i>ER is Negative</i>)	THEN (<i>Chemo is Maybe</i>)
4	IF (<i>NPI is Medium high</i>) and (<i>ER is not Negative</i>)	THEN (<i>Chemo is No</i>)
5	IF (<i>NPI is Medium high</i>) and (<i>ER is Negative</i>)	THEN (<i>Chemo is Yes</i>)
6	IF (<i>NPI is High</i>) and (<i>ER is not Negative</i>)	THEN (<i>Chemo is Maybe</i>)
7	IF (<i>NPI is High</i>) and (<i>ER is not Negative</i>) and (<i>Age is Young</i>)	THEN (<i>Chemo is Yes</i>)
8	IF (<i>NPI is High</i>) and (<i>ER is not Negative</i>) and (<i>VI is Yes</i>)	THEN (<i>Chemo is Yes</i>)
9	IF (<i>NPI is High</i>) and (<i>ER is Weak</i>)	THEN (<i>Chemo is Yes</i>)
10	IF (<i>NPI is High</i>) and (<i>ER is not Negative</i>) and (<i>Age is Old</i>)	THEN (<i>Chemo is No</i>)
11	IF (<i>NPI is High</i>) and (<i>ER is not Negative</i>) and (<i>LN is Negative</i>)	THEN (<i>Chemo is No</i>)
12	IF (<i>NPI is High</i>) and (<i>ER is Negative</i>)	THEN (<i>Chemo is Yes</i>)

4.1.5. Output Processing

The final step is to convert the crisp value obtained from the fuzzy inference system into a

categorical agreement in the set $\{No, Maybe, Yes\}$. The crisp value was compared to the fixed intersection points of the *No*, *Maybe* and *Yes* membership functions of the Chemo variable. The intersection of *No* and *Maybe* occurs at 55, while the intersection of *Maybe* and *Yes* occurs at 56, as described above. Hence, the label of *No*, *Maybe* or *Yes* is assigned according to whether the crisp value obtained through defuzzification falls in the interval $[0, 55]$, $(55, 56]$ or $(56, 100]$, respectively. Finally, in order to compare the performance of the fuzzy system against the actual decisions recommended in clinical practice, an agreement confusion matrix is generated; an example is shown in Table 5. In this paper, the absolute categorical agreement is used: for instance, in the example shown in Table 5, the agreement is:

$$\frac{\sum_{i=1}^3 C_{i,i}}{\sum_{i,j=1}^3 C_{i,j}} = \frac{1000 + 3 + 115}{1310} = 85.34\%$$

Table 5. An example of a confusion matrix of decision agreement

		Clinician Decision		
		No	Maybe	Yes
Decision by Fuzzy System	No	1000	37	72
	Maybe	2	3	4
	Yes	65	12	115

4.2 Fuzzy Model Design using Non-Stationary Fuzzy Sets for Breast Cancer Treatments

There were two main mechanisms used to explore the effects of alternative fuzzy models in this decision making scenario. Firstly, the choice of alternative membership functions and, secondly, the introduction of various amounts of variability into the inference through the use of non-stationary fuzzy sets. The selection of the set of input variables, the membership functions of all associated terms, the fuzzy rules and the amount (and nature) of variability to be incorporated in the non-stationary fuzzy sets within the system is collectively referred to as the selection of the ‘fuzzy model’. The specification of the membership functions of all the terms is, thus, one

component of the fuzzy model. In this paper we use the term ‘fuzzy foundation’ or just ‘foundation’ to refer to the selection of membership functions.

4.2.1. Alternative Foundations

The use of alternative membership functions was explored through the determination of an alternative set of membership functions for four of the five input variables, namely *NPI*, *Age*, *VI* and *LN*. The *ER* variable was not altered as its points of intersection were fixed by the presence of accepted clinical thresholds (at 20 and 100): they were so well established that alteration of the membership functions would be clinically counter-intuitive. Alternative forms for the membership functions for four variables were variously combined to give sixteen foundations. The parameters of the alternative membership functions are given below, and the resultant membership functions are illustrated in Figure 5.

A) NPI

The parameters of the four membership functions of the alternative fuzzy sets for NPI were set to be:

- *Low*: (0, 0, 2.6, 3.4)
- *Medium low*: (3, 3, 3.8)
- *Medium high*: (3.5, 3.4, 5.4)
- *High*: (4.4, 4.4, 10, 10)

The intersection points remain the same as for the initial membership functions.

B) Age

The parameters of the three membership functions of the alternative fuzzy sets for Age were set to be:

- *Young*: (0, 0, 35, 45)

- *Middle age*: (35, 45, 55, 65)
- *Old* : (55, 65, 90, 90)

The intersection points remain the same as for the initial membership functions.

C) VI (Vascular Invasion)

The parameters of the three membership functions of the alternative fuzzy sets for VI were set to be:

- *No*: (0, 1, 2)
- *Maybe*: (1, 2, 3)
- *Yes*: (2, 3, 4)

The intersection points remain the same as for the initial membership functions.

D) LN (Number of Positive Lymph Nodes)

The parameters of the two membership functions of the alternative fuzzy sets for LN were set to be:

- *Negative*: (0, 0, 0.2)
- *Positive*: (0, 0.2, 1, 1)

Hence, *Negative* now intersects *Positive* at 0.1, corresponding to a lymph node ratio of 1 in 10, rather than 1 in 36. We use '*NPI-F*' to represent the '*foundation*' in which only NPI uses the new membership functions and the remainder of the variables are as the original. The list of foundation names and their descriptions are shown in Table 6. Obviously there are many other alternative foundations that are available for selection. The use of automated tuning methodologies such as the ANFIS approach [2] below, or heuristic tuning methods such as genetic algorithms (e.g. [8]), simulated annealing (e.g. [33]), etc., will explore the range of alternative membership functions further, but all require the system to be trained against a specific reference set of cases.

Table 6: The notation used to describe each foundation system.

<i>Foundation Name</i>	<i>Description</i>
Original-F	Original foundation
NPI-F	Only NPI use the new membership function setting
Age-F	Only Age use the new membership function setting
VI-F	Only VI use the new membership function setting
LN-F	Only LN use the new membership function setting
NPI-Age-F	Only NPI and Age use the new membership function setting
NPI-VI-F	Only NPI and VI use the new membership function setting
NPI-LN-F	Only NPI and LN use the new membership function setting
Age-VI-F	Only Age and LN use the new membership function setting
Age-LN-F	Only VI and LN use the new membership function setting
VI-LN-F	Only VI and LN use the new membership function setting
Age-NPI-LN-F	Only LN <i>does not</i> use the new membership function setting
Age-VI-LN-F	Only VI <i>does not</i> use the new membership function setting
NPI-VI-LN-F	Only Age <i>does not</i> use the new membership function setting
Age-NPI-VI-F	Only NPI <i>does not</i> use the new membership function setting
Age-VI-LN-NPI-F	All four inputs use the new membership function setting

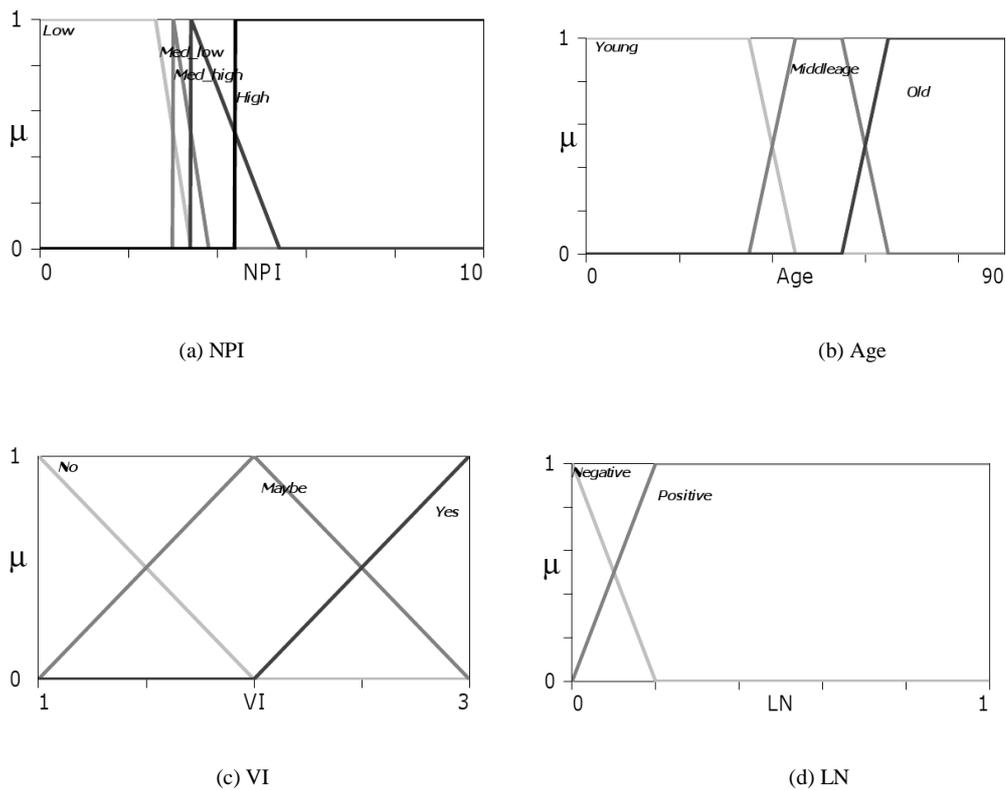


Figure 5: Alternative membership functions for the NPI, Age, VI and LN input variables

4.2.2. Alternative Variability

The purpose of this paper is to evaluate the effects of incorporating variability into a real-world decision making scenario, through the mechanism of the non-stationary fuzzy sets introduced by Garibaldi [24]. In his original paper, Garibaldi describes different types of non-stationarity (e.g. by shifting membership functions left-right or up-down), different forms of perturbation function (e.g. generating random, sinusoidal or chaotic variations) and different sizes (amounts) of variation. In this study, we restrict the perturbations investigated to shifts in membership function location (i.e. left-right shifts), using normally distributed random perturbations (with mean zero). Only the amount of variation, governed by the standard deviation (σ) of the normally distributed perturbations, was varied. As the standard deviation is relative to the universe of discourse of each variable, a standard deviation of (for example) 0.01 corresponds to a variation of 1/100th of the universe of discourse.

Once the standard deviation has been set, and the foundation fuzzy system has been selected, a complete iteration consists of randomly perturbing the membership functions of all variables and then running the resultant fuzzy inference system (termed an instantiation of the system) on all the 1310 cases in our database. The fuzzy system for each instantiation will therefore obtain an agreement percentage for the 1310 cases. The average of the agreement percentages from all the instantiations of each foundation can then also be calculated. The complete procedure is described in the pseudo code in Figure 6.

```

Require:  $Iter$ : number of iterations
Require:  $SD_{start}$ : initial standard deviation
Require:  $SD_{end}$ : final standard deviation
Require:  $F$ : current 'foundation'
Require:  $p_1, p_2, \dots, p_m$ : parameters of current mf
1: Set  $Iter, SD_{start}, SD_{end}$  and  $F$ :
2: for  $\sigma = SD_{start}$  to  $SD_{end}$  do:
3:   for  $t = 1$  to  $Iter$  do:
4:     for each input  $i = 1$  to 5 do:
5:        $k_i \leftarrow$  universe of discourse of input  $i$ 
6:        $n_{mf} \leftarrow$  number of  $mfs$  in input  $i$ 
7:       for  $mf = 1$  to  $n_{mf}$  do:
8:         Generate a number  $r \sim N(0, \sigma)$ .
9:          $p_i(t) \leftarrow p_i + k_i r$ 
10:      end for
11:    end for
12:    for each output do:
13:       $k_0 \leftarrow$  universe of discourse of output
14:       $n_{mf} \leftarrow$  number of  $mfs$  in output
15:      for  $mf = 1$  to  $n_{mf}$  do:
16:        Generate a number  $r \sim N(0, \sigma)$ .
17:         $p_i(t) \leftarrow p_i + k_i r$ 
18:      end for
19:    end for
20:    Apply the perturbed mfs to the fuzzy inference system  $F$ 
21:    Run  $F$  on each case and obtain output
22:    Calculate the agreement percentage
23:    Store results and record the best
24:  end for  $t$  loop
25: end for  $\sigma$  loop

```

Figure 6: Pseudo code of the proposed methodology for creating a non-stationary fuzzy inference system

4.3. Fuzzy Model Tuned by ANFIS

As a comparison, the ANFIS approach was also used to tune a fuzzy inference system based on the same clinical data. An ANFIS system constructs a Takagi-Sugeno type fuzzy inference model in practice. In contrast to Mamdani type fuzzy inference systems as described above, the consequent parts of Takagi-Sugeno fuzzy rules are local linear functions or constants. The overall output of the system is the three recommendations of Chemo: *No*, *Maybe* and *Yes*.

5. Results

The performances of the designed fuzzy decision support systems were evaluated in terms of the rates of agreements with the decisions made by clinicians. Each foundation system consists of 12 fuzzy rules as described in Table 4, but the fuzzy sets used in different foundation systems have different settings. The second column of Table 7 summarizes the results of each foundation system without perturbations. The level of agreement with clinician’s decision generally increases with the use of the alternative membership functions, and the best overall performance (without perturbations) was obtained by the NPI-Age-F system (i.e. the system featuring only alternative NPI and Age membership functions).

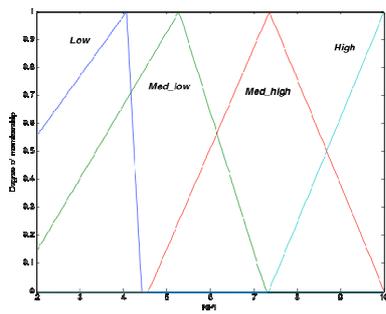
We then applied perturbations to each foundation system on a range of settings for σ over 1000 iterations. The value of σ was set to be 0.01, 0.02, 0.03, 0.04, 0.05 and 0.06 separately. The best agreements obtained for each foundation system have been identified in bold in Table 7. Through perturbations, the best performance of non-stationary fuzzy systems reached a mean of around 88.09% agreement in breast cancer treatment recommendations ($p < 0.001$, 95% CI: 87.96-88.21%). It can be seen from Table 7 that hardly any performance gain was achieved with small perturbations, particularly 0.01, and that the best performance was usually obtained at around 0.05 variation. While we do not believe this result can be generalized, it may support the idea that very small variations are unlikely to lead to any performance gains.

Table 7: The agreements obtained by the non-stationary fuzzy models for a range of σ vs those obtained by each foundation system without perturbations. The best result for each foundation is highlighted in bold.

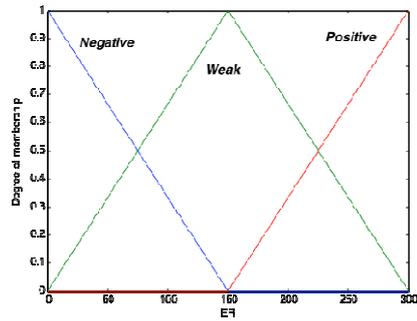
Foundation Name	σ in perturbation function						
	None	0.01	0.02	0.03	0.04	0.05	0.06
Original-F	84.58	85.42	87.25	87.94	88.02	88.09	87.86
NPI-F	84.12	85.57	87.25	87.56	87.71	87.71	88.02
Age-F	84.89	85.88	87.63	88.02	88.17	88.17	88.24
VI-F	83.28	85.27	87.71	87.86	87.79	87.94	87.86
LN-F	82.98	85.19	87.02	87.56	87.71	87.79	87.71
NPI-Age-F	85.50	85.95	87.56	88.09	87.94	88.09	87.86
NPI-VI-F	84.96	85.50	87.25	87.63	87.63	87.71	87.63
NPI-LN-F	84.89	85.50	87.10	87.79	87.79	87.56	88.17
Age-VI-F	84.73	85.95	87.56	88.17	88.09	88.02	88.02
Age-LN-F	84.66	85.73	87.48	87.86	88.24	88.24	88.17

VI-LN-F	84.20	85.27	87.48	87.79	87.56	87.86	88.02
Age-NPI-LN-F	85.27	85.95	88.02	88.02	88.24	88.09	88.09
Age-VI-LN-F	84.50	85.95	87.71	87.71	88.09	88.63	88.17
NPI-VI-LN-F	84.73	85.34	87.48	87.48	87.63	87.71	87.63
Age-NPI-VI-F	85.34	85.80	87.40	88.09	88.24	88.17	87.94
Age-VI-LN-NPI-F	85.11	85.88	87.71	88.02	88.09	88.02	88.09

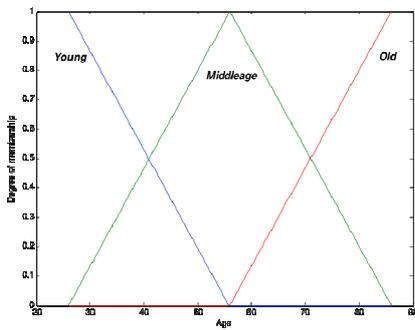
As a comparison, ANFIS was also used to tune a fuzzy inference system based on the same clinical data. From the whole data set, 800 cases were randomly chosen as training data and the remaining 510 cases as the testing data. The total number of parameters that need to be trained was 261 (including 216 linear and 45 nonlinear parameters) and the number of generated fuzzy rules was 216. The agreement rate of tuned fuzzy model applied to testing cases is 85.88%. Figure 7 shows the corresponding membership functions generated by the ANFIS. It can be seen that our constructed fuzzy inference system can achieve performance comparable to the ANFIS fuzzy model. Importantly, the membership functions generated by the ANFIS method suffer from poor interpretation, particularly the NPI and LN fuzzy sets whose membership functions could not be well interpreted according to the clinician knowledge as described in the clinical protocol. On the other hand, the ANFIS fuzzy model generated a rule-base with 216 rules (covering all the combinations of the terms of the input variables), these fuzzy rules do not match the clinical guidelines in practice. Considering the fact that our proposed fuzzy inference model only uses 12 fuzzy rules to make breast cancer treatment recommendation, our proposed model achieves much better parsimony in model construction and better model interpretability than the ANFIS model in making breast cancer treatment recommendations.



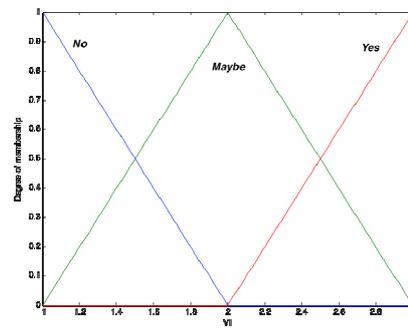
(a) NPI



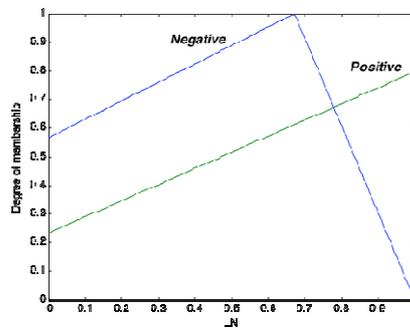
(b) ER



(c) Age



(d) VI



(e) LN

Figure 7: Tuned input membership functions from ANFIS

In terms of the detailed decision making, we examined the confusion matrices for the original foundation system and a representative of the mean best performing of the non-stationary systems. The confusion matrices for the original foundation (1108 agreements, 84.58%) and the mean best

non-stationary system (1154 agreements, 88.09%) are given in Table 8. It can be seen that the best non-stationary system increases the overall number of agreements while increasing both true negatives and true positives.

Table 8: The confusion matrices corresponding to the original system and a representative result from the best non-stationary systems.

		<i>Clinical Decision</i>		
		<i>No</i>	<i>Maybe</i>	<i>Yes</i>
<i>Fuzzy System Decision</i>	<i>No</i>	982	35	64
	<i>Maybe</i>	2	2	3
	<i>Yes</i>	83	15	124

(a) The original system confusion matrix

		<i>Clinical Decision</i>		
		<i>No</i>	<i>Maybe</i>	<i>Yes</i>
<i>Fuzzy System Decision</i>	<i>No</i>	1011	30	49
	<i>Maybe</i>	3	4	3
	<i>Yes</i>	53	18	139

(b) The best non-stationary confusion matrix.

6. Discussion

Making predictions in individuals under uncertainty has become a critical activity in healthcare [9]. From the perspective of biomedical informatics, an important way of assisting healthcare providers in making decisions under uncertainty is to develop for them computer manageable decision models with good interpretation and predictive performance. Fuzzy models constructed by a set of easily understandable *if-then* rules have become a popular choice for incorporating human knowledge into problem solving process under uncertainty. In this paper, we have shown how we have taken into account a clinical protocol and successfully represented the decision making under uncertainty within a fuzzy inference system. The initial linguistic terms expressed by fuzzy sets were used with a standard fuzzy system to capture the essential knowledge encoded within a clinical protocol and associated clinical knowledge. This system performed well to make breast cancer treatment recommendations with an overall agreement to clinical practice of 84.58%. By interactively tuning a range of alternative membership functions which retained the essential crisp clinical thresholds, the performance of the system could be increased further to 85.50%.

As part of ongoing research into incorporating variability into fuzzy inference, in order to permit ensemble decision making, results in this paper demonstrate that introducing random left-right shifts of the membership functions (alterations in the location of the membership functions) can significantly increase the system performance further to around 88.09% (95% CI: 87.96-88.21%, $p < 0.001$). In this way, it may be possible that an effective fuzzy inference system can be designed without any tuning, simply by creating a reasonable foundation system, applying a small variation and by creating effective ensemble decision making.

The use of fuzzy inference systems to model expert decision making (in clinical or other contexts) is a pragmatic approach intended to simulate the results of human reasoning. It does not mean that the modeler is making any assumption that fuzzy inference is somehow taking place inside the expert's brain. Similarly, using randomly generated non-stationary fuzzy sets to model expert variability does not imply any assumption on the origin of such variability. The non-stationary approach adopted in this paper models two situations simultaneously: (i) the inter-expert variability exhibited in any panel of experts, and (ii) the intra-expert variability exhibited in any single expert (that is not attributable to learning). While these two underlying causes of variability are obviously completely different, they are both implicitly modeled in our approach. As an aside, we note that the permanent alteration of fuzzy membership functions, such as is the case in a 'conventional' ANFIS approach (see Section 3.4), is more akin to the process of learning, which may be defined in a Pavlovian sense as a permanent change in stimulus-response (input-output) mapping.

In summary, based on our research in this paper, a trustworthy procedure for designing an effective non-stationary fuzzy system for decision recommendation can include the following steps:

- 1) create a reasonable (acceptable) static fuzzy inference system;
- 2) incorporate variability into the membership functions to obtain alternative decisions (tentatively, variability of around 5% variation in location may be suitable);

- 3) run the inference system multiple times to obtain a range of alternative acceptable decisions; and
- 4) employ ensemble techniques to recommend the set of the best from the range of alternatives.

This is a new approach to fuzzy inference system design, complementary to other standard approaches. While not necessarily any better than any other approach, our method may be generally applicable to a wide range of applications.

7. Conclusions and Future Work

In this paper, we have presented a methodology of designing effective non-stationary fuzzy inference systems for breast cancer treatment recommendations. Our research demonstrated that incorporating variability into a conventional fuzzy inference system can enhance the decision making of that system.

Further study of representing intra- and/or inter- expert variation within fuzzy inference systems is being conducted. The decisions by a number of clinical experts on a subset of the same group of cases are currently being collected over a sequence of time. This will allow the variability introduced into the fuzzy inference system described here to be compared directly to the variability observed in real clinical experts. For the decisions made from a group of medical experts, a consensus model will be generated so that the final decision can be most representative of the opinions of the multidisciplinary panel. To this end, a new type of operator that is able to aggregate the decisions modeled by fuzzy sets with variability into an overall one is needed. The type-1 OWA operators and type-2 OWA operators [31][32][33] could benefit this research.

8. Acknowledgment

This work has been supported by the EPSRC Research Grant EP/C542215/1 and EP/C542207/1.

9. References

- [1] Limberg J, Seising R. Fuzzy set theory and the philosophical foundations of medicine, R. Seising (Ed.): *Views on Fuzzy Sets and Systems*, 2009; *STUDFUZZ* 243: 321–344, Springer-Verlag Berlin Heidelberg.
- [2] Jang JSR. ANFIS: adaptive-network-based fuzzy inference system. *IEEE T Syst Man Cyb* 1993; 23 (3): 665~685.
- [3] Zhou SM, Gan JQ. Low-level interpretability and high-level interpretability: a unified view of interpretable fuzzy system modelling from data. *Fuzzy Set Syst* 2008; 159(23): 3091-3131.
- [4] Mendel JM. *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Prentice-Hall, Upper Saddle River, NJ, 2001.
- [5] Beliakov G, Warren J. Fuzzy logic for decision support in chronic care. *Artif Intell Med* 2001; 21 (1-3): 209~213.
- [6] Zhou SM, Gan JQ. Constructing L2-SVM-based fuzzy classifiers in high-dimensional space with automatic model selection and fuzzy rule ranking. *IEEE T Fuzzy Syst* 2007; 15(3): 398~409.
- [7] Pena-Reyes CA, Sipper M. A fuzzy-genetic approach to breast cancer diagnosis. *Artif Intell Med* 1999; 17(2): 131~155.
- [8] Chan W, Naghdy F. Prognosis of body fluid level by fuzzy logic technique. *Methods Inf Med* 2001; 40(1): 52–58.
- [9] Visweswaran S, Angus DC, Hsieh M. et al, Learning patient-specific predictive models from clinical data, *J Biomed Inform* 2010; 43(5): 669–685.
- [10] Cooper GF, Aliferis CF, Ambrosino R. et al. An evaluation of machine-learning methods

- for predicting pneumonia mortality, *Artif Intell Med* 1997; 9(2): 107–138.
- [11] Garibaldi JM, Ifeachor EC. Application of simulated annealing fuzzy model tuning to umbilical cord acid-base interpretation. *IEEE T Fuzzy Syst* 1999; 7(1): 72-84.
- [12] Lee KM, Kwak DH., Lee-Kwang H. Fuzzy inference neural network for fuzzy model tuning. *Ieee T Syst Man Cyb* 1996; 26 (4):637~645.
- [13] Amini M, Mueller K, Abbaspour KC, et al. Statistical modeling of global geogenic fluoride contamination in groundwaters. *Env Sci & Tech* 2008; 42(10): 3662~3668.
- [14] Caydas U, Hascalik A, Ekici S. An adaptive neuro-fuzzy inference system (ANFIS) model for Wire-EDM. *Expert Syst Appl* 2009; 36(3): 6135~6139.
- [15] Avci E, Turkoglu I. An intelligent diagnosis system based on principle component analysis and ANFIS for the heart valve diseases. *Expert Syst Appl* 2009; 36(2): 2873~2878.
- [16] Riverol C, Sanctis CD. A fuzzy filter for improving the quality of the signal in adaptive-network-based fuzzy inference systems (ANFIS). *Appl Soft Comput* 2009; 9(1): 305~307.
- [17] Zhou SM, Gan JQ. Extracting Takagi-Sugeno fuzzy rules with interpretable submodels via regularization of linguistic modifiers, *IEEE T Knowl Data En* 2009; 21(8): 1191-1204.
- [18] Cancer Research UK, CancerStats Key Facts on Breast Cancer (December 2008). URL [http, pp://info.cancerresearchuk.org/cancerstats](http://pp://info.cancerresearchuk.org/cancerstats)
- [19] Torosian MH, *Breast Cancer: A Guide to Detection and Multidisciplinary Therapy*, Humana Press, 2002.
- [20] Houssami N, Sainsbury R. Breast cancer: Multidisciplinary care and clinical outcomes. *Eur J Cancer* 2006; 42 (15):2480~2491.
- [21] Whelan JM, Griffith CD, Archer T. Breast cancer multi-disciplinary team in England: much achieved but still more to be done. *The Breast* 2006; 15(1):.119~122.
- [22] Jerez-Aragones JM, Pelaez Sanchez JI, Dona JEA.. A neuro-fuzzy decision model for prognosis of breast cancer relapse. in: R. Conejo, M. Urretavizcaya, J. Páerez de la Cruz (Eds.), *Current Topics in Artificial Intelligence*, Lecture Notes in Computer Science,

- Springer, Berlin / Heidelberg 2003; 3040: 638~645.
- [23] Nabil E, Badr A, Farag I, et al. A hybrid artificial immune genetic algorithm with fuzzy rules for breast cancer diagnosis. *Proc of the 6th Int Conf on Comp and Inform (INFOS)*, Cairo, Egypt, 2008:HBI31~HBI39.
- [24] Garibaldi JM, Ozen T. Uncertain fuzzy reasoning: A case study in modelling expert decision, *IEEE T FUZZY SYST* 2007; 15(1): 16~30.
- [25] Garibaldi JM, Jaroszewski M, Musikasuwana S. Non-stationary fuzzy sets. *Ieee T Fuzzy Syst* 2008; 16 (4): 1072~1086.
- [26] Garibaldi JM, Musikasuwana S, Ozen T. The association between non-stationary and interval type-2 fuzzy sets: A case study. *Proc of the IEEE Int Conf on Fuzzy Sys (FUZZ-IEEE 2005)*, Reno, USA, 2005: 224~229.
- [27] Garibaldi JM, Jaroszewski M, Musikasuwana S. New concepts related to non-stationary fuzzy sets. *Proc of the IEEE Int Conf on Fuzzy Sys (FUZZ-IEEE)*, London 2007:1684~1689.
- [28] Haybittle JL, Blamey RW, Elston CW, et al. A prognostic index in primary breast cancer. *Br J cancer* 1982; 45(3):361-366.
- [29] National Cancer Institute, online dictionary of cancer terms (2009). URL <http://www.cancer.gov/dictionary>
- [30] Pathology Reporting of Breast Disease, A Joint Document Incorporating the Third Edition of the NHS Breast Screening Programme's Guidelines for Pathology Reporting in Breast Cancer Screening and the Second Edition of The Royal College of Pathologists' Minimum Dataset for Breast Cancer Histopathology, *NHSBSP Publication* No 58, January 2005.
- [31] Zhou SM, Chiclana F, John RI, Garibaldi JM. Type-1 OWA operators for aggregating uncertain information with uncertain weights induced by type-2 linguistic quantifiers. *Fuzzy Set Syst* 2008; 159(24): 3281-3296.
- [32] Zhou SM, John RI, Chiclana F, Garibaldi JM. On aggregating type-2 fuzzy sets by type-2 OWA operators for soft decision making. *Int J Intell Syst* 2010; 25(6): 540-558.

- [33] Zhou SM, Chiclana F, John RI, Garibaldi JM. Alpha-level aggregation: a practical approach to type-1 OWA operation for aggregating uncertain information with applications to breast cancer treatments. *IEEE T Knowl Data En* 2011; 23(10): 1455 – 1468.
- [34] Gort M, Broekhuis M, Otter R et al, Improvement of best practice in early breast cancer: actionable surgeon and hospital factors. *Breast Cancer Res Tr* 2007; 102: 219-226.
- [35] Bueno-de-Mesquita JM, Nuyten DS, Wesseling J et al, The impact of inter-observer variation in pathological assessment of node-negative breast cancer on clinical risk assessment and patient selection for adjuvant systemic treatment. *Ann Oncol* 2010; 21: 40-47.