

# Adaptive Neuro-Fuzzy Inference System (ANFIS) in Modelling Breast Cancer Survival

Hazlina Hamdan and Jonathan M. Garibaldi, *Member, IEEE*

**Abstract**—Medical prognosis is the prediction of the future course and outcome of a disease and an indication of the likelihood of recovery from that disease. Soft-computing approaches including artificial neural networks and fuzzy inference have been used widely to model expert behaviour. In this paper, we propose the use of an adaptive fuzzy inference system (ANFIS) technique in the estimation of survival prediction. This paper describes the methodology by which ANFIS was used to model survival and presents a comparison of this new method with existing methods in the capability to predict the survival rate in a given medical data set concerning survival of patients following operative surgery for breast cancer.

## I. INTRODUCTION

Breast cancer is one of the most common cancers to afflict the female population. It is estimated that one in nine women in the UK will develop breast cancer at some point in their life [1]. Breast cancer is a malignant tumour that develops from uncontrolled growth of cells in the breast. A malignant tumour is composed of cells that invade or spread to other parts of the body. The exact cause of the breast cancer is not really known, but is most likely to be a combination of genetic and environmental factors. However, in general, earlier diagnosis and treatment should increase the survival rates, as the disease is much easier to control if it has not spread to other parts of the body.

Medical prognosis is the estimation of cure, complication, disease recurrence or survival for a patient or group of patients after treatment [2]. Prognosis is the principal factor in determining the treatment that will immediately follow the diagnosis of the disease. It is important because it is used to guide the type and intensity of the medication administered to patients.

Survival analysis describes the analysis of data that corresponds to the time from when an individual enters a study until the occurrence of some particular event or end-point. It is concerned with the comparison of survival curves for different combinations of risk factors. Analysis of survival data is complicated by the presence of censorship (patients leaving the study). Statistical methods are commonly used in the analysis of survival data and lately artificial intelligence techniques have been considered as alternative methods.

In medical contexts, artificial neural networks (ANNs) have been the subject of great interest, following the discovery of the backpropagation algorithm, and recently have

become very popular in medical survival prediction [3], [4], [5]. Neural networks work by detecting patterns in data, learning from the relationships and adapting to them. This knowledge is then used to predict the outcome for new combinations of data.

Fuzzy inference is a process of mapping from a given input to an output using the theory of fuzzy sets. Knowledge is encoded as a set of explicit linguistic rules, which can be easily understood by people without technical expertise. In addition, the exploration on the use of fuzzy approach has rarely been used to date in cancer prognosis.

This paper presents a hybrid methodology which combines the advantages of ANNs and fuzzy inference in modelling survival. By adapting the method proposed by [6] known as the *partial logistic ANN* (PLANN) model that can estimate the smooth discrete hazard rate, we propose to use an adaptive fuzzy inference system (ANFIS) in predicting the hazard curve and survival curve of breast cancer patients. A specific form of data pre-processing has to be performed before a standard ANFIS model can be used for prognostic prediction of survival.

## II. BACKGROUND

### A. Survival Analysis

In the study of cancer and other diseases, it is important to measure the time between response to treatment and recurrence or disease-free survival time, rather than just time to death [7]. The recording of what the event is and when the period of observation starts and finish is necessary in the analysis of survival. An individual with cancer cannot all be observed for the same length of time, because some individual are diagnosed at the beginning of the period under study, some near the end and others may be diagnosed at any time in the study. Basically, survival data contains uncensored and censored observations. Uncensored observations involved patients who are observed until they reach the end of the study. Censored observations on the other hand, involve only patients who survive beyond the end or who are lost to follow-up at some point.

There are two reasons for modelling survival data: the first reason is to determine which combination of potential explanatory variables affect the form of hazard function and the second reason is to estimate the hazard function for an individual in addition to their survival function [8].

The survival function is defined as the probability that an individual survives at least up to a certain time  $t$ , where  $T$  denotes a positive random variable associated with the

Hazlina Hamdan is with the Intelligent Modelling and Analysis (IMA) Research Group, School of Computer Science, University of Nottingham, UK (phone: +44 115 9514229; email: hzh@cs.nott.ac.uk).

Jonathan M. Garibaldi is with the Intelligent Modelling and Analysis (IMA) Research Group, School of Computer Science, University of Nottingham, UK (phone: +44 115 9514216; email: jmg@cs.nott.ac.uk).

survival time, represented as [6]:

$$S(t) = P(T > t) \quad (1)$$

On the other hand, the hazard function, also known as conditional failure probability, is the probability an individual will die at a certain time  $t$  (conditioned on survival up to that time) and so denotes the instantaneous death rate. It can be shown in this form:

$$h_l = P(T \in A_l | T > t_{l-1}) = \left( \frac{S(t_{l-1}) - S(t_l)}{S(t_{l-1})} \right) \quad (2)$$

where the time interval  $l = 1, 2, \dots, L$  forms disjoint intervals  $A_l = (t_{l-1}, t_l]$ .

The survival and hazard functions are related to each other, in that the estimation of survival function can be written as:

$$S(t) = \prod_{l:t_l \leq t} (1 - h_l) \quad (3)$$

Statistical methods, such as the Kaplan-Meier estimate, are usually used to explain the data and to model the disease progression with the ability to handle censored data. The Kaplan-Meier method, also called the Product Limit Estimator, is a model where a clear interval is created for each occurrence of death. It involves computing the number of people who died at a certain time, divided by the number of people who were still in the study at that time and multiply these probabilities by any earlier computed probabilities.

Lately, the potential uses of various Artificial Intelligence (AI) techniques in medical has become popular — the motivation for this work was to investigate whether the use of AI techniques in predictive survival can improve upon the predictive power of proportional hazards in breast cancer as in other forms of cancer. A wide range of AI techniques such as neural networks, fuzzy systems, decision trees, genetic algorithms, support vector machines and various combinations of intelligent technologies have been used in modelling medical prognosis. Comparisons of the predictive capability between the AI techniques have been performed by [9], [10], [11], [12] to examine the potential of each technique in terms of accuracy, sensitivity and specificity. Comparisons also have been made between AI techniques and statistical method [13], [14], [15], [16], [17].

### B. Partial Logistic Artificial Neural Network (PLANN)

An artificial neural network (ANN) is defined as an information processing system inspired by the structure of the human brain [18]. In the medical field, ANNs have been used since the late 1980s, initially to identify accuracy of survival prediction, the time to recurrence of disease after surgery or the probability of disease free survival, for example [19], [20], [16], [21], [13].

Multi-layer perceptrons that use backpropagation training are most commonly used [5], [4], [3]. Backpropagation networks are networks where signals travel in one direction from input neuron to an output neuron without returning to

its source. Backpropagation networks consist of at least three layers of units: an input layer, at least one hidden layer and an output layer. The output from the input layer is connected as an input into the hidden layer. Similarly, the output from the hidden layer is connected as an input into the output layer to produce the final output of the ANN.

The transfer function is a function used to transform the activation level of a neuron into an output signal. The behaviour of the ANN depends on both the weights and the activation function that is specified for the neuron. More importantly, each layer has its own transfer function. Some examples of common activation functions are the step function, sigmoid function and linear function.

Special organisation of the ANN structure and the data is required in order for an ANN to be used for modelling survival. An approach in which the time interval is included as one of the inputs to a standard back-propagation network has been proposed by Biganzoli et al. to estimate smooth discrete hazards. The technique, known as the *partial logistic artificial neural network* (PLANN) model, is thoroughly described in [6]. Adapting the standard feed forward neural network, the PLANN architecture, as shown in Figure 1, contains three layers: the input layer combining the multiple explanatory variables with the survival time, one hidden layer and a single output layer, interpreted as the conditional probability of failure for the given time. The relationship of the layers can be written as:

$$y_k(x_i, w) = f_o(\alpha_k + \sum_{h=1}^H w_{hk} f_h(\alpha_h + \sum_{j=1}^J w_{jh} x_{ij})) \quad (4)$$

The logistic function is used as activation function  $f$  for both the hidden nodes and the single output node.  $w_{jh}$  denotes the weight for the connection between input and the hidden units and  $w_{hk}$  are the weights of the connections between the hidden and output units.  $\alpha_h$  and  $\alpha_k$  are the weights of the connections of the bias unit with the hidden and output units, respectively.

In order for the output of the ANN to be interpreted as conditional probabilities of failure, requires a likelihood term that reflects the status of each patient  $x_i$  at each time interval. The negative logarithm of the likelihood or the error function is expressed by:

$$E = - \sum_{i=1}^n \sum_{l=1}^{l_i} \{d_{il} \log h_l(x_i, a_l) + (1-d_{il}) \log [1-h_l(x_i, a_l)]\} \quad (5)$$

where  $d_{il}$  denotes the event attribute and  $a_l$  is time when the patient was last observed.

Briefly, the implementation of the PLANN model requires the following process:

- Choose the relevant variables plus the survival time as the input of the network.
- Transform the variables into suitable inputs, depending on the variable type.

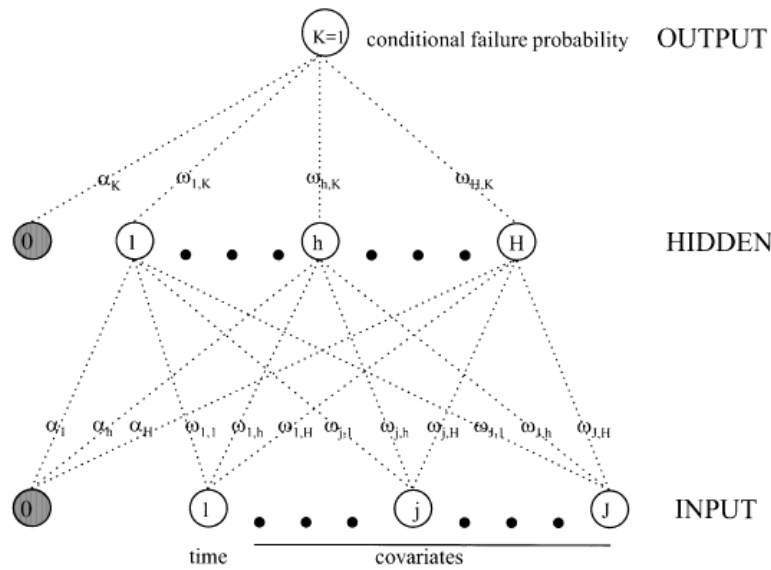


Fig. 1. PLANN architecture [6]

- Replicate all the inputs until the maximum survival time.
- Create the ANN structure by setting the hidden nodes and weight decay.
- Train the replicated inputs where the output of the network is the estimation of hazard rate or conditional failure probability.
- Predict the hazard rate for each time interval as the test data.
- Calculate the estimation of survival for each time interval using equation (3).

See section III and IV for more details of each step.

### C. Adaptive Neuro-Fuzzy Inference System (ANFIS)

Fuzzy logic was introduced by Zadeh in 1965 [22] to represent and manipulate data and information in which there are various forms of uncertainty. Fuzzy rule-based systems use linguistic variables to reason using a series of logical rules that contain IF-THEN rules which connect antecedent(s) and consequent(s), respectively. An antecedent is a fuzzy clause with certain degree of membership (between 0 and 1). Fuzzy rules can have multiple antecedents connected with AND or OR operators, where all parts are calculated simultaneously and resolved into a single number. Consequents can also be comprised of multiple parts, which are then aggregated into a single output of a fuzzy set [18].

Fuzzy inference is a process of mapping from a given input to an output using the fuzzy set methods. Figure 2 shows an example of a fuzzy inference system with five functional blocks. The fuzzification component transforms each crisp input variable into a membership grade based on the membership functions defined. The inference engine then conducts the fuzzy reasoning process by applying the appropriate fuzzy operators in order to obtain the fuzzy set to be accumulated in the output variable. The defuzzifier transforms the fuzzy output into a crisp output by applying a specific defuzzification method.

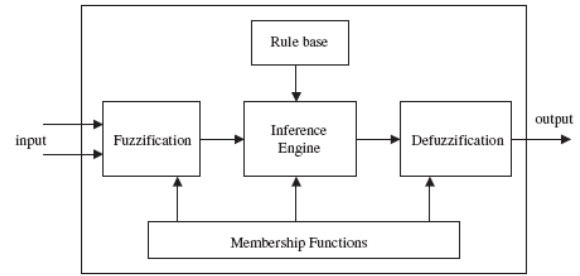


Fig. 2. Fuzzy Inference System

The adaptive neuro-fuzzy inference system (ANFIS) proposed by Jang in 1993 [23], implements a Sugeno fuzzy inference method. The ANFIS architecture contains a six-layer feed-forward neural network as shown in Figure 3. Layer 1 is the input layer that passes external crisp signals to Layer 2, known as the fuzzification layer, to determine the membership grades for each input implemented by the given fuzzy membership function, for example the bell-shaped (6) or gaussian curve (7) membership function is given by:

$$\mu_{A_i}(x) = \frac{1}{1 + \left[\left(\frac{x-c_i}{a_i}\right)^2\right]^{b_i}} \quad (6)$$

$$\mu_{A_i}(x) = \exp\left\{-\left(\frac{x-c_i}{a_i}\right)^2\right\} \quad (7)$$

where  $\{a_i, b_i, c_i\}$  are the parameters of the membership function. Layer 3 of ANFIS is the rule layer, which calculates the firing strength of the rule as the product of the membership grades. Layer 4 is called the 'normalised firing strengths', in which each neuron in the layer receives inputs from all neurons in Layer 3, and calculates the ratio of the firing strength of a given rule to the sum of firing strengths of all rules. Layer 5 is the defuzzification layer that yields the

TABLE I  
CATEGORY OF NPI SCORE

| NPI score           | Category | Comment                    |
|---------------------|----------|----------------------------|
| Less than 3.41      | Good     | more chance of survival    |
| Between 3.41 to 5.4 | Moderate | medium chances of survival |
| Over 5.4            | Poor     | less chances of survival   |

parameters of the consequent part of the rule. A single node in Layer 6 calculates the overall output as the summation of all incoming signals. Full details of the ANFIS process can be found in [23], [18].

ANFIS training can use alternative algorithms to reduce the error of the training. A combination of the gradient descent algorithm and a least squares algorithm is used for an effective search for the optimal parameters. The main benefit of such a hybrid approach is that it converges much faster, since it reduces the search space dimensions of the backpropagation method used in neural networks [23]. In the medical context, fuzzy approaches have been used in many areas, including in the prediction of patients' survival rate [11], [12] and for relapse probability [24], [25].

### III. DATA AND METHODS

#### A. Data

A set of breast cancer data collected by the *Breast Cancer Pathology Research Group* in the University of Nottingham were used to model the survival curve using the ANFIS model. In this study, 958 patients are assigned into prognostic groups using a prognostic index. The 'Nottingham Prognostic Index' (NPI) has been widely accepted in clinical practice to categorise patients into high (78%), intermediate (50%) or low (20%) risk groups. This index is based on pathological size, grade of tumor and the number of axillary nodes effected are identified significant in the prediction of survival [26]. The formula for the NPI is:

$$\text{NPI} = 0.2 * \text{pathological tumor size(cm)} + \text{lymph node stage} + \text{histological grade}$$

where, the categories of NPI score as shown in Table I.

In this study, three groups of NPI were compared, namely good, moderate and poor (represented as 1, 2, and 3, respectively) with an observation time of 120 months (ten year period).

#### B. Pre-processing

Pre-processing is a process that converts the raw inputs and outputs (target values) into a form understandable or acceptable before the training process. Often, this is used to reduce the dimensionality of input data and to optimise the generalization performance [27].

In order to be used within the PLANN model, if the variable is of categorical type, it must be transformed into an indicator variables or a 1-of- $n$  coding as shown in Table II, while continuous variables must be transformed by rescaling into the range  $[0, 1]$  or  $[-1, 1]$ . In addition,

TABLE II  
PRE-PROCESSING OF CATEGORICAL VARIABLES

| NPI category |   |   |   |
|--------------|---|---|---|
| Good         | 1 | 0 | 0 |
| Moderate     | 0 | 1 | 0 |
| Poor         | 0 | 0 | 1 |

TABLE III  
EXAMPLE OF THREE PATIENTS DATA SETS

|           | Time interval | NPI category | Event |
|-----------|---------------|--------------|-------|
| Patient 1 | 3             | 2            | 1     |
| Patient 2 | 5             | 1            | 0     |
| Patient 3 | 2             | 3            | 1     |

the implementation of weight decay only makes sense if the inputs range from 0 to 1, in order to be comparable with the output of hidden units [28]. In this study, the NPI group and time interval were used for the inputs. After the NPI score have been identified into the three categorical groups, it will be transformed into an indicator variable. The time interval is in months, with a time observation until 120 months.

In both models used in this study (PLANN and ANFIS), for training purposes, each patient is replicated for all the intervals in which the patient is observed, using the event indicator as the target. The input of the network (survival time and NPI groups) is replicated into  $t$  times which is the maximum survival time of an individual patient. The event attribute as a target of the network is also replicated and assigned as zero until the last time value is reached, where the event is 1 for occurrence and zero for censored. An example of replication is shown in Table III which shows the original data of three patients and Table IV which shows the replicated data suitable for input into the PLANN and ANFIS model.

### IV. EXPERIMENTAL RESULTS AND DISCUSSION

Data from 958 breast cancer patients were subjected to the pre-processing described above before being passed to the

TABLE IV  
REPLICATION FOR ALL INTERVALS OBSERVED FOR EACH PATIENT

|           | Time interval | NPI category | Event |
|-----------|---------------|--------------|-------|
| Patient 1 | 1             | 2            | 0     |
|           | 2             | 2            | 0     |
|           | 3             | 2            | 1     |
| Patient 2 | 1             | 1            | 0     |
|           | 2             | 1            | 0     |
|           | 3             | 1            | 0     |
|           | 4             | 1            | 0     |
|           | 5             | 1            | 0     |
| Patient 3 | 1             | 3            | 0     |
|           | 2             | 3            | 1     |

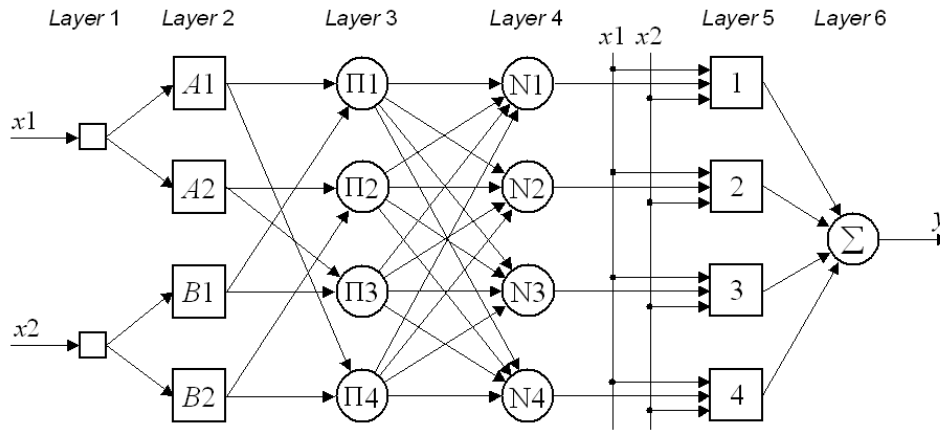


Fig. 3. Adaptive Neuro-Fuzzy Inference System (ANFIS) [18]

training process. This section presents the result of two AI models, the existing PLANN model and our ANFIS model, in modelling survival rate and a comparison of two models is made according to the Kaplan-Meier method.

For the PLANN model, several neural network configurations were tested with the number of hidden nodes ranging from 2 to 15 and weight decay value of 0.025, 0.05, 0.075 and 0.1. Biganzoli [6] outlines how the network information criteria (NIC) can be used to identify the best model from different network configurations. Twelve hidden nodes and 0.075 weight decay were found to give a smooth shape of the conditional hazard function such that it appears neither underfitted nor overfitted. Figure 4 shows the smooth hazard function, while Figure 5 shows an example of a hazard curve obtained with lower degree of smoothing indicating a tendency to overfit the empirical data.

For the ANFIS model, initial parameters of the fuzzy inference system have to be established before the training process commences. The number of membership functions assigned to each input was chosen empirically by trial and

error, and the performance measure (the root mean square error (RMSE)) was used to ensure the model did not cause overfitting. Gaussians were used for the membership functions and constants were used for the rule outputs (a zeroth-order Sugeno model). Hybrid learning, the combination of gradient descent and least squares algorithm, was selected as the learning algorithm. Several ANFIS model were configured with different numbers of membership functions for the survival time, ranging from 3 to 7, and between 10 and 200 epochs were used in the training process.

Observations were made as to the effect that different ANFIS model configurations had on the conditional hazard function curve. It is necessary to obtain a smooth curve with non-negative values in any of the time intervals. Figure 6 shows an example of a hazard curve that has negative values at the beginning of the time interval, while Figure 7 shows a negative hazard curve at the end of the interval. Neither of these are permitted. Figure 8 shows a lower degree of smoothing of the conditional hazard function.

Three membership functions for each input were finally

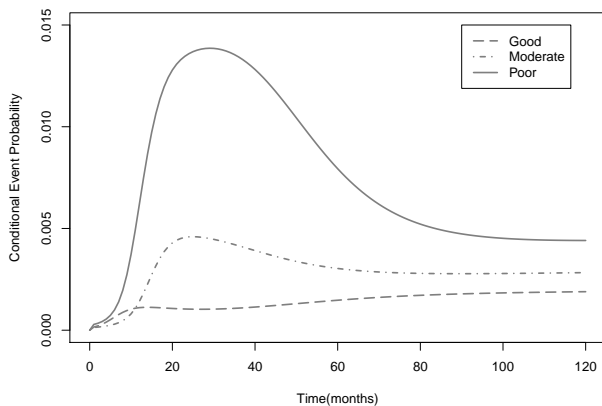


Fig. 4. Conditional Event Probability (hidden=12 and weight decay=0.075)

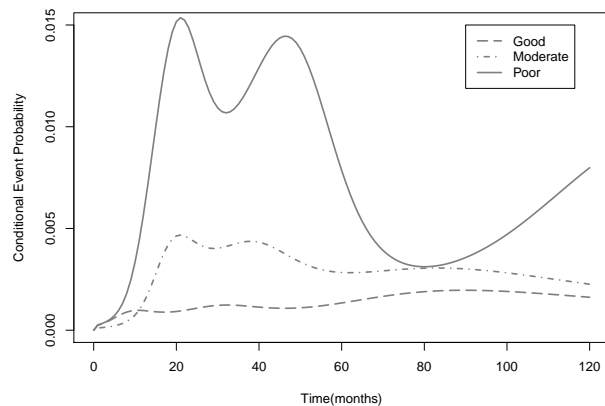


Fig. 5. Conditional Event Probability (hidden=12 and weight decay=0.05)

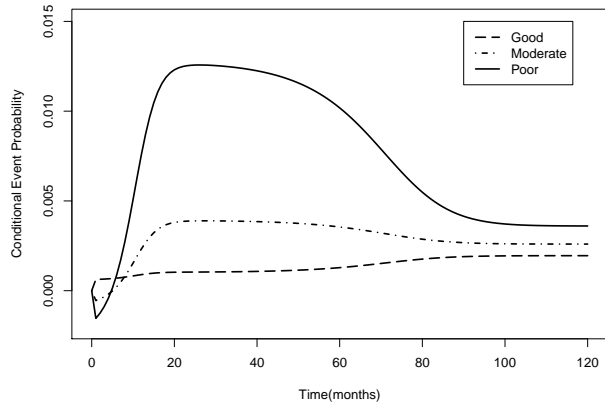


Fig. 6. Conditional Event Probability (mf=[3 3] and epochs=50)

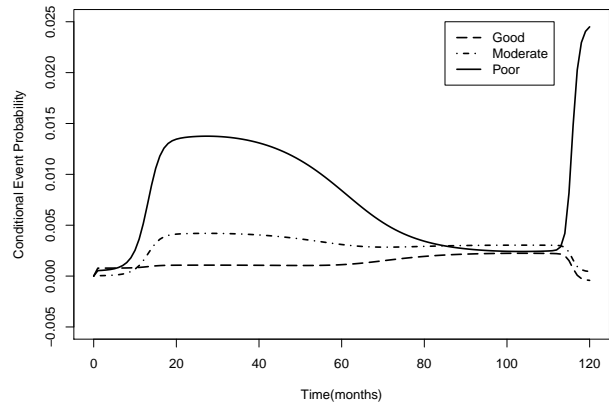


Fig. 7. Conditional Event Probability (mf=[5 3] and epochs=150)

selected as it was observed that these provided a smooth conditional hazard function. Figure 9 shows the initial membership functions for both inputs (the survival time and NPI group). The parameters for training options have been selected empirically as follows; 150 epochs were used to stop the training process and the step size was 0.1. Figure 10 shows the final membership functions obtained for both inputs after 150 epochs of training.

The testing data of 120 month observation for each of the three NPI groups are presented. The output of the testing is the estimation of conditional failure probability for each time interval (i.e. the hazard function). From this, the estimation of survival function using equation (3) can be plotted. Figure 11 shows the estimates of conditional failure probability of three categories of NPI after applying the PLANN (grey line) and ANFIS (black line) model for 120 months. Figure 12 shows the estimates of survival function of both the PLANN and ANFIS models against the Kaplan-Meier plot for the original (observed) data. It can be seen that, as desired, the curves obtained from the ANFIS model are close to those of the PLANN model. Given that the ANFIS model features explicit linguistic terms in its variables, we hope that this will lead to greater model interpretability.

## V. CONCLUSION

The PLANN and ANFIS models have been applied to the Nottingham Breast Cancer data set in order to obtain the conditional failure probability and the survival curve. Pre-processing of the raw data had to be carried out before the model can be used to get a smooth estimation of discrete hazard rate. Both models have demonstrated their predictive power in producing proportional hazard rates, and thus has confirmed their suitability as alternative method(s) in modelling survival in the presence of censorship.

Given the representation of fuzzy inference systems, in which knowledge is encoded as a set of explicit linguistic rules that can be easily understood by people without technical expertise, it is hoped that this will allow the incorporation

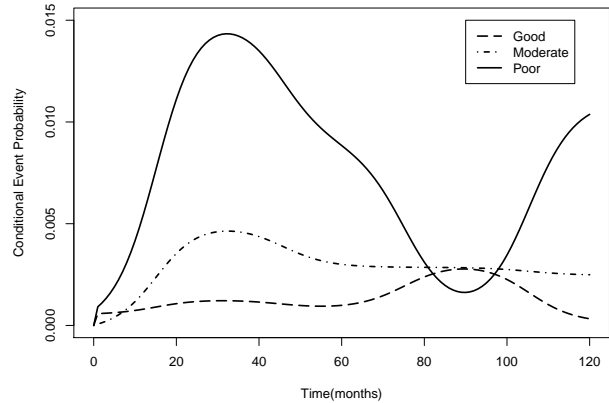


Fig. 8. Conditional Event Probability (mf=[5 3] and epochs=10)

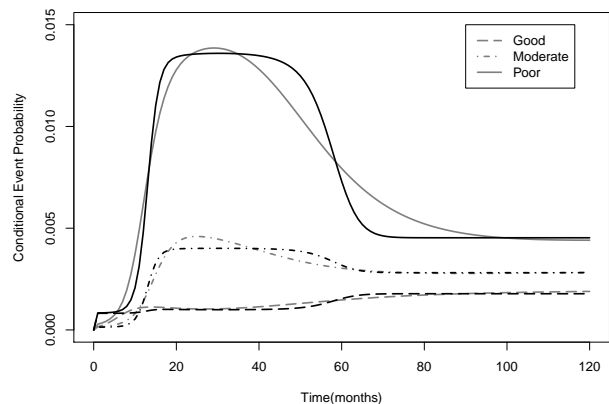


Fig. 11. Estimates of conditional failure probability of NPI groups with the best ANFIS model (black lines) and PLANN model (grey lines)

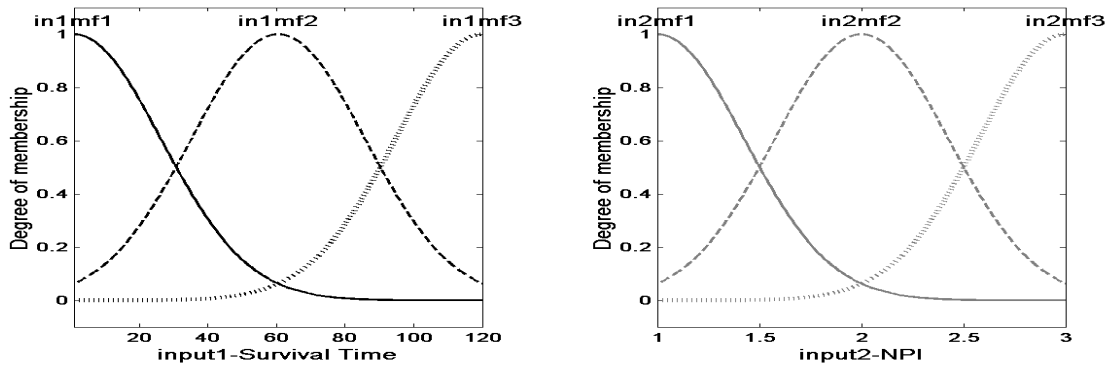


Fig. 9. Membership functions before training

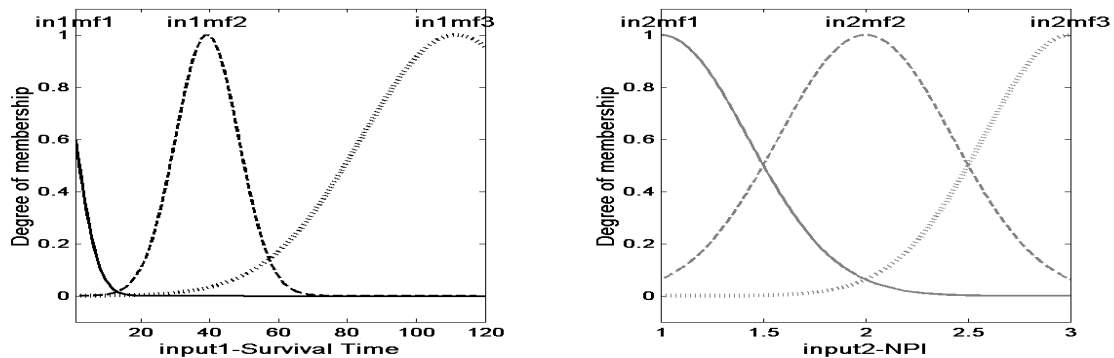


Fig. 10. Membership functions after training

of expertise from clinicians into the selection of inputs and the modelling of rules. Thus, it is hoped that such a technique may better address real clinical needs.

## VI. FUTURE WORK

In the future, we aim to repeat the study with the NPI variable represented as a real number, utilising between 3 and 7 membership functions. By doing so, we will be able to examine whether the membership functions can be trained to better match the data, rather than using the existing fixed clinical cut-offs presented in Table I.

As indicated above, we aim to then present the results to our clinical collaborators in order to investigate whether explicit clinical knowledge can be incorporated into the ANFIS model, in order to refine the prediction of survival.

We also aim to create ANFIS models for other clinical data sets — we have recently obtained data for a cohort of over 400 colorectal cancer patients with ten year follow-up survival data.

## ACKNOWLEDGEMENTS

The authors thank all members of the *Nottingham Breast Cancer Pathology Research Group*, and particularly Prof. Ian Ellis, Dr Andy Green and Dr Des Powe, for their help in preparing and providing the data set used in this study.

This study was supported by the Ministry of Higher Learning, Malaysia and Universiti Putra Malaysia (UPM).

## REFERENCES

- [1] Cancer Research UK, "Uk breast cancer incidence statistics," 2009.
- [2] L. Ohno-Machado, "Modeling medical prognosis: Survival analysis techniques," *Journal of Biomedical Informatics*, vol. 34, no. 6, pp. 428–439, 2001.
- [3] A. C. Joseph and S. W. David, "Applications of machine learning in cancer prediction and prognosis," *Cancer Informatics*, vol. 2, pp. 59–78, 2006.
- [4] P. J. G. Lisboa, "A review of evidence of health benefit from artificial neural networks in medical intervention," *Neural Networks*, vol. 15, no. 1, pp. 11–39, 2002.
- [5] H. Burke, P. Goodman, D. Rosen, D. Henson, J. Weinstein, F. Harrell, J. Marks, D. Winchester, and D. Bostwick, "Artificial neural network improve the accuracy of cancer survival prediction," *Cancer*, vol. 79, no. 4, pp. 857–862, 1997.
- [6] E. Biganzoli, P. Boracchi, L. Mariani, and E. Marubini, "Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach," *Statistics in Medicine*, vol. 17, no. 10, pp. 1169–1186, 1998.
- [7] T. G. Clark, M. J. Bradburn, S. B. Love, and D. G. Altman, "Survival analysis part i: Basic concepts and first analyses," *British Journal of Cancer*, vol. 89, no. 2, pp. 232–238, 2003.
- [8] D. Collet, *Modelling Survival Data In Medical Research*. London: Chapman & Hall, 1994.
- [9] M. Ilias, Z. Elias, and A. Ioannis, "An intelligent system for automated breast cancer diagnosis and prognosis using svm based classifiers," *Applied Intelligence*, vol. 30, no. 1, pp. 24–36, 2009.
- [10] I. Anagnostopoulos and I. Maglogiannis, "Neural network-based diagnostic and prognostic estimations in breast cancer microscopic instances," *Medical & Biological Engineering & Computing*, vol. 44, no. 9, pp. 773–784, 2006.
- [11] F. B. Oras and A. K. Sameem, "Anfis models for prognostic and survival rate analysis: Nasopharyngeal carcinoma," in *Management of*

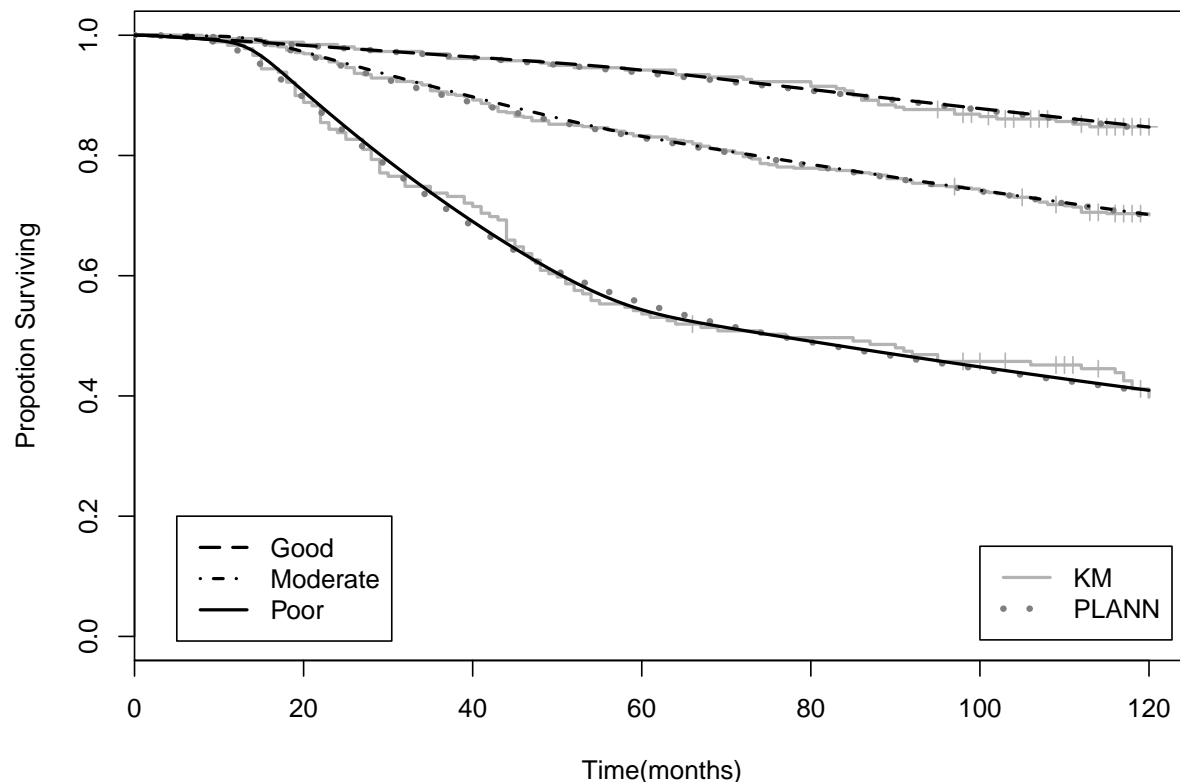


Fig. 12. Survival curve of actual Kaplan-Meier (solid grey lines) estimated against the best ANFIS model (black lines) and PLANN model (grey dotted lines)

*Innovation and Technology, 2008. ICMIT 2008. 4th IEEE International Conference on*, pp. 537–541, 2008.

[12] K. Muhammad Umer, C. Jong Pill, S. Hyunjung, and K. Minkoo, “Predicting breast cancer survivability using fuzzy decision trees for personalized healthcare,” in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pp. 5148–5151, 2008.

[13] D. Delen, G. Walker, and A. Kadam, “Predicting breast cancer survivability: a comparison of three data mining methods,” *Artificial Intelligence in Medicine*, vol. 34, no. 2, pp. 113–127, 2005.

[14] J. M. Jerez, L. Franco, E. Alba, A. Llombart-Cussac, A. Lluch, N. Ribelles, B. Munarriz, and M. Martin, “Improvement of breast cancer relapse prediction in high risk intervals using artificial neural networks,” *Breast Cancer Research and Treatment*, vol. 94, no. 3, pp. 265–272, 2005.

[15] L. Ohno-Machado, “A comparison of cox proportional hazards and artificial neural network models for medical prognosis,” *Computers in Biology and Medicine*, vol. 27, no. 1, pp. 55–65, 1997.

[16] R. M. Ripley, A. L. Harris, and L. Tarassenko, “Neural network models for breast cancer prognosis,” *Neural Computing & Applications*, vol. 7, no. 4, pp. 367–375, 1998.

[17] A. F. G. Taktak, A. C. Fisher, and B. E. Damato, “Modelling survival after treatment of intraocular melanoma using artificial neural networks and bayes theorem,” *Physics in Medicine and Biology*, vol. 1, p. 87, 2004.

[18] M. Negnevitsky, *Artificial Intelligence: a guide to intelligent systems*. Essex, England: Pearson Education Limited, 2005.

[19] P. M. Ravdin and G. M. Clark, “A practical application of neural network analysis for predicting outcome of individual breast cancer patients,” *Breast Cancer Research and Treatment*, vol. 22, no. 3, pp. 285–293, 1992.

[20] W. N. Street, “A neural network model for prognosis prediction,” in *Proceedings on the Fifteenth International Conference on Machine Learning*, pp. 540–546, Morgan Kaufmann, 1998.

[21] P. J. G. Lisboa, H. Wong, P. Harris, and R. Swindell, “A bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer,” *Artificial Intelligence in Medicine*, vol. 28, no. 1, pp. 1–25, 2003.

[22] L. A. Zadeh, “Fuzzy sets,” *Information and Control*, vol. 8, no. 3, pp. 338–&, 1965.

[23] J.-S. Jang, “Anfis adaptive-network-based fuzzy inference system,” *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 23, pp. 665–685, May/Jun 1993.

[24] J. M. Jerez, P. S. Ignacio, J. M. Dona, and E. Alba, “A neuro fuzzy decision model for prognosis of breast cancer relapse,” in *Current Topics in Artificial Intelligence*, pp. 638–645, Springer Berlin or Heidelberg, 2004.

[25] M. F. Abbod, J. W. F. Catto, M. Chen, D. A. Linkens, and F. C. Hamdy, “Artificial intelligence techniques for the prediction of bladder cancer progression,” in *Proceedings of the 13th European Symposium On Artificial Neural Networks Bruges (Belgium)*, pp. 109–114, 2005.

[26] M. Galea, R. Blamey, C. Elston, and I. Ellis, “The nottingham prognostic index in primary breast cancer,” *Breast Cancer Research and Treatment*, vol. 22, no. 3, pp. 207–219, 1992.

[27] C. M. Bishop, *Neural Networks for Pattern Recognition*. UK: Oxford University Press, 2007.

[28] B. D. Ripley, *Pattern recognition and neural networks*. Cambridge: Cambridge University Press, 1996.