

Artificial Neural Network Model for Predicting Lung Cancer Survival

Hansapani Rodrigo, Chris P. Tsokos

Department of Mathematics and Statistics, University of South Florida, Tampa, USA

Email: sarasepa@mail.usf.edu, ctsokos@usf.edu

How to cite this paper: Rodrigo, H. and Tsokos, C.P. (2017) Artificial Neural Network Model for Predicting Lung Cancer Survival. *Journal of Data Analysis and Information Processing*, 5, 33-47.
<https://doi.org/10.4236/jdaip.2017.51003>

Received: September 21, 2016

Accepted: February 19, 2017

Published: February 22, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The object of our present study is to develop a piecewise constant hazard model by using an Artificial Neural Network (ANN) to capture the complex shapes of the hazard functions, which cannot be achieved with conventional survival analysis models like Cox proportional hazard. We propose a more convenient approach to the PEANN created by Fornili *et al.* to handle a large amount of data. In particular, it provides much better prediction accuracies over both the Poisson regression and generalized estimating equations. This has been demonstrated with lung cancer patient data taken from the Surveillance, Epidemiology and End Results (SEER) program. The quality of the proposed model is evaluated by using several error measurement criteria.

Keywords

Survival Analysis, Hazard Prediction, Artificial Neural Network, Piecewise Exponential Survival Model, Censored Data, Lung Cancer

1. Introduction

Precise prediction of the survival and the hazard has been a challenging task through-out the past years. Research scientists have used parametric methods quite often to serve this purpose. However, they impose certain distributional assumptions on the hazard functions [1]. Cox proportional hazard [2] model is an another well-known approach which has been used extensively in the survival analysis. Though this model allows flexible modeling of the hazard with unspecified baseline hazard function, the assumption of time-independence of the hazard ratio may not always be correct [3]. Use of this model without verification of those assumptions can lead to misleading results. A more intuitive approach is to assume the hazard ratio to be independent of time, just for smaller periods, partitioning the whole time period into several intervals and introducing the piecewise constant hazard model. This has been advocated as a flexible and parsimonious tool in the literature [4] [5] [6] [7] and is generally useful for

interpreting cancer survival and to facilitate the treatments and diagnoses [8]. There exist several other techniques for flexible modeling of the hazard function. For example, Boracchi *et al.* [9] have developed a model with cubic splines, whereas Diehl *et al.* [10] have developed a nonparametric model based on kernel density estimation approach.

Artificial intelligence neural networks (ANNs) have been extremely popular in almost every field, including computer science, engineering and in the biomedical field among others. They have the strength of making predictions based on both individual attributable variables and possible complex interactions among them. In addition to that, ANNs have the capability of handling nonlinear functions and non-additive effects. Moreover, they are free of any statistical assumptions. Thus, ANN based survival analysis models serve as efficient alternatives to the conventional survival analysis models with enhanced predictive power. One of the earliest work in survival analysis with ANN was introduced by Faraggi and Simon [11] where they have used ANN as a basis for a nonlinear proportional hazard model. Another work, Biganzoli *et al.* [12] have used ANN to predict the smoothed discrete hazards as conditional probabilities of failure. Ravdin and Clark [13] have shown that ANN can be used to predict the patient outcome with censored survival data including time as a covariate. ANN has also been used in modeling cause-specific hazards [14]. Modeling of the piecewise exponential model (piecewise constant hazard) using ANN (PEANN) is proposed by Fornili *et al.* [15]. This method accommodates a greater flexibility in modeling the complex hazard functions. However, when using this model for a large amount of data, the analysis becomes difficult or even impossible due to the high data redundancy involved with the modeling.

In the present study, we have modified the PEANN model by combining it with another ANN model introduced by Mani *et al.* [16] to develop a more efficient model. As we mentioned earlier, the proposed model has the capability of handling a large amount of data. Importantly, it improves the prediction accuracies. This has been demonstrated by using lung cancer patients' data, and their hazards were predicted in the presence of competing risks. A comprehensive evaluation of the proposed model is conducted by using several error measurements including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Percentage Error (MPE) and Root Square Error (RSE). We compare our model results with the Generalized Estimating Equations (GEE). The results of the proposed model are much better than the competing models.

This paper is structured as follows. In Section 2, we introduce the new ANN system and related theory along with other models which we used for comparison. Following to that, we present our results. The final section discusses the implication and limitations.

2. Materials and Methodology

2.1. The Piecewise Constant Hazard Model

Let T be the survival or the follow-up time for subjects $i = 1, 2, \dots, N$ where $T =$

$\min \{\text{Survival Time, Censoring Time}\}$, and \mathbf{X} be the covariates. We consider different types of competing risks, R , which causes the subject to observe the same event of interest [17]. Equation (1) defines the hazard function for the r^{th} risk,

$$\lambda(r, t, \mathbf{X}) = \lim_{\Delta t \rightarrow 0^+} \frac{P(t < T \leq t + \Delta t, R = r | T \geq t, \mathbf{X})}{\Delta t} \quad (1)$$

Then the corresponding survival function and the probability density function can be obtained by Equation (2) and Equation (3) as given below.

$$S(t, \mathbf{X}) = \exp\left(-\int_0^t \lambda(., u, \mathbf{X}) du\right) \quad (2)$$

and

$$f(t | \mathbf{X}) = \lambda(., t, \mathbf{X}) S(t, \mathbf{X}) \quad (3)$$

where $\lambda(., t, \mathbf{X}_i) = \sum_{r=1}^R \lambda(r, t, \mathbf{X}_i)$ even without assuming independence among the competing risks. Thus, for independent observations, the likelihood function L can be written by Equation (4) assuming non-informative censoring,

$$L = \prod_{i=1}^N f(t_i | \mathbf{X}_i)^{\delta_i} S(t_i, \mathbf{X}_i)^{1-\delta_i} = \prod_{i=1}^N \frac{\lambda(., t_i, \mathbf{X}_i)}{\exp\left(\int_0^{t_i} \lambda(., u, \mathbf{X}_i) du\right)} \quad (4)$$

where δ_i is equal to 0 if the subject i is censored and 1 otherwise. Under the piecewise constant hazard model, the follow up time T is divided into several disjoint time intervals, a_0, a_1, \dots, a_j , where $a_0 = 0$ and $a_j \rightarrow \infty$ and the hazard function for the r^{th} risk is assumed to be constant in the j^{th} time period $[a_{j-1}, a_j)$. Hence, we have,

$$\lambda(., t, \mathbf{X}_i) = \lambda(., j, \mathbf{X}_i)$$

where $\lambda(., j, \mathbf{X}_i) = \sum_{r=1}^R \lambda(r, j, \mathbf{X}_i)$ for each subject. Then, the modified likelihood function can be written as in Equation (5),

$$L = \prod_{i=1}^N \frac{\prod_{j=1}^{J_i} \lambda(., j, \mathbf{X}_i)^{\delta_{ij}}}{\exp\left(\sum_{j=1}^{J_i} \lambda(., j, \mathbf{X}_i) \tau_{ij}\right)} = \frac{1}{\prod_{i=1}^N \prod_{j=1}^{J_i} \tau_{ij}^{\delta_{ij}}} \prod_{i=1}^N \prod_{j=1}^{J_i} \frac{\left(\lambda(., j, \mathbf{X}_i)^{\delta_{ij}} \tau_{ij}^{\delta_{ij}}\right)}{\delta_{ij}! \exp\left(\lambda(., j, \mathbf{X}_i) \tau_{ij}\right)} \quad (5)$$

where

$$\delta_{ij} = \begin{cases} 1, & \text{if the } i^{\text{th}} \text{ subject is deceased during the } j^{\text{th}} \text{ interval} \\ 0, & \text{otherwise} \end{cases}$$

J_i is the last interval that the subject i is observed and τ_{ij} is the corresponding exposure time which is defined by

$$\tau_{ij} = \begin{cases} a_j - a_{j-1}, & \text{if } t_i \geq a_j \\ t_i - a_{j-1}, & \text{if } a_{j-1} < t_i \leq a_j \\ 0, & \text{if } t_i \leq a_{j-1} \end{cases}$$

The kernel given in Equation (5) corresponds to the likelihood of the Poisson random variable δ_{ij} with mean $\mu_{ij} = \lambda(., j, \mathbf{X}_i) \tau_{ij}$. Applying the logarithm on

both sides, we obtain

$$\log(\mu_{ij}) = \log(\lambda(., j, X_i)) + \log(\tau_{ij}). \quad (6)$$

It has been shown that, $\lambda(., j, X_i)$ in Equation (6) can be modeled with a Poisson log linear model of the form $\log(\lambda(., j, X_i)) = \alpha_j + x'_i \beta$ as given in [18] [19]. However, this model is not effective in handling large amount of subjects over a longer period. Sometimes, the analysis even becomes impossible, due to the vast amount of δ_{ij} observations that is created.

An alternative method is to group the exposure times and the similar δ_{ij} values for each interval j and then use the Poisson regression. Nevertheless, overdispersion might be a problem with this kind of a Poisson model due to the correlated nature of the data. As a possible solution, we can use generalized estimating equations, which is an extension of the generalized linear model [20].

2.2. The Proposed ANN Model

In this section, we introduce an efficient method of modeling the hazard function with artificial neural networks. ANNs allow flexible modeling of the hazard function without any probability distributional assumptions. Moreover, it captures the nonlinear effects of the risk factors.

Preceding our model, Fornili *et al.* [15] have used ANN to model the $\lambda(., j, X_i)$ in Equation (6) where they referred it as PEANN. However, this model uses the same data structure as in the Poisson log linear model, and hence, not very effective due to the high data redundancy. As a solution, we introduce a new ANN model with a different network architecture. The new ANN model has several output nodes; each corresponds to a different time interval. This structure is similar to the ANN model used by Mani *et al.* [16]. The proposed model is a competent alternative to the PEANN model, especially when we need to deal with large number of subjects or/and longer follow up periods.

2.2.1. Data Preprocessing

Prior to using the proposed ANN model, data need to be preprocessed. This process can be explained using a simple example. Consider three subjects, called A , B and C who have been observed for J number of years. Suppose we have information about their risk factors X_1 and X_2 , the risk type, their survival time and whether they were censored or not, as given in Table 1. We have considered two different risk types, R_1 or R_2 where each subject can decease due to one of them. The “censor” variable indicates whether a subject has lost follow up somewhere during the study period or not. Hence, for all deceased subjects during the study period, it is set to zero. As can be seen, subject A and B were deceased due to risk types, R_1 and R_2 after 3 and 4 years respectively. According to Table 1, subject C has lost follow-up after 2 years.

In order to use the new ANN model, this information needs to be pre-processed as given in Table 2. An ANN consists of several layers; input, hidden and output. Input and the output nodes are usually determined by the nature

Table 1. Sample data.

Subject	X_1	X_2	Survival Time	Risk Type	Censor
A	1	0	3	R_1	0
B	1	1	4	R_2	0
C	1	1	2	R_1 or R_2	1

Table 2. Preprocessed data.

Subject	X_1	X_2	R_1	R_2	$h(1)$	$h(2)$	$h(3)$	$h(4)$	$h(5)$...	$h(J)$
A	1	0	1	0	0	0	1	1	1	...	1
B	1	1	0	1	0	0	0	1	1	...	1
C	1	1	1	0	0	0	0.31	0.24	0.12	0.01
C	1	1	0	1	0	0	0.42	0.57	0.45		0.63

of the data. In this example, we have four inputs, the covariates X_1 and X_2 and two indicator variables R_1 and R_2 which we used to denote the risk type of the subject. For censored subjects like C, we need to consider the possibility of exposing into each of these risk types. Hence, those data are repeated twice as given in **Table 2**. Assuming a constant hazard for each year, we have J number of output nodes in the ANN. For each subject i , we have a vector of outputs with the following values.

$$h(j) = \begin{cases} 0, & \text{subject is alive} \\ 1, & \text{subject is deceased in the } j^{\text{th}} \text{ time interval due to } r^{\text{th}} \text{ risk} \\ \frac{d_{rj}}{n_j}, & \text{subject is censored} \end{cases}$$

where n_j , is the total number of subjects alive at the beginning of j^{th} time interval and, d_{rj} is the number of deceased subjects in the j^{th} time interval due to the r^{th} risk. The ratio, $\frac{d_{rj}}{n_j}$, is the Kaplan-Meier [21] hazard probabilities of r^{th}

risk for the period j . In summary, if the subject is alive, then $h(j) = 0$, if the subject is deceased, then $h(j) = 1$ and if the subject is censored, then $h(j) = \frac{d_{rj}}{n_j}$. With this approach, we can significantly reduce the data redundancy.

2.2.2. Network Training

In developing the proposed ANN model, we used the hyperbolic tangent and the exponential activation functions in the hidden and the output layers. The proposed ANN structure is represented in **Figure 1**. The network output, $y(j|r, X)$, gives the hazard for each time interval j as in Equation (7)

$$y(j|r, X) = \lambda(r, j, X) = \exp \left(w_{oj}^{(2)} + \sum_{k=1}^H \tanh \left(w_{ok}^{(1)} + \sum_{l=1}^p w_{lk}^{(1)} x_l \right) w_{kj}^{(2)} \right) \quad (7)$$

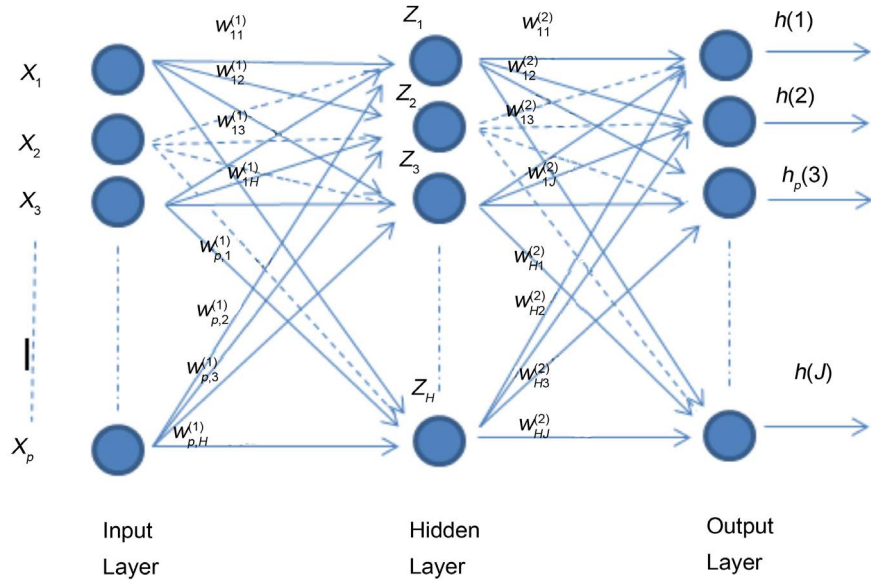


Figure 1. The proposed ANN structure.

where $j = 1, 2, \dots, J$. Moreover, x_1, \dots, x_p are the inputs, and $w_{ik}^{(1)}$ and $w_{kj}^{(2)}$ are the hidden and output layer weights.

During the training, we minimized the regularized canonical error function given by Equation (8), where α is the non-negative weight decay penalty term. By adding a weight decay term, we expect to minimize the network overfitting, [22]. As per [23], we trained several ANN models with different weight decay values.

$$E = -\sum_{i=1}^N \sum_{j=1}^J \left(h(j) \log y(j|r, \mathbf{X}) \right) - y(j|r, \mathbf{X}) \tau_{ij} + \alpha \left(\sum_{k=0}^H \sum_{l=0}^p \left(w_{lk}^{(1)} \right)^2 + \sum_{k=0}^H \sum_{j=0}^J \left(w_{kj}^{(2)} \right)^2 \right) \quad (8)$$

We used a k -fold cross validation technique to find the optimal number of hidden nodes in each network. When using a k -fold cross validation technique, the training dataset is divided into k folds, where $k-1$ folds are used to train the model while the remaining set is used for the validation. This process is repeated k times until each fold is used for validation. Other model selection criteria like Akaike Information Criterion and Bayesian Information Criterion are not suitable for model selection as the error function, E is not a linear function of the network weights. The optimal networks are selected based on the minimum average validation error, and each of them is used to make the hazard rate predictions in the testing data set. Scale conjugate algorithm is used for weight optimization. Then, the corresponding survival probabilities are obtained using Equation (9). We evaluated the models using several error measurements calculated based on the predicted median survival time and the actual survival time of the non-censored subjects in the testing data.

$$S_i(t) = \exp \left(- \left(\sum_{j=1}^J y(j|r, \mathbf{X}_i) \right) \right) \quad (9)$$

2.3. The Lung Cancer Data

The data for our study is selected from the Surveillance, Epidemiology and End Results (SEER) program [24], and it contains details of 38,262 white lung cancer patients data who have been diagnosed from 2004 to 2009. Among these, 23,332 subjects were deceased due to lung cancer and 4652 were deceased due to some other causes. The rest were considered as censored due to missing information or lost in the follow-up.

In our analysis, four risk factors were used: age at diagnosis, tumor size, histology and the stage of the cancer. As can be seen from **Table 3**, a higher amount of patients were between the ages of 65 - 75 and most of them had distant metastasis. The majority of the patients were diagnosed with adeno or squamous cell carcinoma. The overall median follow-up time for males was 1.33 years and 2 years for females, while median tumor size is about 38 mm and 32 mm for the two groups respectively.

Table 3. Lung cancer patient information.

	Male	Female
Cause of Death		
Lung	13,029 (64%)	10,303 (58%)
Other	2724 (13%)	1928 (11%)
Censored	4767 (23%)	5511 (31%)
Age at Diagnosis		
45 - 49 years	635 (3%)	705 (4%)
50 - 54 years	1320 (6%)	1161 (7%)
55 - 59 years	2206 (11%)	1747 (10%)
60 - 64 years	3208 (16%)	2515 (14%)
65 - 69 years	3757 (18%)	3127 (18%)
70 - 74 years	3723 (18%)	3086 (17%)
75 - 79 years	3187 (16%)	2837 (16%)
80 - 84 years	1793 (9%)	1826 (10%)
85+ years	691 (3%)	738 (4%)
Stage of the Cancer		
Localized	5536 (27%)	5525 (31%)
Regional	7028 (34%)	5816 (33%)
Distant	7956 (39%)	6401 (36%)
Histology Type		
Adeno	9162 (45%)	10,056 (57%)
Squamous	8492 (41%)	5054 (28%)
Large Cell	917 (4%)	691 (4%)
Small-cell	1949 (10%)	1941 (11%)
Total	20,520	17,742

We found that the survival time between males and females to be significantly different from each other, which was already a known fact [25], and hence, two separate analysis were conducted for each of them. In order to develop the piecewise constant hazard model, we partitioned the time into six disjoint intervals each with a 12-month period (1 year). Then we carried our analysis using GEE and ANN models using in SAS and MATLAB.

3. Results

For both males and females, we created a training data set of 70% and a testing data set of 30%. The training set was used to train the models while the testing dataset was used to evaluate the prediction accuracies of the proposed models.

We started our analysis with Poisson regression models. However, according to the deviance and the Pearson chi-square statistics, none of those models were adequate [26]. This might be due to the high correlation among data. We even tried to use several other models, a Poisson model with an overdispersion parameter and a negative binomial model. There was no significance difference; results remain unchanged with models being inadequate. Therefore, we chose an alternative method, generalized estimating equations. Using this approach, we came up with two different statistical models for males and females. We found the interaction between the stage and histology to be significant in both males and females. Applying these two models, we were able to predict the hazard and to obtain the corresponding survival probabilities for our lung cancer testing data.

Following to that, we proceeded with building ANN models. We created both PEANN and our proposed ANN models. As mentioned earlier, we considered five different weight decay values: 0.01, 0.025, 0.05, 0.075, 0.1 and 10-fold cross validation was used to find the optimal number of hidden nodes in each case. The optimal network is selected based on the minimum average validation error. By using each optimal network, we predicted the hazard and corresponding survival probabilities for the testing data. In order to evaluate the prediction accuracies of ANN and GEE, we used the actual survival times and their predicted median survival times of non-censored subjects. For a better comparison, several prediction errors were considered, including the root mean square error (RMSE): average differences between actual and the predicted values, mean absolute error (MAE): average of the absolute errors, mean percentage error (MPE): average of percentage errors, and relative squared error (RSE): total squared error normalized by the total squared error of the simple predictor for both males and females as given in Table 4 and Table 5. We can get a better idea of using these different error measurements as they serve various aspects of the model predictions. For example, RMSE and MAE are conventional measures of prediction accuracies while MPE acts as a good measure of bias in the predictions. RSE gives the relative error to what it would have been if a simple predictor (the average of the actual values) had been used.

As can be seen from Table 4 and Table 5, the proposed ANN method is better

Table 4. Model evaluation for males.

Male	GEE	Weight Decay 0.01		Weight Decay 0.025		Weight Decay 0.05		Weight Decay 0.075		Weight Decay 0.1	
		NewANN	PEANN	NewANN	PEANN	NewANN	PEANN	NewANN	PEANN	NewANN	PEANN
RMSE	4.0986	2.3253	3.5967	2.2416	3.6277	2.2693	3.5190	2.2144	3.7136	2.3444	3.5561
MAE	3.5155	1.69	2.8767	1.6226	2.9106	1.6412	2.7791	1.6174	3.0070	1.7292	2.8226
RSE	8.4539	2.721	6.4604	2.5287	6.5724	2.5916	6.1844	2.4676	6.8873	2.7659	6.3154
MPE	-2.5349	-0.6645	-1.6856	-0.6137	-1.7456	-0.6125	-1.5952	-0.5819	-1.8934	-0.7077	-1.6561
Data Count	4659	4659	4659	4659	4659	4659	4659	4659	4659	4659	4659

Table 5. Model evaluation for females.

Female	GEE	Weight Decay 0.01		Weight Decay 0.025		Weight Decay 0.05		Weight Decay 0.075		Weight Decay 0.1	
		NewANN	PEANN	NewANN	PEANN	NewANN	PEANN	NewANN	PEANN	NewANN	PEANN
RMSE	4.3146	2.5209	3.9056	2.5232	3.9682	2.4737	3.8961	2.4969	3.9611	2.4871	3.9256
MAE	3.8683	1.8927	3.2999	1.8896	3.3834	1.8529	3.2912	1.8700	3.3714	1.8655	3.3240
RSE	8.6342	2.9475	7.0751	2.9529	7.3036	2.8383	7.0407	2.8916	7.2776	2.869	7.1476
MPE	-2.9081	-0.8038	-1.9276	-0.811	-2.0216	-0.7844	-1.9000	-0.7757	-2.0242	-0.7689	-1.9494
Data Count	3568	3568	3568	3568	3568	3568	3568	3568	3568	3568	3568

than both GEE and PEANN with respect to RMSE and MAE for both genders. In addition to that, RSEs for new ANNs are smaller than those two types of models. Although the predictions of new ANNs have negative biases which indicate underestimations of the survival, it is significantly less than the other two models. In particular, we found the smallest error values for the new ANN models with weight decay 0.05 and 0.075 for females and males respectively. Further analysis of the hazard rates was carried out using those two models.

Figure 2 and **Figure 3** represent the hazard variation among male and female patients according to different tumor sizes while keeping the other categorical risk factors in their mode categories. It is important to note that we have chosen different tumor size ranges for males and females depending on the available training data. We followed this precaution to increase the validity of our results, as ANNs are data driven models, and hence depend heavily on the amount of training data with each tumor size. According to **Figure 2**, we can see that the hazard rate for males increases over the follow-up years for smaller tumor sizes while it slightly decreases for higher tumor sizes. Males seem to have a high risk within the first two years of diagnosis. As opposed to males, there are significant differences in the hazard rate with the tumor sizes among females (**Figure 3**). For smaller tumors, hazard rate increases over the years while for larger tumors, it increases during the first three years and then decreases. The above graphs reveal that our ANN models are capable of capturing the complex shapes of the hazard functions, which was one of the main advantages of ANN over conventional survival analysis models. Despite some unusual findings, overall these hazard functions capture the indwelling patterns in the data.

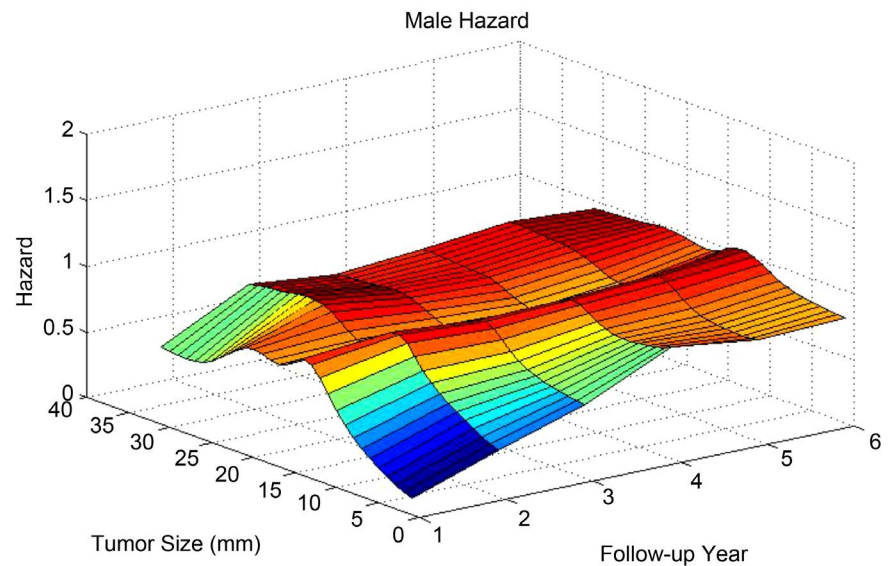


Figure 2. Hazard variation according to tumor size (mm) for males.

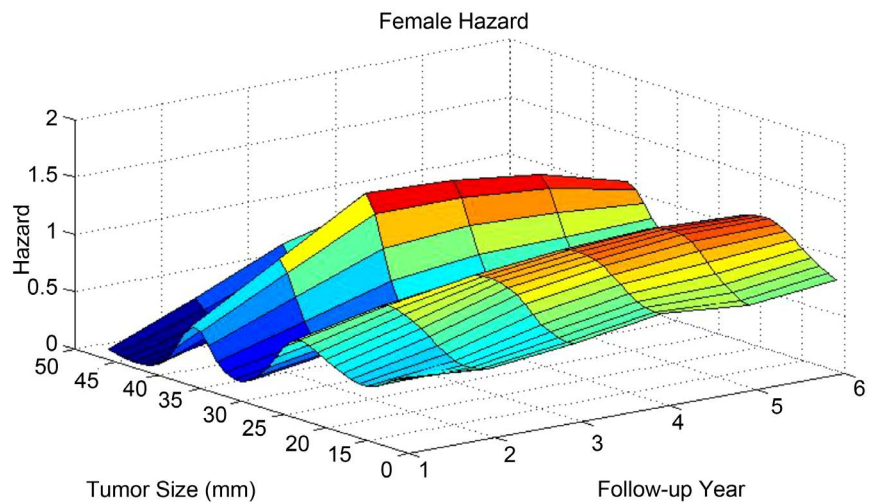


Figure 3. Hazard variation according to tumor size (mm) for females.

Figure 4 represents the variation in the hazard rates according to the age group and histology types, for both males and females. From that, we can observe that there are noticeable differences in the hazard. Nevertheless, we did not find the interaction between age group and the histology to be significant from our GEE model. This depicts the capability of detecting nonlinear patterns by ANN models. In accordance with [27] as, we observed a comparatively higher hazard for elderly patients than for younger patients, regardless of their gender. Usually, for elder patients, hazard seems to be elevated soon after the diagnosis while for younger patients, hazard tends to increase over time.

Figure 5 and **Figure 6** represent the variation in the hazard function according to the histology and stage for males and females respectively. These graphs reveal the fact that there exist significant differences in the hazard according to the histology and stages among males and females. We can see that the hazard for

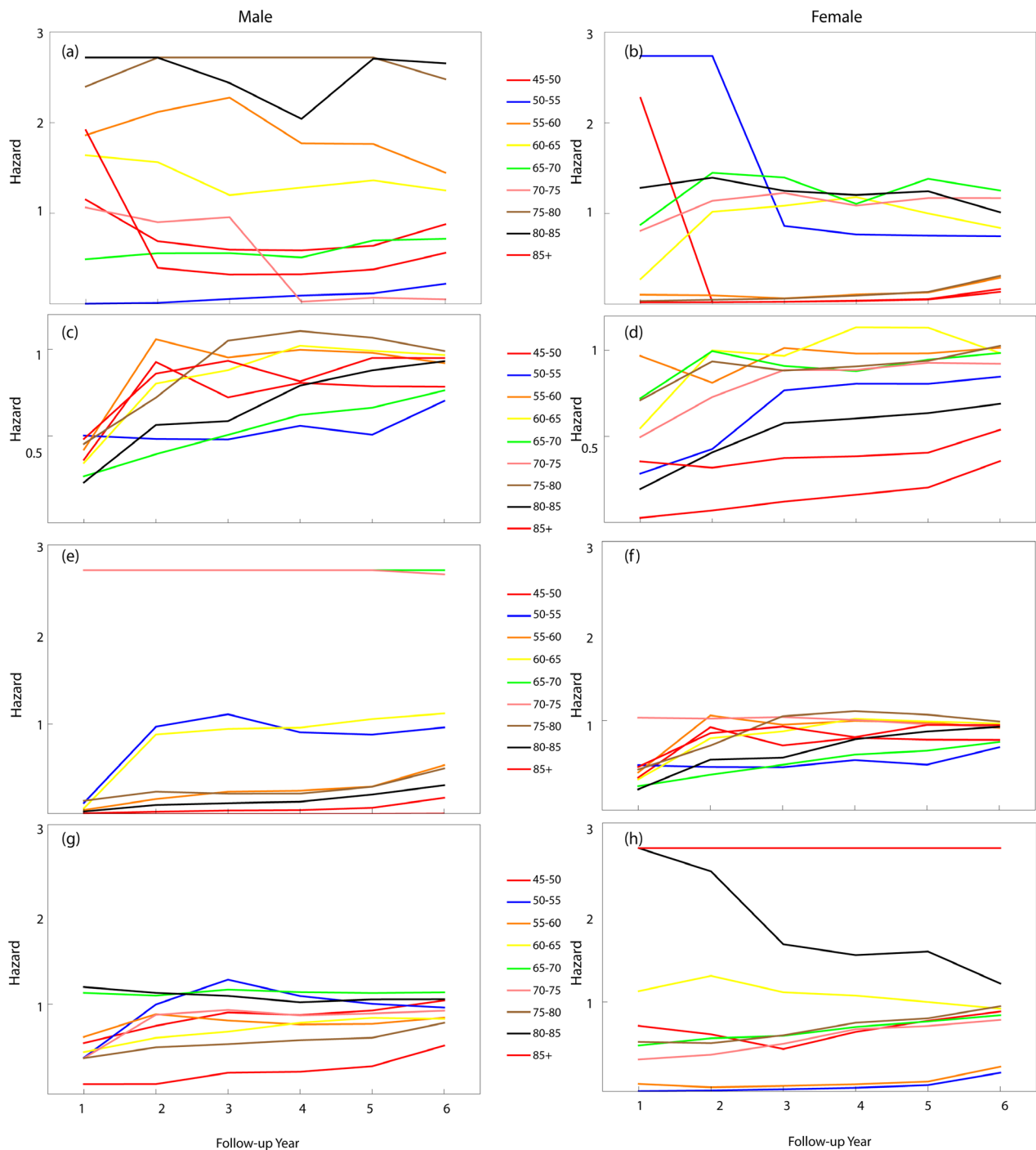


Figure 4. Hazard variation according to age group and histology types: (a) Small male, (b) Small female, (c) Adeno male, (d) Adeno female, (e) Large male, (f) Large female, (g) Squamous male, (h) Squamous female.

small cell carcinoma is higher in females than in males for all stages as confirmed by the [28].

4. Discussion

We have introduced a new neural network architecture to model the piecewise constant hazard model. This provides a more convenient approach to handle a

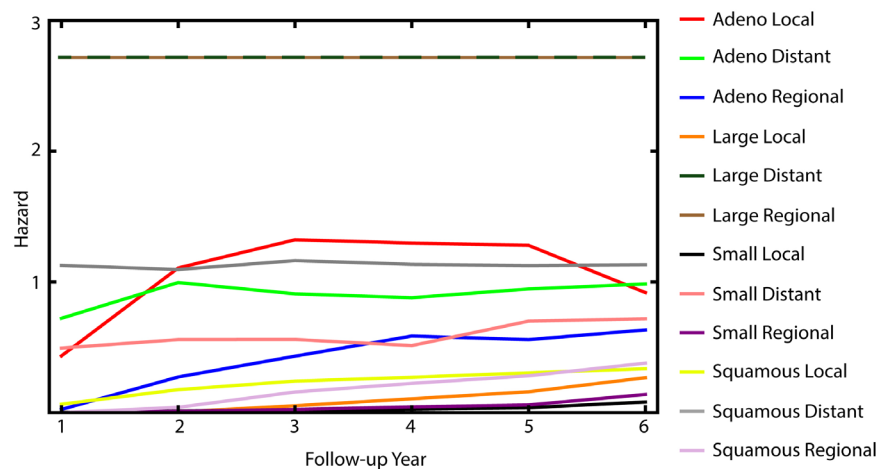


Figure 5. Hazard variation among males according to histology and stage.

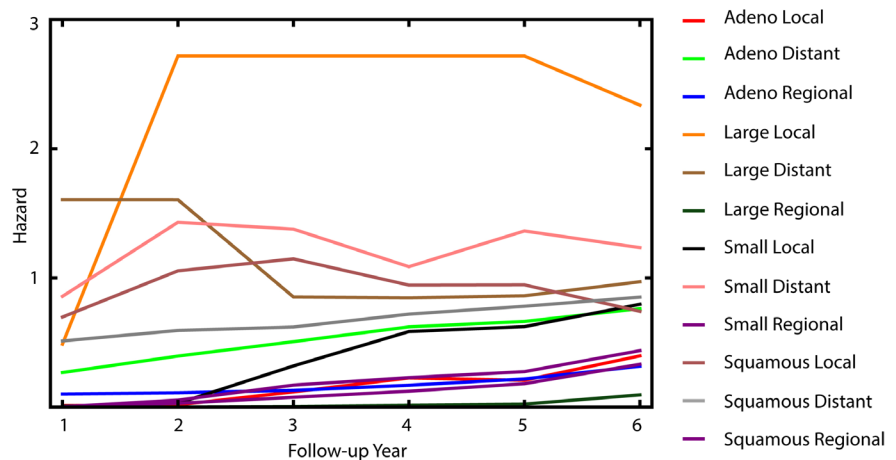


Figure 6. Hazard variation among females according to histology and stage.

large amount of observations over longer periods. In particular, our ANN model captures the complex shapes of the hazard functions, in the presence of competing risks. Moreover, these ANN models can handle nonlinear and non-additive effects among the risk factors. The new method overcomes several limitations associated with the traditional piecewise constant hazard model. Our ANN model is capable of modeling the hazard functions even with a large amount of data where the equivalent Poisson regression model of the piecewise constant hazard model fails. Importantly, the prediction accuracy of the survival times given by the proposed ANN model is higher than both generalized estimating equation models and the PEANN model. However, PEANN model is much suitable than our ANN model when there are time-dependent risk factors, as it is specially designed to deal with that kind of data. Our findings confirm the fact that elder patients have relatively higher hazard compared to younger patients. Their hazard is usually at the greatest around the first two years of diagnosis while for younger patients it tends to vary. However, we advise doing further analysis before making any clinical decisions.

In developing the proposed ANN model, we use the nonlinear Poisson regression model used in [29]. Network parameters are trained using the back propagation algorithm. In order to compute the Hessian matrix, which contains the derivatives of the error function with respect to the weights and biases, we used a special algorithm developed by Pearlmutter [30]; a similar approach is that of Nabney [31]. We have used the cross-validation technique to identify the optimal number of hidden nodes in the neural networks. There is a possibility of incorporating Bayesian learning techniques in our model selection, like the evidence procedure. This is one of possible future research.

References

- [1] Wienke, A. (2010) *Frailty Models in Survival Analysis*. CRC Press, Boca Raton. <https://doi.org/10.1201/9781420073911>
- [2] Cox, D.R. (1972) Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**, 187-220.
- [3] Moolgavkar, S.H. (2006) Fine Particles and Mortality. *Inhalation Toxicology*, **18**, 93-94. <https://doi.org/10.1080/08958370500419207>
- [4] Breslow, N. (1974) Covariance Analysis of Censored Survival Data. *Biometrics*, **30**, 89-99. <https://doi.org/10.2307/2529620>
- [5] Aitkin, M., Laird, N.M. and Francis, B. (1983) A Reanalysis of the Stanford Heart Transplant Data. *Journal of the American Statistical Association*, **78**, 264-274. <https://doi.org/10.1080/01621459.1983.10477959>
- [6] Clark, D.E. and Ryan, L.M. (2002) Concurrent Prediction of Hospital Mortality and Length of Stay from Risk Factors on Admission. *Health Services Research*, **37**, 631-645. <https://doi.org/10.1111/1475-6773.00041>
- [7] Han, G., Schell, M.J. and Kim, J. (2014) Improved Survival Modeling in Cancer Research Using a Reduced Piecewise Exponential Approach. *Statistics in Medicine*, **33**, 59-73. <https://doi.org/10.1002/sim.5915>
- [8] Goodman, M.S., Li, Y. and Tiwari, R.C. (2011) Detecting Multiple Change Points in Piecewise Constant Hazard Functions. *Journal of Applied Statistics*, **38**, 2523-2532. <https://doi.org/10.1080/02664763.2011.559209>
- [9] Borachi, P., Biganzoli, E. and Marubini, E. (2003) Joint Modelling of Cause-Specific Hazard Functions with Cubic Splines: An Application to a Large Series of Breast Cancer Patients. *Computational Statistics & Data Analysis*, **42**, 243-262. [https://doi.org/10.1016/S0167-9473\(02\)00122-6](https://doi.org/10.1016/S0167-9473(02)00122-6)
- [10] Diehl, S. and Stute, W. (1988) Kernel Density and Hazard Function Estimation in the Presence of Censoring. *Journal of Multivariate Analysis*, **25**, 299-310. [https://doi.org/10.1016/0047-259X\(88\)90053-X](https://doi.org/10.1016/0047-259X(88)90053-X)
- [11] Faraggi, D. and Simon, R. (1995) A Neural Network Model for Survival Data. *Statistics in Medicine*, **14**, 73-82. <https://doi.org/10.1002/sim.4780140108>
- [12] Biganzoli, E., Boracchi, P., Mariani, L. and Marubini, E. (1998) Feed Forward Neural Networks for the Analysis of Censored Survival Data: A Partial Logistic Regression Approach. *Statistics in Medicine*, **17**, 1169-1186. [https://doi.org/10.1002/\(SICI\)1097-0258\(19980530\)17:10<1169::AID-SIM796>3.0.CO;2-D](https://doi.org/10.1002/(SICI)1097-0258(19980530)17:10<1169::AID-SIM796>3.0.CO;2-D)
- [13] Ravdin, P.M. and Clark, G.M. (1992) A Practical Application of Neural Network Analysis for Predicting Outcome of Individual Breast Cancer Patients. *Breast Cancer Research and Treatment*, **22**, 285-293. <https://doi.org/10.1007/BF01840841>

- [14] Boracchi, P., Biganzoli, E. and Marubini, E. (2001) Modelling Cause-Specific Hazards with Radial Basis Function Artificial Neural Networks: Application to 2233 Breast Cancer Patients. *Statistics in Medicine*, **20**, 3677-3694. <https://doi.org/10.1002/sim.1112>
- [15] Fornili, M., Ambrogi, F., Boracchi, P. and Biganzoli, E. (2014) Piecewise Exponential Artificial Neural Networks (PEANN) for Modeling Hazard Function with Right Censored Data. In: Formenti, E., Tagliaferri, R. and Wit, E., Eds., *Computational Intelligence Methods for Bioinformatics and Biostatistics*, Springer, Berlin, 125-136. https://doi.org/10.1007/978-3-319-09042-9_9
- [16] Mani, D., Drew, J., Betz, A. and Datta, P. (1999) Statistics and Data Mining Techniques for Lifetime Value Modeling. *KDD'99 Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Diego, 15-18 August 1999, 94-103.
- [17] Satagopan, J.M., Ben-Porat, L., Berwick, M., Robson, M., Kutler, D. and Auerbach, D. (2004) A Note on Competing Risks in Survival Data Analysis. *British Journal of Cancer*, **91**, 1229-1235. <https://doi.org/10.1038/sj.bjc.6602102>
- [18] Holford, T.R. (1980) The Analysis of Rates and of Survivorship Using Log-Linear Models. *Biometrics*, **36**, 299-305. <https://doi.org/10.2307/2529982>
- [19] Laird, N. and Olivier, D. (1981) Covariance Analysis of Censored Survival Data Using Log-Linear Analysis Techniques. *Journal of the American Statistical Association*, **76**, 231-240. <https://doi.org/10.1080/01621459.1981.10477634>
- [20] Liang, K.-Y. and Zeger, S.L. (1986) Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika Trust*, **73**, 13-22. <https://doi.org/10.1093/biomet/73.1.13>
- [21] Kaplan, A.E.L. and Meier, P. (1958) Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, **53**, 457-481. <https://doi.org/10.1080/01621459.1958.10501452>
- [22] Bishop, C.M. (1995) *Neural Networks for Pattern Recognition*. Oxford University Press, New York.
- [23] Ripley, B.D. (1996) *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511812651>
- [24] Surveillance Epidemiology and End Results (SEER) Program, Research Data (1973-2009), Division of Cancer Control and Population Sciences, National Cancer Institute, Surveillance Research Program, Surveillance Systems Branch (2012). <http://www.seer.cancer.gov/>
- [25] Wisnivesky, J.P. and Halm, E.A. (2007) Sex Differences in Lung Cancer Survival: Do Tumors Behave Differently in Elderly Women? *Journal of Clinical Oncology*, **25**, 1705-1712. <https://doi.org/10.1200/JCO.2006.08.1455>
- [26] Pedan, A. (2001) Analysis of Count Data Using the SAS System. *Proceedings of the 26th Annual SAS Users Group International Conference*, Long Beach, 22-25 April 2001.
- [27] Tas, F., Cifci, R., Kilic, L. and Karabulut, S. (2013) Age Is a Prognostic Factor Affecting Survival in Lung Cancer Patients. *Oncology Letters*, **6**, 1507-1513.
- [28] Osann, K.E., Lowery, J.T. and Schell, M.J. (2000) Small Cell Lung Cancer in Women: Risk Associated with Smoking, Prior Respiratory Disease, and Occupation. *Lung Cancer*, **28**, 1-10. [https://doi.org/10.1016/S0169-5002\(99\)00106-3](https://doi.org/10.1016/S0169-5002(99)00106-3)
- [29] Fallah, N., Gu, H., Mohammad, K., Seyyedsalehi, S.A., Nourijelyani, K. and Eshrahgian, M.R. (2009) Nonlinear Poisson Regression Using Neural Networks: A Simulation Study. *Neural Computing and Applications*, **18**, 939-943. <https://doi.org/10.1007/s00521-009-0277-8>

- [30] Pearlmutter, B.A. (1994) Fast Exact Multiplication by the Hessian. *Neural Computation*, **6**, 147-160. <https://doi.org/10.1162/neco.1994.6.1.147>
- [31] Nabney, I.T. (2002) Netlab—Algorithms for Pattern Recognition. Springer, Gateshead.



Scientific Research Publishing

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact jdaip@scirp.org

