Improving Gastric Cancer Outcome Prediction Using Single Time-Point Artificial Neural Network Models

Hamid Nilsaz-Dezfouli[1], Mohd Rizam Abu-Bakar[1], Jayanthi Arasan[1], Mohd Bakri Adam[1] and Mohamad Amin Pourhoseingholi[2]

[1]Institute for Mathematical Research, Universiti Putra Malaysia, Serdang, Malaysia.
[2]Gastroenterology and Liver Diseases Research Center, Research Institute for Gastroenterology and liver Diseases, Shahid Beheshti University of Medical Sciences,Tehran, Iran.

Cancer Informatics
1–11

DOI: 10.1177/1176935116686062

ABSTRACT: In cancer studies, the prediction of cancer outcome based on a set of prognostic variables has been a long-standing topic of interest. Current statistical methods for survival analysis offer the possibility of modelling cancer survivability but require unrealistic assumptions about the survival time distribution or proportionality of hazard. Therefore, attention must be paid in developing nonlinear models with less restrictive assumptions. Artificial neural network (ANN) models are primarily useful in prediction when nonlinear approaches are required to sift through the plethora of available information. The applications of ANN models for prognostic and diagnostic classification in medicine have attracted a lot of interest. The applications of ANN models in modelling the survival of patients with gastric cancer have been discussed in some studies without completely considering the censored data. This study proposes an ANN model for predicting gastric cancer survivability, considering the censored data. Five separate single time-point ANN models were developed to predict the outcome of patients after 1, 2, 3, 4, and 5 years. The performance of ANN model in predicting the probabilities of death is consistently high for all time points according to the accuracy and the area under the receiver operating characteristic curve.

KEYWORDS: Gastric Cancer, Survival Analysis, Single Time-Point Artificial Neural Networks, Censored Data

RECEIVED: May 25, 2016. **ACCEPTED:** September 18, 2016.

PEER REVIEW: Five peer reviewers contributed to the peer review report. Reviewers' reports totaled 1451 words, excluding any confidential comments to the academic editor.

TYPE: Original Research

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Hamid Nilsaz-Dezfouli, Institute for Mathematical Research, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia. Email: hamid.nilsaz@gmail.com

Introduction

Background

Gastric cancer accounts for higher morbidity and mortality rates worldwide.[1,2] The latest rate of gastric cancer is 7.4 per 100 000 per year. The death rate is 3.3 per 100 000 men and women per year. Based on the estimates of the American Cancer Society for gastric cancer in the United States, an estimated 26 370 cases will be diagnosed with gastric cancer in 2016. It is estimated that due to this disease, 10 730 deaths will occur this year. The incidence of gastric cancer varies worldwide. Although gastric cancer is decreasing in the Western countries, it is still considered one of the most common types of cancers worldwide. Considering these facts, gastric cancer is one of the most preferred fields for investigation.[3,4]

To facilitate clinical decision making, the methods of survival analysis have been widely applied in gastric cancer studies, and a number of parametric regression models as well as the Cox proportional hazard models have been used. Most studies that used parametric methods or the Cox models have aimed at finding the relative importance of the prognostic factors in the development of the disease. The findings of these studies have revealed that some histologic, biomedical, and clinical variables have prognostic utility; however, these models have not qualified to provide an outcome prediction for patients with a given set of variables.[5] Thus, for an individual patient, the prediction of the disease outcome still remains a challenging task.[6]

In recent years, a lot of attention has been paid to the application of artificial neural network (ANN)–based methods for developing prognostic models in medicine. Artificial neural network models have been used in diagnosis, prognosis, and outcome prediction in a number of cancer studies. The usefulness of the ANN methodology is justified by the fact that ANN models do not assume a certain prior functional form and do not require fulfilling the assumptions required by a statistical technique. In this regard, multilayer feed-forward ANN models, also known as universal function approximators,[7] can overcome the proportionality and linearity constraints imposed by the conventional survival analysis techniques. A comparison of the performance of ANN models with other conventional statistical methods is demonstrated in a comprehensive study,[6] wherein the use of ANN models for the diagnosis and prognosis of a number of gastrointestinal diseases is investigated.[8] The findings demonstrated that for all diagnostic, prognostic, and prediction tasks, ANN models outperformed the conventional statistical methods.

Censoring is the distinguishing feature of survival analysis. In principle, censoring occurs when we have some information about the survival time of an individual, but we do not know the exact survival time. For example, a patient may neglect to follow up during the time of the study, withdraw from the

clinical trial, or experience a different event such as accidental death. In all such situations, the precise survival time of the patient will remain unknown. The standard structure of a neural network model does not allow a direct modelling of the censored data. This implies that the development of an ANN model for survival predictions requires the introduction of some strategies for incorporating censored data in constructing ANN survival models.

To estimate the survival of patients with cancer, the ANN models were first used by Ravdin and Clark.[9] They were also the first who questioned how to handle the censored data in neural network implementation for survival analysis. An ANN was developed and validated on a real clinical data set of 1373 patients with node-positive breast cancer. In their model, the time was recorded as a predictor. The follow-up time was split into different time intervals. The input vector of each failed subject was replicated for all time intervals. However, the censored subjects were replicated only for the observation of intervals. In addition, Ravdin and Clark reported that the network output roughly corresponds to the probability of unconditional events estimated using the Kaplan-Meier (K-M) method. However, the survival curves generated by this approach may not be monotonically decreasing.

Liestbl et al[10] integrated the survival analysis theory with the ANN methodology in a discrete multiple classification framework. They grouped the observed time T with k distinct intervals and specified the output nodes for each interval. For subject i, the kth target represents the survival status of the subject in the kth interval, and it is unspecified after failing or the censoring time. The proportional odds model is expressed as follows:

$$\frac{h_k(X_i)}{1-h_k(X_i)} = \frac{h_k(0)}{1-h_k(0)} \exp\left(\beta^T X_i\right) \quad k = 1, 2, \ldots, K, \qquad (1)$$

where X_i is the covariate vector, β is the vector of regression coefficients, and $h_0(t)$ is the baseline hazard rate for individuals with $X = 0$.

By defining

$$\theta_k = \log\left[\frac{h_k(0)}{1-h_k(0)}\right], \qquad (2)$$

equation (2) can be expressed as follows:

$$h_k(X_i) = \frac{\exp\left(\theta_k + \beta^T X_i\right)}{\exp\left(\theta_k + \beta^T X_i\right)} \qquad (3)$$

Thus, the distinct hazard rates are modelled with a logistic regression model that is a linear combination of covariate values.

The method proposed by Brown et al[11] is similar to that proposed earlier.[10] For the uncensored subjects, the target value is set to 0 as long as the subject is alive, and it is set to 1 when the subject encounters the event. The output values are unconstrained for all subsequent time intervals. However, a censored subject is considered to have survived to the end of the interval only if the subject has survived for more than half of the interval before being censored.

This approach is measurable and provides monotonic survival curves. However, some outputs are coded as 'not defined' and do not affect the error function.[12,13]

As a multiple time-point model, Street[14] proposed a neural network model for predicting the survival at 10 different time intervals, which makes the appropriate use of the censored observations. The output layer has a hyperbolic tangent activation function so that all output nodes vary between −1 and +1. In the output layer, each node represents the probability of survival for a corresponding time period. The first output unit represents recurrent patients in 1 year or less than 1 year following the surgery, whereas the second output unit represents recurrent patients between 1 and 2 or 3 years, up to 10 years.

Chi et al[15] proposed a variation in the Street method and compared the prediction results obtained from 2 breast cancer data sets. They used a multiple output neural network model with a sigmoid activation function for predicting the probability of recurrence at different time intervals for each patient. The target vector for the censored patients is derived using the K-M survival curves. The researchers also used the model to differentiate patients with 'good' (more than 5 years) and 'bad' (less than 5 years) prognoses. However, the study did not address a strategy for evaluating the performance of model in predicting the outcomes at different time intervals.

There are also a number of other approaches for modelling the hazard function. In this regard, an ANN model for predicting hazard as a function of covariates and time has been proposed.[16] Uncensored cases were repeated only for the time intervals in which they were actually observed. As a result, the failed subjects were not considered after the time interval of death. This approach, also referred to as the partial logistic artificial neural network (PLANN) model, has been applied by several authors. Other scholars have used the PLANN model in the Bayesian framework and applied it for oncological prognosis.[17–19] In addition, the PLANN methodology has been used for competing risks.[20–22]

There have been a significant number of publications that applied the ANN approaches in gastric cancer diagnosis and prognosis. Nevertheless, no study has been conducted on constructing an ANN strategy for modelling the censored survival data. A literature review of ANN models applied in gastric cancer prognosis demonstrated that in many studies the censored patients have been excluded from the data set or no clear strategies have been addressed for dealing with censoring.[23–26]

Accordingly, this study proposed a single time-point ANN model for predicting the probability of death for patients with gastric cancer, and a strategy is also suggested for dealing with censored data. Using the proposed strategy, the censored

subjects are not excluded from the data set. Five sets of multilayer feed-forward ANN models were constructed for predicting the survival of patients at 1, 2, 3, 4, and 5 years after surgery. Because the neural networks used are defined as single-output networks that predict the survival at only 1 time point, each of the time points is analysed by a separate neural network.

Materials and Methods

Data set description

The data set used in this study was obtained from a retrospective study on patients with confirmed gastric cancer conducted at the Research Center for Gastroenterology and Liver Disease at Shahid Beheshti University of Medical Sciences, Tehran, Iran. The data set consists of the records of 452 patients who underwent surgery at Taleghani Hospital, Tehran, Iran. The event of interest is defined as death, and losses or failure to follow-up is considered as censorship.

Patients who died had a survival time up to the point of death, and the survival time recorded for live patients spans from surgery (total or subtotal gastrectomy) to the last known follow-up date. For each patient, the survival time is recorded in months. The selection criterion includes patients with known survival time and survival status.

The data set was compiled using the demographic information of patients, which includes gender, age, marital status, education, ethnicity, main activity, as well as medical history and clinical and pathologic information concerning the tumour and its growth rate, such as tumour size, grading, pathologic stage, histologic type, and lymph node metastasis. The data set contains 20 predictor variables. A total of 161 (35.6%) patients died because of gastric cancer during the follow-up time, and 291 (64.4%) patients were censored by their last follow-up time. The median follow-up time was 11.90 months (range, 1-123 months).

Single time-point ANN model

This study proposes a single time-point ANN approach that can predict the survival probability of a patient at any predetermined point of time. A single time-point ANN model is efficient when prognosis is of interest at a specified time point. For example, in cancer studies, the survival rate is generally reported within 5 years of treatment. In public health studies involving the development of certain diseases or conditions, the event occurrence is generally observed within 10 years.[27] In such cases, a neural network structure can be developed for producing the outcome estimates at specific follow-up times. The proposed single time-point ANN approach has the advantage that no temporal structure such as proportional hazards is assumed in the model, which allows for flexible survival data modelling.

The single time-point ANN model presented in this study, also called a binary classification ANN, is essentially a logistic

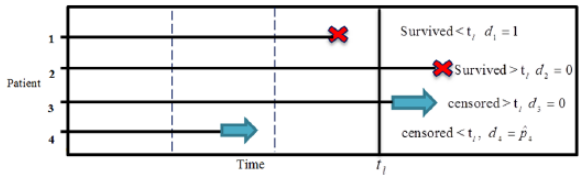

Figure 1. Target values for censored and uncensored patients.

regression analogue for binary classification tasks. In a binary classification problem, an input pattern is classified into 1 of the 2 nonoverlapped classes, say C_1 and C_2. The decision is based on estimating the conditional probability of observing an individual with class level C_1 given a set of covariate values.[28] In a linear logistic regression model, a single target variable Y is defined such that $Y = 1$ denotes class C_1 and $Y = 0$ denotes class C_2.

Let $p(x) = P(Y = 1 \mid X = x)$, the logistic regression model is expressed as follows:

$$p(x) = \frac{\exp(\beta_0 + \beta x)}{1 + \exp(\beta_0 + \beta x)} = \frac{1}{1 + \exp(\beta_0 + \beta x)} \quad (4)$$

Equivalently, the log odds, called the logit, show the linear relationship as follows:

$$\text{logit}[p(x)] = \log \frac{p(x)}{1 - p(x)} = \beta_0 + \beta_x, \quad (5)$$

where the unknown parameters β are called the 'regression coefficients' and are estimated using the maximum likelihood procedure. This equates the logit link function to the linear predictor. To minimise the misclassification rate, we should predict $Y = 1$ when $P \geqslant .5$ and $Y = 0$ when $P < .5$.

To implement the model using an ANN model, we considered the network that is presented in Figure 1. The network is set up in 3 interconnected layers: an input layer with nodes corresponding to the prognostic covariates, a hidden layer for modelling nonlinearity, and a single output layer that represents survival at a certain time point.

The network is composed of p input units, H hidden units, and 1 output unit. The units in the input layer $x = (x_1, \ldots, x_p)'$ correspond to prognostic covariates. Each unit in the hidden layer estimates a weighted sum of the input variables and the bias. By selecting a common activation function Λ for the hidden units, the output of the hidden unit h is given by

$$O_h = \Lambda\left(w_{0h} + \sum_{i=1}^{p} w_{ih} x_i\right) \quad i = 1, \ldots, p,\ h = 1, \ldots, H, \quad (6)$$

where w_{ih} is the weight from input i to the hidden unit h and w_{0h} is the bias for the hidden unit $h, h = 1, \ldots, H$. The

activation function used here is a logistic function given by $\Lambda(u) = 1/(1+\exp(-u))$.

During transmission to the output units, o_h is multiplied by weights w_h and a bias parameter w_0 is added to provide the net value:

$$\left(w_0 + \sum_{h=1}^{H} w_h o_h \right) \tag{7}$$

where w_h is the weight from the hidden node h to the output unit and w_0 is the bias for the output unit. The output unit value is calculated by applying another logistic activation function $\Lambda(\cdot)$ to the weighted sum of the hidden unit values and the bias. The network output is obtained as follows:

$$\begin{aligned} y = f(x,w) &= \Lambda\left(w_0 + \sum_{h=1}^{H} w_h o_h \right) \\ &= \Lambda\left(w_0 + \sum_{h=1}^{H} w_h \cdot \Lambda\left(w_{0h} + \sum_{i=1}^{p} w_i x_i \right) \right) \end{aligned} \tag{8}$$

Extending the single time-point ANN for modelling survival data. This section demonstrates how a single time-point ANN model can be developed for predicting the probability of death for survival data. We begin with some notation. Let T be a random variable that represents the time to death with the probability density function $f(t)$ and the cumulative distribution function $F(t)$. The survival function is given by $S(t) = \Pr(T > t) = 1 - F(T)$. The predicted probability of a patient dying before a fixed time t is then given by $\hat{F}(t) = 1 - \hat{S}(t)$.[29] Now, the primary objective is to develop an ANN model for predicting the probability of death in a single time interval $A_l = (0, t_l]$ as a function of the covariates x_i. The probability of an event in interval A_l is given by

$$P(T \leqslant t_l \mid x_i) = 1 - S(t_l \mid x_i) \tag{9}$$

In particular, a single time-point ANN model is applied for predicting the probability that a patient will die before the specific time t_l. To extend the model for predicting the probability of death for survival data, a binary target variable d_n is defined as the indicator of death occurring for the nth patient, which takes a value of 1 if the death occurs within the particular time period $A_l = (0, t_l]$, and 0 otherwise. The model corresponds to fitting of the logistic regression models to survival data.[30]

When no hidden layer is introduced in the calculation, the corresponding model is given by

$$\begin{aligned} f(x,w) &= P(T \leqslant t_l \mid x_i) \\ &= \Lambda\left(w_0 + \sum_{h=1}^{H} w_h \cdot \Lambda\left(w_{0h} + \sum_{i=1}^{p} w_{ih} x_i \right) \right) \end{aligned} \tag{10}$$

The rule of prediction for the nth patient with the covariate vector $x^{(n)} = (x_1^{(n)}, \ldots, x_p^{(n)})'$ is expressed as follows:

$$\begin{aligned} y^{(n)} &= f(x^{(n)}, w) = P(T \leqslant t_l \mid x_i) \\ &= \Lambda\left(w_0 + \sum_{h=1}^{H} w_h \cdot \Lambda\left(w_{0h} + \sum_{i=1}^{p} w_{ih} x_i^{(n)} \right) \right) \end{aligned} \tag{11}$$

The network output is defined as the probability of death before a specific time t_l. This model can be considered as a network that has been trained with death probabilities, and the predicted outputs are the probabilities of dying before a specific time point t_l. Obviously, such models can be repeatedly applied for the prediction of survival probability before fixed time points $t_1 < t_2 < \cdots < t_L$. Because the objective of this study is to predict 1-, 2-, 3-, 4-, and 5-year survival in patients with gastric cancer, 5 separate single time-point ANN models were developed for predicting the outcome of patients after 1, 2, 3, 4, and 5 years. The model was fitted using written codes in R-language. We used the nnet library described in an earlier work[12] with amendments to cope with the censored data.[28]

Modelling of censored data

Although the proposed ANN model can be applied to survival data, it has the drawback that all patients need to be observed until the event occurs. If patients are censored within A_l, their removal from the data set leads to biased estimates.[30] To deal with the problem of censored data, the following procedure is used.

In the absence of censoring, the target variable d_n represents the actual outcome (status) of the nth patient at a specific follow-up time. However, the existence of censoring requires some modifications in the fundamental method so that the target variable can be presented to the ANN model. For an uncensored patient, the target value is considered 1 if the patient dies before the given time point t_l, and 0 otherwise.

For a censored observation, the actual outcome is unknown, but only the censoring time is known. Due to this partial information, complications arise and it is unclear how to proceed because there is no target variable for training in this case.[31] For example, according to the study, it is known that the patient was alive for 1 year, but there is no information what happened after that. The simplest way to deal with censoring is to exclude the censored subjects for whom the censoring time is shorter than t_l and develop the model with the remaining noncensored data. However, excluding the censored patients reduces the number of training cases and influences the results.[12,32] Alternatively, an attempt is made for estimating the outcome of a patient whose censoring time was less than t_l. The Cox model was thus used for estimating the probability of death of the censored patients. For a prediction in interval $A_l = (0, t_l]$, the target variable can be expressed as follows:

$$d_n = \begin{cases} 1 & \text{if the event time} \leqslant t_l \\ 0 & \text{if the event or censoring time} > t_l \\ \hat{p}_n & \text{if the censoring time} \leqslant t_l \end{cases} \tag{12}$$

where $\hat{p}_n$ is the estimated probability of dying before t_l for the nth patient who was censored before time point t_l. The process of defining the target vector is explained in Figure 1. Figure 1 presents an example with 4 observations. For the first patient who failed before t_l, the target variable was equal to 1. For the second patient who survived beyond t_l, the target variable was equal to 0, indicating no event before t_l. For the third patient who was censored after t_l, the target value was equal to 0, indicating no event occurred before t_l. For the fourth patient, the censoring time was shorter than t_l, so the target value was imputed using the Cox model.

Training the Network

To optimise the performance of the network, the process of training a neural network adjusts the weight and bias parameter values.[33] The weights are estimated by minimising an appropriate error function. The sum of squares is the most commonly applied error function in many related studies. However, the cross-entropy error function is a more appropriate option for an error function,[7] which is given by equation (13), where d_n is the actual outcome for the nth patient and $y_n = f(w, x^{(n)})$ denotes the probability of death predicted by the ANN model for the nth patient:

$$E = -\sum_{n=1}^{N}\left[d_n \log y_n + (1-d_n)\log(1-y_n)\right] \qquad (13)$$

To minimise the cross-entropy error function, the Broyden-Fletcher-Goldfarb-Shanno quasi-Newton training algorithm[28] was used and thereby its efficiency and speed were proven in practice. In the quasi-Newton algorithm, the error function always decreases at each iteration unless the weight vector arrives at a local or a global minimum.[7]

K-fold cross-validation

In the *K*-fold cross-validation, the data set is randomly divided into *K* mutually exclusive subsets of approximately equal size. The most commonly used values of *K* are 5 and 10.[34] In this study, the 5-fold cross-validation was performed on our original data set for determining the error rate.[35] For the 5-fold cross-validation, the data set was divided into 5 disjoint subsets of approximately equal size. Four subsets were used for training the network in each run. The subset that was left out was used for testing the model and the results were recorded. Then, a different subset was selected to be left out, and the network was trained with all the other 4 subsets and tested with the excluded one. This process was repeated 5 times so that every subset was used once as a test set and all patients were tested. Finally, all results were combined to approximate the true error.

Hence, all data were used for training the network. Furthermore, the cross-validation is known to provide an unbiased estimation of the generalisation error rate.[36]

Model selection

The number of units in the hidden layer needs to be determined, which plays an important role in model accuracy.[37] In the proposed model, the optimal number of hidden units was determined by a constructive learning process. Training began with a network with no hidden units. One or more new hidden units were added to the hidden layer at a time.[6,38] The number of hidden units was increased progressively until the network performance began to deteriorate.[39]

In the development of the ANN model, overfitting is another challenge. The inability of the network to generalise the unseen data is known as overfitting. For an overfitted ANN model, the network performance is satisfactory for the training set; however, it significantly deteriorates when unseen data are presented to the network.

To generalise well and to avoid overfitting, the model was trained using a combination of regularisation and early stopping techniques.[7,40] If applied properly, early stopping and regularisation can ensure network generalisation.[33] As discussed above, for the proposed model, the weights were estimated by minimising the cross-entropy error function as presented in equation (13). Using weight decay as a regularisation technique, the error function was modified by adding a penalty term $\lambda\sum_{i,j} w_{ij}^2$, which is a multiple of the sum of the squares of weights w_{ij}. The coefficient λ is called the weight decay parameter. Adding a penalty term penalises large weights and improves the convergence of the optimisation algorithm.[16] Finally, a cross-validation technique was used for finding the best combination of the weight decay parameter λ and the number of hidden units at the same time. The weight decay parameter varied according to the number of hidden units in the model, usually over a range of 0.001 to 1.[29]

The network was fitted using the λ values between 0.001 and 0.1, whereas the number of hidden nodes ranged between 0 and 30. The combination of the number of hidden units and the weight decay parameter that achieved the highest accuracy was selected. The combination of the number of hidden units and the weight decay parameter that achieved the highest accuracy was selected. While selecting the best number of hidden units and the amount of weight decay, the model was trained several times on the data using different starting weights. Eventually, a network structure with 10 units in the hidden layer and a weight decay coefficient of 0.01 were selected to achieve the best performance. An alternative to weight decay as a means of controlling the effective network complexity is the procedure of early stopping. For early stopping, the maximum number of iterations was set to 1000.[41]

Prognostic variables

The sensitivity analysis has been used to find the best set of prognostic factors. As noted by many researchers in this field, most of the time, ANN models may offer better predictive

Table 1. Final set of predictor variables used by the artificial neural network (ANN) model.

VARIABLE	ATTRIBUTES	NO.	PERCENT
Age at diagnosis	⩽45	77	17
	>45	375	83
Tumour size	<35 mm	332	73.5
	⩾35 mm	120	26.5
Extent of wall penetration	T1	17	3.8
	T2	72	15.9
	T3	253	56
	T4	110	24.3
Regional lymph node metastasis	N1	126	27.9
	N2	257	56.9
	N3	69	15.3
Pathologic distance metastasis	M0	371	82.1
	M1	81	17.9
Tumour grade	Well differentiated	88	19.5
	Moderately differentiated	116	25.7
	Poorly differentiated	145	32.1
	Undifferentiated	103	22.8

ability but not much explanatory value.[42] This observation is generally true, but the sensitivity analysis can be performed for providing information about the relative importance of the input variables in predicting the output.

The fundamental notion is that the sensitivity analysis measures the predictor variables based on the change in model's performance that occurs if a predictor variable is not included in the model.[43] Table 1 illustrates the final set of the input variables that were selected through the sensitivity analysis for constructing the single time-point ANN model. The present sensitivity analysis results are based on 5 different ANN models developed for 5 data folds.

Using the sensitivity analysis, the most important leading factor in predicting the gastric cancer survivability was determined to be 'distant metastasis' followed by 'tumour size', 'regional lymph node metastasis', and 'age'. This result is consistent with some earlier studies that demonstrated these are important prognostic factors in the prediction of gastric cancer survival.[25,44–46] The 5 ANN models constructed for predicting the outcome at 5 different time points are in agreement with the variables that are most important for the prediction of outcome.

Validation Methods

In assessing the traditional survival analysis models, the detection of the departure from the underlying model assumptions and the selection of the best model with respect to the training data have been emphasised.[29] Regarding the neural network models, the main interest is to evaluate the prediction accuracy of the model.

Different criteria were used for evaluating the performance of the single time-point ANN model. According to this study, patients were classified as 'alive' if the event (death) did not occur during the time period considered and were classified as 'dead' otherwise. Several terms are commonly used for measuring the accuracy of the diagnosis model: true-positive (TP), true-negative (TN), false-negative (FN), and false-positive (FP).[47] These results are summarised in a confusion matrix, which represents the classification results.[48] For a 2-class prediction problem, the upper left cell (TP) denotes the number of the patients correctly classified as 'dead' and the lower right cell (TN) represents the number of the patients correctly classified as 'alive'.

The other 2 cells are FN and FP, which indicate the number of the misclassified patients.

Once the confusion matrix was formed, the accuracy (equation (14)), sensitivity (equation (15)), and specificity (equation (16)), as well as the positive predictive values (equation (17)) and the negative predictive values (equation (18)), were calculated to recognise the ability of the ANN model to classify the patients correctly[48,49]:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{14}$$

Table 2. Predicted and actual outcomes within 1, 2, and 3 years using single time-point ANN models.

CONFUSION MATRICES							
		1YEAR		2YEAR		3YEAR	
		ACTUAL		ACTUAL		ACTUAL	
		'DEAD'	'ALIVE'	'DEAD'	'ALIVE'	'DEAD'	'ALIVE'
Predicted by ANN	'Dead'	75	13	184	16	266	17
	'Alive'	31	333	34	218	31	138

Abbreviation: ANN, artificial neural network.

$$Sensitivity = \frac{TP}{TP + FN} \quad (15)$$

$$Specificity = \frac{TN}{TN + FP} \quad (16)$$

$$Positive\ predicted\ value = \frac{TP}{TP + FP} \quad (17)$$

$$Negative\ predicted\ value = \frac{TN}{TN + FN} \quad (18)$$

The proportion of the TPs that are correctly identified by the model is called sensitivity. It shows how good the model is at detecting a death. For example, 90% sensitivity means that when a test is conducted (using the ANN model) on a patient who has died, there is a 90% chance this patient will be identified as 'dead'. The proportion of the TNs correctly identified by the model is called specificity. It shows how well the model identifies the 'alive' (negative) condition. Accuracy is defined as the ratio of the correctly classified cases to the total number of the patients. A positive predictive value provides the proportion of the patients with the positive test results (predicted as 'dead') who have actually died.

The receiver operating characteristic (ROC) curve was also used to evaluate the predictive accuracy of the proposed ANN model,[50,51] which is generally used to measure the discriminatory power.[52] The ROC curve is indicated by a plot of sensitivity against 1 minus the specificity for different threshold values.[53] Overall, the diagnostic accuracy of an ROC curve is indicated by the area under the ROC curve (AUROC). The AUROC provides a method of measuring the accuracy of a diagnostic test. The larger the area, the more accurate the diagnostic test. Therefore, the AUROC values closer to 1 indicate a perfect model, whereas a value of 0.5 indicates that the predictor is no better than chance.[6,29,54] The cross-validation estimate of the overall AUROC is calculated as the average of the 5 individual AUROC measures obtained from 5 test sets.

The model calibration assesses a different aspect of prediction than the AUROC criteria.[55] The model calibration was checked using the Hosmer-Lemeshow (H-L) goodness-of-fit test.[56] The H-L test is a chi-square goodness-of-fit test that determines the degree of agreement between the probabilities generated by the model and the actual outcomes. All subjects were sorted based on their predicted probabilities of death and were divided into deciles. The chi-square statistic test was then calculated based on the differences found between the observed and the expected survival probabilities in each decile. The χ^2 and P value are calculated for a given significant level α. A value of P greater than .05 indicates no significant difference between the predicted and the observed survival probabilities,[57] whereas a higher P value defines a model with higher calibration.

Results and Discussion

The objective of the single time-point model is to predict the probability of death at specific time points. Five separate single time-point ANN models were used to predict the probability of death within 1 to 5 years. The occurrence of an event (death) was coded as 1 and the absence of the event was coded as 0. The model outputs included a set of estimated probabilities of death for each patient in the data set. At a given time point, the patients were classified as 'alive' if no death occurred during that time period or as 'dead' otherwise. For a particular patient, according to the classification rule, 'dead' means that the network output is greater than a cut-off level, say 0.5, or 'alive' otherwise.

First, the model was evaluated for a number of misclassifications. Table 2 displays the confusion matrix for predictions within 1 to 3 years.

Table 2 demonstrates that out of 452 patients, 106 (75 + 31) patients died before 1 year and 346 (333 + 13) patients survived past 1 year. The model accurately classifies 333 nonsurvivors and 75 survivors with an accuracy of 90.3%. Table 2 shows the predicted numbers of deaths before 2 years with an accuracy of 88.9%.

In 2 years, 218 (184 + 34) patients died and 234 (16 + 218) patients survived. The ANN model accurately predicted 184 nonsurvivors and 218 survivors with an accuracy of 88.9%. In 3 years, 297 (266 + 31) patients died and 155 (138 + 17) patients survived. The model accurately classified 266 nonsurvivors and 138 survivors with an accuracy of 89.4%. Table 3 shows the predicted numbers of deaths within 4 and 5 years.

Table 3. Predicted and actual outcomes within 4 and 5 years using single time-point ANN models.

CONFUSION MATRICES					
		4 YEARS		5 YEARS	
		ACTUAL		ACTUAL	
		'DEAD'	'ALIVE'	'DEAD'	'ALIVE'
Predicted by ANN	'Dead'	309	21	371	17
	'Alive'	30	92	30	34

Abbreviation: ANN, artificial neural network.

Table 4. Performance of single time-point ANN model in predicting survival at 1, 2, and 3 years obtained from the aggregation of 5 test folds.

MEASURE	PREDICTION OF 1-YEAR SURVIVAL	PREDICTION OF 2-YEAR SURVIVAL	PREDICTION OF 3-YEAR SURVIVAL
Accuracy	0.903	0.889	0.894
95% CI for accuracy	(0.871-0.928)	(0.857-0.917)	(0.862-0.921)
Sensitivity	0.707	0.844	0.896
Specificity	0.962	0.932	0.890
Positive predictive value	0.852	0.920	0.940
Negative predictive value	0.915	0.865	0.817
AUROC	0.962	0.958	0.973
H-L test	$P > .05$	$P > .05$	$P > .05$

Abbreviations: ANN, artificial neural network; AUROC, area under the receiver operating characteristic curve; CI, confidence interval; H-L, Hosmer-Lemeshow goodness-of-fit statistics.

The models were also evaluated based on the accuracy measures. The results were obtained using 5-fold cross-validation and are based on the average results obtained from the 5 test sets.

As shown in Table 4, for 1-year prediction, the proposed ANN model obtained the classification accuracy of 90.3% with a sensitivity of 70.7% and a specificity of 96.2%, which indicates that the model was more successful in predicting surviving than nonsurviving patients. The positive and negative predictive values were 85.2% and 91.5%, respectively. The AUROC curve was 0.96, which is an example of a perfect model.[54]

For 2-year prediction, the single time-point ANN model obtained 88.9% accuracy with a sensitivity of 84.4% and a specificity of 93.2%. The positive predictive value obtained was 0.92. The negative predictive value obtained was 86.5%, indicating that among patients predicted by the model as 'alive', 86.5% were actually alive. The AUROC value was 0.958.

The network constructed to predict survival within 3 years achieved 89.4% classification accuracy with a sensitivity of 89.6%, a specificity of 89.0%, and an AUROC value of 0.973. The positive and negative predictive values obtained indicate that among those predicted as 'dead' 94.0% had actually died and among those predicted as 'alive' 81.7% were actually alive. Table 5 shows the performance of the model in predicting survival within 4 and 5 years.

The single time-point model obtained an accuracy of 88.7% for the prediction within 4 years. The sensitivity and specificity were 91.1% and 81.4%, respectively. The AUROC value obtained was 0.950, which is an example of a perfect model. The proposed model also obtained 89.6% accuracy for prediction within 5 years. The sensitivity and specificity were 92.5% and 66.7%, respectively, and AUROC value obtained was 0.943.

According to the accuracy and the AUROC values, the performance of the single time-point ANN model in predicting the probabilities of death appears to be consistently high for all time points.

An improvement was observed in the sensitivity and positive predictive values of the third and fourth models (predictions within 3 and 4 years) compared with the 2 former models. This can be justified by the fact that the higher number of patients who experienced death at later time points reinforced the ability of the models to identify death (the event) and also enhanced the sensitivity of the predictive models. For the 5-year time point, the ANN model showed low specificity and negative predictive value, which indicates that the model was less successful in identifying no event ('alive') rather than an event ('dead'). This could be due to the lower number of the patients who were still alive by the fifth year. In addition, the results of the H-L test indicate that all models were well fitted ($P > .05$) and there was no evidence of lack of fit.

Table 5. Performance of single time-point ANN model in predicting survival within 4 and 5 years obtained from the aggregation of 5 test folds.

MEASURE	PREDICTION OF 4-YEAR SURVIVAL	PREDICTION OF 5-YEAR SURVIVAL
Accuracy	0.887	0.896
95% CI for accuracy	(0.854-0.915)	(0.864-0.923)
Sensitivity	0.911	0.925
Specificity	0.814	0.667
Positive predictive value	0.936	0.956
Negative predictive value	0.754	0.531
AUROC	0.950	0.943
H-L statistic	$P > .05$	$P > .05$

Abbreviations: ANN, artificial neural network; AUROC, area under the receiver operating characteristic curve; CI, confidence interval; H-L, Hosmer-Lemeshow goodness-of-fit statistics.

To refer a reliable model, the survival rates predicted by the ANN model should be monotonically decreasing at 1, 2, 3, 4, and 5 years. In this regard, a survival curve was obtained from the aggregation of 5 single time-point ANN models, which displays the predicted 1- to 5-year survival rates. As shown in Figure 2, the survival curve obtained from the aggregation of the 5 single time-point ANN models resulted in a monotonically decreasing survival curve.

The proposed ANN model can provide survival predictions for individual patient. In the medical context, such information is valuable for both clinicians and patients. Patients at a high risk of dying could be followed up more frequently than those at a lower risk so that the valuable resources are aligned to those who need them the most. Obtaining information about the prognosis of the patients is also valuable for planning their life.[17]

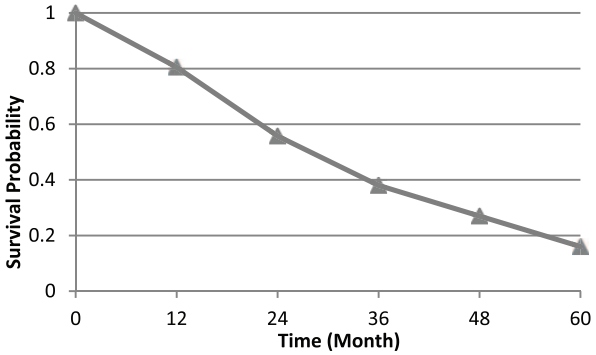

Figure 2. Survival curve obtained from the aggregation of 5 single time-point artificial neural networks.

Conclusions

Considering many diseases, the prediction of the probable survival of patients can be a challenging objective.[31] For different types of cancer, the estimated risk of the patient can directly affect the choice of treatment. However, some investigations usually aim only at finding the relative importance of the prognostic factors or comparing the performance of ANN models with the conventional analysis methods; little effort has been put to apply the ANN methodology to censored data modelling. This study focuses on developing an efficient ANN structure with the ability to handle censored data. Five sets of single time-point feed-forward ANN models were developed for predicting the outcome for gastric cancer patients at 1, 2, 3, 4, and 5 years after surgery. Each network had a single output that represents survival at a certain time point. Applying a 5-fold cross-validation technique helped using all patients in the data set for both model training and testing.

As a result, the ANN prediction models displayed accuracy ranging from 88.7% to 90.2% with sensitivity ranging from 70.2% to 92.5% and specificity ranging from 66.7% to 96.2%. The reported AUROC values more than 0.9 showed that the model was consistently accurate in predicting the survival of patients with gastric cancer within 1 to 5 years. The proposed strategy also indicates that the predictions were based on individual predictor variables, as well as the possible complex interactions between covariates because the interaction terms were intrinsically incorporated in the neural network architecture.

One of the concerns regarding the aggregation of single time-point ANN models is that the survival curve obtained from aggregating several ANN models may not be a monotonic function of time.[12] For example, the estimated survival rate within 4 years may be lower than that for 5 years. However, as shown in Figure 2, the aggregations of 5 single time-point neural networks resulted in a monotonically decreasing survival curve.

For an individual patient, the predicted survival probability is rarely equal to the observed probability. However, the predicted probabilities should follow the observed probabilities depending on how well the model performs. The H-L test results showed a high degree of agreement between the predicted and the observed survival rates.

This study demonstrates the feasibility of applying ANN models in medical decision support systems that use clinical

data sets for predicting the survival of patients with gastric cancer. The clinical application of the proposed ANN prediction models can potentially improve prognosis accuracy and treatment decisions for the patients. The development of better clinical decision support systems for gastric cancer prognosis could decrease uncertainty in prognosis, thus allowing treatment to be focused on patients with the worst expected survival chances.

Author Contributions

HND implemented the software and designed the experiments. HND and MRA wrote the first draft of the manuscript. MBA, JA, and MAP contributed to the writing and revising of the manuscript. MRA made critical revisions and approved the final draft. All authors reviewed and approved the final manuscript.

Disclosures and Ethics

As a requirement of publication, authors have provided to the publisher signed confirmation of compliance with legal and ethical obligations including, but not limited to, the following: authorship and contributorship, conflicts of interest, privacy and confidentiality, and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any disclosures are made in this section. The external blind peer reviewers report no conflicts of interest.

REFERENCES

1. McLean MH, El-Omar EM. Genetics of gastric cancer. *Nat Rev Gastroenterol Hepatol*. 2014;11:664–674.
2. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer*. 2013;49:1374–1403.
3. Bass AJ, Thorsson V, Shmulevich I, et al. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. 2014;513:202–209.
4. Bethesda MD. SEER cancer statistics review (CSR) 1975–2013. http://seer.cancer.gov/csr/1975_2013/. Published November 2015, Updated April 2016.
5. Ohno-Machado L. Sequential use of neural networks for survival prediction in AIDS. *Proc AMIA Annu Fall Symp*. 1996;1996:170–174.
6. Jerez JM, Franco L, Alba E, et al. Improvement of breast cancer relapse prediction in high risk intervals using artificial neural networks. *Breast Cancer Res Treat*. 2005;94:265–272.
7. Bishop CM. *Pattern Recognition and Machine Learning*. New York, NY: Springer; 2006.
8. Grossi E, Mancini A, Buscema M. International experience on the use of artificial neural networks in gastroenterology. *Dig Liver Dis*. 2007;39:278–285.
9. Ravdin PM, Clark GM. A practical application of neural network analysis for predicting outcome of individual breast cancer patients. *Breast Cancer Res Treat*. 1992;22:285–293.
10. Liestbl K, Andersen PK, Andersen U. Survival analysis and neural nets. *Stat Med*. 1994;13:1189–1200.
11. Brown SF, Branford AJ, Moran W. On the use of artificial neural networks for the analysis of survival data. *IEEE Trans Neural Netw*. 1997;8:1071–1077.
12. Baesens B, Van Gestel T, Stepanova M, Van den Poel D, Vanthienen J. Neural network survival analysis for personal loan data. *J Oper Res Soc*. 2005;56:1089–1098.
13. Yang Y. *Neural Network Survival Analysis* [master's dissertation]. Gent, Belgium: Universiteit Gent; 2010.
14. Street WN. A neural network model for prognostic prediction. In: Proceedings of the Fifteenth International Conference on Machine Learning; July 24-27, 1998; Madison, WI.
15. Chi C, Street WN, Wolberg WH. Application of artificial neural network-based survival analysis on two breast cancer datasets. *AMIA Annu Symp Proc*. 2007;2007:130–134.
16. Biganzoli E, Boracchi P, Mariani L, Marubini E. Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach. *Stat Med*. 1998;17:1169–1186.
17. Eleuteri A, Aung MSH, Taktak AFG, Damato B, Lisboa PJG. Continuous and discrete time survival analysis: neural network approaches. *Conf Proc IEEE Eng Med Biol Soc*. 2007;2007:5420–5423.
18. Lisboa PJG, Wong H. Are neural networks best used to help logistic regression? an example from breast cancer survival analysis. *Proc Int J Conf Neural Netw* (Cat No. 01CH37222). 2001;4:2472–2477.
19. Lisboa PJG, Vellido A, Wong H. Bias reduction in skewed binary classification with Bayesian neural networks. *Neural Netw*. 2000;13:407–410.
20. Lisboa PJG, Etchells TA, Jarman IH, et al. Partial logistic artificial neural network for competing risks regularized with automatic relevance determination. *IEEE Trans Neural Netw*. 2009;20:1403–1416.
21. Arsene CTC, Lisboa PJ, Biganzoli E. Model selection with PLANN-CR-ARD. In: Rojas I, Joya G, Cabestany J, eds. *Advances in Computational Intelligence*. Berlin, Germany: Springer; 2011:210–219.
22. Arsene CTC, Lisboa PJ. Bayesian neural network applied in medical survival analysis of primary biliary cirrhosis. In: Proceedings of the 2012 UKSim 14th International Conference on Computer Modelling and Simulation (UKSim); March 28-30, 2012; Cambridge, UK.
23. Biglarian A, Hajizadeh E, Kazemnejad A, Zali MR. Application of artificial neural network in predicting the survival rate of gastric cancer patients. *Iran J Public Health*. 2011;40:80–86.
24. Amiri Z, Mohammad K, Mahmoudi M, Parsaeian M, Zeraati H. Assessing the effect of quantitative and qualitative predictors on gastric cancer individuals survival using hierarchical artificial neural network models. *Iran Red Crescent Med J*. 2013;15:42–48.
25. Zhu L, Luo W, Su M, et al. Comparison between artificial neural network and Cox regression model in predicting the survival rate of gastric cancer patients. *Biomed Rep*. 2013;1:757–760.
26. Ohno-Machado L. Modeling medical prognosis: survival analysis techniques. *J Biomed Inform*. 2001;34:428–439.
27. Venables WN, Ripley BD. *Modern Applied Statistics With S-PLUS*. Berlin, Germany: Springer Science + Business Media; 2013.
28. Ripley RM, Harris AL, Tarassenko L. Neural network models for breast cancer prognosis. *Neu Comput Appl*. 1998;7:367–375.
29. Schwarzer G, Vach W, Schumacher M. On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology. *Stat Med*. 2000;19:541–561.
30. Biganzoli E, Boracchi P, Marubini E. A general framework for neural network models on censored survival data. *Neural Netw*. 2002;15:209–218.
31. Kalderstam J. *Neural Network Approaches to Survival Analysis* [dissertation]. Lund, Sweden: Lund University; 2015.
32. Ripley RM, Harris AL, Tarassenko L. Non-linear survival analysis using neural networks. *Stat Med*. 2004;23:825–842.
33. Beale MH, Hagan MT, Demuth HB. *Neural Network Toolbox 7 User's Guide*. Natick, MA: MathWorks; 2010.
34. Hassan SS, Ruusuvuori P, Latonen L, Huttunen H. Flow cytometry-based classification in cancer research: a view on feature selection. *Cancer Inform*. 2016;14:75–85.
35. Stansfield JC, Rusay M, Shan R, et al. Toward signaling-driven biomarkers immune to normal tissue contamination. *Cancer Inform*. 2016;15:15–21.
36. Dreyfus G. *Neural Networks: Methodology and Applications*. Berlin, Germany: Springer Science + Business Media; 2005.
37. Joshi R, Reeves C, Johnston C. Probabilistic artificial neural networks for malignant melanoma prognosis. In: Ribeiro B, Albrecht R, Dobnikar A, Pearson D, Steele N, eds. *Adaptive and Natural Computing Algorithms*. Vienna, Austria: Springer; 2005:538–541.
38. Ripley BD. *Pattern Recognition and Neural Networks*. Cambridge, UK: Cambridge University Press; 1996.
39. Cervellera C, Macciò D, Muselli M. Deterministic learning for maximum-likelihood estimation through neural networks. *IEEE Trans Neural Netw*. 2008;19:1456–1467.
40. Dolgobrodov SG, Moore P, Marshall R, Bittern R, Steele RJC, Cuschieri A. Artificial neural network: predicted vs. observed survival in patients with colonic cancer. *Dis Colon Rectum*. 2007;50:184–191.
41. Yu D, Yao K, Su H, Li G, Seide F. KL-divergence regularized deep neural network adaptation for improved large vocabulary speech recognition. In:

Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); May 26, 2013; Vancouver, BC.
42. Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. *Artif Intell Med*. 2005;34:113–127.
43. Oztekin A, Delen D, Turkyilmaz A, Zaim S. A machine learning-based usability evaluation method for eLearning systems. *Decis Support Syst*. 2013;56:63–73.
44. Biglarian A, Hajizadeh E, Kazemnejad A, Zayeri F. Determining of prognostic factors in gastric cancer patients using artificial neural networks. *Asian Pac J Cancer Prev*. 2010;11:533–536.
45. Pourhoseingholi MA, Moghimi-Dehkordi B, Safaee A, Hajizadeh E, Solhpour A, Zali MR. Prognostic factors in gastric cancer using log-normal censored regression mode. *Indian J Med Res*. 2009;129:262–267.
46. Kim JG, Ryoo B-Y, Park YH, et al. Prognostic factors for survival of patients with advanced gastric cancer treated with cisplatin-based chemotherapy. *Cancer Chemother Pharmacol*. 2008;61:301–307.
47. Zhu W, Zeng N, Wang N. Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. In: Proceedings of the SAS Conference, *November 14 2010*; Baltimore, MD. 2010.
48. Oztekin A, Delen D, Kong ZJ. Predicting the graft survival for heart-lung transplantation patients: an integrated data mining methodology. *Int J Med Inform*. 2009;78:84–96.
49. Mofidi R, Deans C, Duff MD, de Beaux AC, Paterson Brown S. Prediction of survival from carcinoma of oesophagus and oesophago-gastric junction following surgical resection using an artificial neural network. *Eur J Surg Oncol*. 2006;32:533–539.
50. Bernstein IH. *Applied Multivariate Analysis*. Berlin, Germany: Springer Science + Business Media; 2012.
51. Swets J, Pickett R. *Evaluation of Diagnostic Systems*. Amsterdam, The Netherlands: Elsevier; 1982.
52. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *J Biomed Inform*. 2002;35:352–359.
53. Bourdès V, Bonnevay S, Lisboa P, et al. Comparison of artificial neural network with logistic regression as classification models for variable selection for prediction of breast cancer patient outcomes. *Adv Artif Neural Syst*. 2010;2010:1–11.
54. Dhande JD, Gulhane SM. Design of classifier using artificial neural network for patients survival analysis. *Int J Eng Sci Innovative Technol*. 2012;1:278–282.
55. Lin RS, Horn SD, Hurdle JF, Goldfarb-Rumyantzev AS. Single and multiple time-point prediction models in kidney transplant outcomes. *J Biomed Inform*. 2008;41:944–952.
56. Hosmer DW Jr, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. Vol 398. John Wiley & Sons; 2013.
57. Jaimes F, Farbiarz J, Alvarez D, Martínez C. Comparison between logistic regression and neural networks to predict death in patients with suspected sepsis in the emergency room. *Crit Care (London, England)*. 2005;9:R150–R156.