Original Article

Institution of MECHANICAL ENGINEERS
Journal of ENGINEERING IN MEDICINE

Proc IMechE Part H:
J Engineering in Medicine
2014, Vol. 228(9) 857–866

DOI: 10.1177/0954411914549980
pih.sagepub.com
SAGE

Bayesian neural network approach for determining the risk of re-intervention after endovascular aortic aneurysm repair

Omneya Attallah[1,2] and Xianghong Ma[2]

Abstract

This article proposes a Bayesian neural network approach to determine the risk of re-intervention after endovascular aortic aneurysm repair surgery. The target of proposed technique is to determine which patients have high chance to re-intervention (high-risk patients) and which are not (low-risk patients) after 5 years of the surgery. Two censored datasets relating to the clinical conditions of aortic aneurysms have been collected from two different vascular centers in the United Kingdom. A Bayesian network was first employed to solve the censoring issue in the datasets. Then, a back propagation neural network model was built using the uncensored data of the first center to predict re-intervention on the second center and classify the patients into high-risk and low-risk groups. Kaplan–Meier curves were plotted for each group of patients separately to show whether there is a significant difference between the two risk groups. Finally, the logrank test was applied to determine whether the neural network model was capable of predicting and distinguishing between the two risk groups. The results show that the Bayesian network used for uncensoring the data has improved the performance of the neural networks that were built for the two centers separately. More importantly, the neural network that was trained with uncensored data of the first center was able to predict and discriminate between groups of low risk and high risk of re-intervention after 5 years of endovascular aortic aneurysm surgery at center 2 ($p = 0.0037$ in the logrank test).

Keywords

Survival analysis, censored data, Kaplan–Meier curve, Bayesian network, endovascular aortic aneurysm repair, back propagation neural network

Date received: 22 October 2013; accepted: 1 July 2014

Introduction

Survival analysis is a statistical method used for analyzing data which contains a variable describing the time of survival,[1] which is also known as lifetime data analysis, reliability analysis, time to event analysis, and event history analysis.[2] Recently, survival techniques have been extensively used in medical applications in order to determine the probability of patients' survival or the time till an event of interest occurs. This event could be death, recurrence of a disease, discharge from hospital, or surgical re-intervention (REINT). However, there is high probability of the presence of censoring in the medical datasets. Censored data means the information about the time to an event for some patients is not available, the only available information is the time till death (when the event of interest is not the death) or the last follow-up which is known as censoring time. Reasons for censoring include patients that die due to other reasons than that of the disease under study, some patients drop out of the study during the follow-up period,[3,4] or the event of interest does not occur at the end of study period. Right and left censoring are two well-known types of censoring.[5] Patients who were

[1]Department of Electronics and Communications Engineering, Arab Academy for Science, Technology & Maritime Transport, Alexandria, Egypt
[2]School of Engineering and Applied Science, Aston University, Birmingham, UK

Corresponding author:
Xianghong Ma, School of Engineering and Applied Science, Aston University, Birmingham B4 7ET, UK.
Email: x.ma@aston.ac.uk

censored before the study observation time are considered to be left censored. Those who were censored during or after the study period are called right-censored cases.

Right censorship is the most common type of censoring that appears in most medical survival datasets. There are two types of right-censored data. In type 1 data, patients die or leave the follow-up during the study duration, so the time to the event is unknown. Type 2 data have patients who have completed their follow-up observations till the end of the study and do not experience the event, so the time to the event is also unknown. Even though the event is less likely to occur at a time after the end of the study period, they are considered as censored as well; as it is not sure whether the event of interest will happen to these patients or not, the only thing we are certain of is that the event did not appear till the end of the observation period. Most of the articles discussing and using survival analysis methods[2,6–13] considered any patient who does not experience the event of interest during the whole duration of the study as censored. In this article, we will follow on this definition. However, some articles such as Zupan et al.[14] and Delen et al.[15] state that patients that either stayed for a long time during the study period or completed their follow-up observations and the event of interest did not occurred to them as event-free patients, and they cannot be considered as censored.

Survival analysis is used with medical datasets to solve the censoring issue. It is useful not to ignore any instance in the dataset even if the event had not occurred (censored) as this may lead to biasing the dataset.[1]

Endovascular aortic aneurysm repair (EVAR) is a type of surgical intervention used to fix abnormality caused by ballooning of aorta (aneurysm). Aorta is the main and biggest blood vessel connected to heart. Sometimes, the aorta wall becomes too weak which leads to its ballooning, this phenomenon is known as aortic aneurysm. EVAR carries significantly lower operative risk than the traditional open-repair surgery; therefore, it is preferred by patients and recommended by medical guidelines as the choice for treating abdominal aortic aneurysm (AAA).[16] Currently, EVAR is followed by lifelong surveillance which not only is expensive but also exposes patients to radiation exposure and contrast nephropathy. Majority of complications requiring treatment are missed despite the cost and regular surveillance schedule.[17,18] This research is aimed to develop and validate a prediction system for aortic complications after EVAR. It will enable clinicians to determine which patients have higher chance of requiring REINT (high-risk group), therefore a more regular monitoring schedule. The lower chance patients can be monitored less regularly (low-risk group). This prediction system is achieved by the machine-learning survival analysis methods including Bayesian network and back propagation (BP) neural network methods.

Two prospectively maintained databases, which included all patients undergoing EVAR of infra-renal AAA at two tertiary vascular centers in the United Kingdom, were collected. Detailed report of the datasets can be found in Karthikesalingam et al.[18] The variables of dataset include aneurysm morphology measured from patients' computed tomography (CT) and physiology parameters. The datasets consist of 24 morphological and 12 physiological features, respectively. Centers 1 and 2 have 464 and 259 patient samples, respectively. Only about 9% of the patients (42 and 22 patients of centers 1 and 2, respectively) did the REINT (which is the event of interest) during the 7 years of observations, and their targets were set equal to one. According to the censorship definition that was used in most of the work done in survival analysis as discussed previously, both datasets are considered to be highly censored, they contain 91% (422 and 237 patients for centers 1 and 2, respectively) of the patients who did not undergo the surgical intervention. Only 62 and 20 of these patients have censoring time greater than 5 years for centers 1 and 2, respectively (can be considered as event free or less likely to do REINT). So, even if we were to consider the event-free patients as uncensored, we would still have the majority (360 and 217 for centers 1 and 2, respectively) patients as censored.

Censored patients with zero targets, meaning REINT, did not occur to those patients due to various reasons. As the worst case they could be dead as a result of surgery or other reason (with short censoring time) or best could be recovering well after the surgery, therefore no need for REINT during the study period (with long censoring time). The long censoring time patients are considered as censored as they might need REINT after the study period and the time for the operation is unknown (they may be considered as patients with lower probabilities to redo the surgery). For the short censoring time patients, we do not know which spectrum they belong to. The only information on hand is their censoring time and the corresponding features till this time. These censored patients cannot be ignored or deleted from the datasets. As by doing this, huge amount of information will be lost which leads to biasing the prediction targets when calculating the probability of surgical REINT of EVAR. Censoring is an important cause for unsuccessful survival models built by standard machine-learning techniques, as censored patients cannot be all considered as real zero targets (which means that the event of interest definitely has not happened). By doing so, it will bias the predictive model toward the zero targets;[19] this phenomenon will also be shown in section "Results" later. The challenge in this work is how to solve the issue of high censoring in the datasets, more specifically, in estimating the probability that each censored patient belongs to either high- or low-risk group (uncensoring the dataset). After

the datasets are uncensored, classification techniques can then be used to build a model capable of predicting the low- and high-risk groups.

Machine-learning techniques have been widely applied in the field of survival analysis in clinical trials. They are preferred over the standard statistical models such as Cox proportional hazard model as they can detect complex relations between data which lead to better prediction. Additionally, they are capable of capturing nonlinearities between the input variables and the targets of a dataset.[20] The two most popular techniques are Bayesian network and artificial neural networks (ANN). Bayesian network is a probabilistic network that uses probability theory to calculate joint probabilities between variables. It is also considered as a graphical model since it looks like a graph representing relations between the nodes (also called vertices) which correspond to input variables.[21] A Bayesian network was used by Sebastiani et al.[22] to build a model that predicts the risk of death from sickle-cell disease within 5 years and by Steele et al.[23] for estimating preoperative risk of *Clostridium difficile* infection following colon surgery. Kaderali et al.[6] combined it with Cox regression model to predict survival times and select the features that relate to survival. The algorithm was applied to two types of cancer patients. Štajduhar and Dalbelo-Bašić (S&DB)[33] proposed a technique that uses Bayesian network and likelihood information to build a model for uncensoring the dataset. It is a preprocessing step after which any machine-learning technique can be used for classification. Naive Bayesian and decision tree classifiers were used to predict breast and skin cancer prognoses.

ANN has been used widely in building prediction models for medical conditions. For example, in the study by Damato et al.,[24] a model was built for predicting the survivability of patients after being treated from choroidal melanoma. While in the study by Taktak et al.,[25] ANN is applied to estimate the survivability curves of intraocular melanoma patients after treatment. Moreover, in several other studies, it was employed for predicting survivability or recurrence for several types of cancers such breast, liver, and colorectal cancers.[7,8,15,26–29]

Although Bayesian network and ANN have been successfully applied in the above applications, it is still a challenge to apply them in this study. The reason is that the previous applications only dealt with low- or medium-censored datasets, while our dataset is highly censored with more than 90% of the total patient samples. Also, in previous studies, performances of the classifiers used were not tested for other datasets collected from different centers or hospitals. Usually, measuring the performance of a classifier was carried out either through a test set collected from the same center and not used in building and training the classifier[3,7,8,15,17–24,26–29] or through a 10-fold cross-validation test applied for the whole dataset.[14] Yet the cross-center testing is important if the model built is going to be valid for wider applications.

This article first presents a new modified approach to solve the censoring problem in the datasets. Then, the uncensored data of the first center were used to build a neural network model in order to determine the risk of REINT after EVAR and classify patients of the second center into high-risk and low-risk groups. Next, Kaplan–Meier curves were plotted for each group of patients separately. Finally, the logrank test was applied to determine whether the neural network model was capable of predicting and distinguishing between the two risk groups.

Methodology

Data acquisition

Follow-up observations to patients going through EVAR surgery were taken, and their data were collected during the study period 2004–2010 from two different vascular hospitals. The two datasets contain details of operative procedure and patient morphological and physiological features. Pre-surgical morphology variables were measured using three-dimensional CT. The CT images had a slice thickness of 0.625 or 1.25 mm and were acquired from the thoracic inlet to the level of the common femoral artery bifurcation. Only morphological features were used for constructing the prediction models as they have greater effect on aortic complications than physiological ones.[30–32]

Uncensoring approach

In trying to solve the problem of censored data, several survival analysis techniques were studied. First, the technique based on S&DB[33] was applied to the EVAR datasets. All patients who did not undergo the surgical REINT are used to build a Bayesian network called censored Bayesian network. This network includes patients with short censoring time and others with longer censoring time greater than 5 years, who have lower chances to do the REINT. Patients who did the REINT are used to build another network called high-risk (re-intervened) Bayesian network. As a result of the highly censored EVAR datasets, the constructed censored Bayesian network expressed both the inherent distributions of low-risk (less chance to do the REINT) and high-risk groups instead of the low-risk group alone. The S&DB algorithm has slightly improved the prediction accuracy of each dataset. However, the corresponding neural network model constructed with the first center was not capable of predicting REINT in the second center which will be shown later in section "Results." In order to overcome this, a new modified approach was proposed to deal with the high censoring of the datasets. The new method is able to determine the risk of surgical REINT to patients after 5 years

from undergoing an EVAR operation and classify them into high- and low-risk groups.

The new approach (Attallah and Ma (A&M) approach) is discussed in detail: first, Kaplan–Meier curves (illustrated in section "Kaplan–Meier curves") are plotted for each EVAR center separately and are shown in Figure 1. Afterward, the censoring time of each EVAR center dataset is used to divide patients into three groups. The first one belongs to patients who experience the REINT at a time lower than or equal to 5 years (re-intervened or high-risk patients which are 42 and 22 patients for centers 1 and 2, respectively). The second group refers to patients who did not need the REINT till a time greater than or equal to 5 years (62 and 20 patients for centers 1 and 2, respectively). Because they have low chance of needing the REINT, we call it low-risk group. Finally, the third group is the rest of the patients who are considered as censored with short censoring time. The main difference between the A&M algorithm and the S&DB approach is that S&DB uses patients with both short and long censoring time (lower and greater than 5 years) to build the so-called censored Bayesian network which is used later for uncensoring the data, while the A&M method only uses patients with long censoring time which is greater than 5 (low-risk group) to construct a Bayesian network called low-risk Bayesian network which is employed later for the uncensoring process to differentiate the short censoring time patients into low- and high-risk groups.

In the new proposed approach, the dataset is discretized using an unsupervised discretization technique. Then, the low- and high-risk groups are used to build two separate Bayesian networks called the low- and high-risk Bayesian networks B^{low} and B^{high}, respectively. Targets of low- and high-risk patients can now be omitted from the dataset, as its knowledge is already embedded in the two networks generated. Next, each censored instance in the third group is compared with the inherent distribution of high-risk group (p^{high}) and inherent distribution of low-risk group (p^{low}). This is done by calculating the likelihood that the instance (class omitted) was sampled from either model. Likelihood $\ell(x_c/p)$ is calculated from equations (1) and (2) by multiplying all the exact probabilities from the probability tables, following the network topology. Both Bayesian networks consist of ξ which indicates that it is a directed acyclic graph (DAG) network, and the number of nodes is represented as V_i. Each V represents a variable of the dataset, and π is the parent of this node V

$$\hat{\ell}(x_c/p^{high}) = \ell(x_c/B^{high}) = p(x_c/\xi^{high}, p^{high}) = \prod_{i=1}^{n} p^{high}(V_i/\pi(V_i)) \quad (1)$$

$$\hat{\ell}(x_c/p^{low}) = \ell(x_c/B^{low}) = p(x_c/\xi^{low}, p^{low}) = \prod_{i=1}^{n} p^{low}(V_i/\pi(V_i)) \quad (2)$$

The posterior probability $P(O/x_c)$, which is the output prediction (O) given that it is censored (x_c), is calculated using equations (3) and (4), then normalized to ignore the effect of probability of a censored instance $P(x_c)$ using equation (5)

$$P(O^{high}/x_c) = \hat{P}(O^{high}) * \frac{\hat{\ell}(x_c/p^{high})}{P(x_c)} \quad (3)$$

$$P(O^{low}/x_c) = \hat{P}(O^{low}) * \frac{\hat{\ell}(x_c/p^{low})}{P(x_c)} \quad (4)$$

$$P(O/x_c) = \hat{P}(O) * \hat{\ell}(x_c/p) = \hat{P}(O) = \prod_{i=1}^{n} P(V_i/\pi(V_i)) \quad (5)$$

Finally, a censoring correction threshold P_{Th} is chosen to relocate the censored instance to either high- or

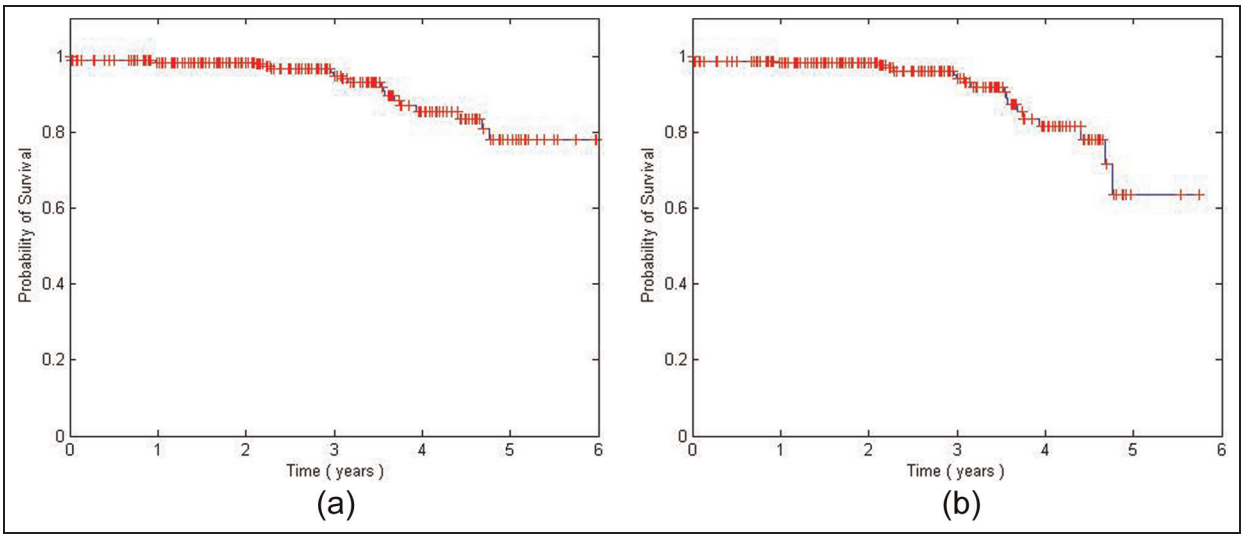

Figure 1. Kaplan–Meier curves for EVAR: (a) center 1 and (b) center 2.

low-risk groups. If $P(O^{high}/x_c)$ is greater than P_{Th}, the censored instance will be labeled as high risk and vice versa.

The two Bayesian networks were learned with hill-climbing structure learning algorithm to produce the low- and high-risk networks. The scoring function used was minimum description.[34] Parameter learning was done using maximum likelihood procedure.[35] The thresholds which showed better classification accuracy results for uncensoring the two centers were in the range between 0.7 and 0.8.

The structure and parameter learning of the Bayesian networks were implemented using Weka. The following neural network model, Kaplan–Meier curve, and logrank test were implemented using MATLAB.

Classification models and evaluation metrics

A three-layer BP neural network model was constructed using the uncensored center 1 data in order to determine the REINT after 5 years from EVAR surgery and classify patients of the center 2 into high-risk and low-risk groups. Log-sigmoid and saturating linear transfer functions were used as activation functions for the two upper layers. Additionally, two neural networks were constructed using the EVAR dataset of centers 1 and 2 separately to test the performance of the proposed uncensoring algorithm; this testing was done for each network using the 10-fold cross-validation test. The number of hidden neurons in each network is determined through iterative approach to produce the minimum mean square error in the test.

Metrics used to evaluate the performance of neural networks built are discussed below:

- Classification accuracy is the percentage of the correctly classified (predicted) instances among all the instances used in the testing set.
- Sensitivity (true positive rate) is the ratio between the correctly classified positive instances and the actual positive instances of the original dataset before classification.
- Specificity (true negative rate) is the ratio between the correctly classified negative instances and the actual negative instances of the original dataset before classification.
- False-positive rate is the ratio between the incorrectly classified positive instances and the actual negative instances of the original dataset before classification.
- False-negative rate is the ratio between the incorrectly classified negative instances and the actual positive instances of the original dataset before classification.
- Receiver operating characteristics curve (ROC) is the curve that plots the sensitivity as a function of specificity. Usually, area under the receiver operating characteristics curve (AUROC) is used to evaluate the performance of the model. The greater area indicates better performance. The maximum area that can be reached is one which indicates that 100% of the data were correctly classified.

Kaplan–Meier curves

Also known as product limit estimate of the survival function,[36,37] Kaplan–Meier curve is a well-known nonparametric survival analysis technique which gives an estimation of the probability of patient's survival at any time for the whole dataset even if it is censored. It is widely used in clinical trials such as the determination of the effectiveness of a specific treatment on illness relapses by calculating the number of patients at risk. Another example is the estimation of the risk of REINT after a surgical intervention.[30]

Logrank test

In clinical trials, usually doctors need to check whether there is a significant difference between two risk groups of patients. They start by drawing survival curves (usually Kaplan curves) to determine the probability of survival using the two risk groups separately. Then, statistical tests are employed in order to compare the survival curves of the two risk groups. The most well-known test is logrank. It is a hypothesis test to compare the survival distributions of the two risk groups. For every time t at which an event of interest has occurred, it calculates the number of the observed (O) events, expected (E) events, and the variance in the expected events (V) in each group and then puts them in a table. Finally, it performs a chi-squared test, which is the summation of $(O-E)/\sqrt{V}$. The result is a statistical coefficient value known as p value which indicates the difference between groups.[38] Usually, a p value lower than 0.05 indicates that there is a significant difference between the two risk groups.

Results

Separate BP neural network models for prediction of center 1 and center 2 datasets, respectively

The number of hidden neurons, learning rate, and momentum of the BP neural network for each center is 13, 0.3, and 0.2, respectively. A comparison between the prediction results of the datasets before and after using the S&DB and A&M uncensoring approaches for centers 1 and 2 are illustrated in Tables 1 and 2, respectively. A 10-fold cross-validation test was used to produce these results in both cases. The number of high-risk patients in center 1 after uncensoring is 69 (S&DB approach) and 204 (A&M approach), while in center 2 after uncensoring, it is 52 (S&DB approach) and 73 (A&M approach). It is clear in Tables 1 and 2 that both techniques were capable of uncensoring the datasets, as the AUROC for the prediction of results after uncensoring has increased than that before uncensoring.

Table 1. Comparing prediction metrics of center 1 before uncensoring and after uncensoring using the S&DB and A&M techniques.

Class	True (%)	False (%)	AUROC
Before uncensoring (center 1)			
High risk (positive)	9.5	90.5	0.572
Low risk (negative)	94.3	5.7	
After uncensoring with S&DB approach (center 1)			
High risk (positive)	44.9	55.1	0.713
Low risk (negative)	93.2	6.8	
After uncensoring with A&M approach (center 1)			
High risk (positive)	66.7	33.3	0.808
Low risk (negative)	75.8	24.8	

S&DB: Štajduhar and Dalbelo-Bašić; AUROC: area under the receiver operating characteristics curve.

Table 2. Comparing prediction metrics of center 2 before uncensoring and after uncensoring using the S&DB and A&M techniques.

Class	True (%)	False (%)	AUROC
Before uncensoring (center 2)			
High risk (positive)	4.5	95.5	0.472
Low risk (negative)	93.2	6.8	
After uncensoring with S&DB approach (center 2)			
High risk (positive)	36.5	63.5	0.664
Low risk (negative)	82.6	17.4	
After uncensoring with A&M approach (center 2)			
High risk (positive)	41.1	58.9	0.702
Low risk (negative)	83.3	16.7	

S&DB: Štajduhar and Dalbelo-Bašić; AUROC: area under the receiver operating characteristics curve.

Also, the prediction model before using the uncensoring techniques was biased toward the low risk (zero targets) with true negative and true positive values of 0.943 and 0.095, respectively. This means that only 9.5% patients who underwent REINT have been identified correctly by the censored model; this shows that censoring is a major cause when standard machine-learning techniques fail to work properly in constructing survival models. Moreover, the A&M technique outperforms the S&DB as the AUROC in the former is 0.808 and 0.702, while that in the later is 0.713 and 0.664 for centers 1 and 2, respectively.

Center 1 neural network model for prediction of center 2 patients

S&DB technique used for uncensoring the datasets. A BP neural network model was built using the dataset of uncensored center 1 to predict the risk of REINT on center 2 patients. The number of hidden neurons used was 3 in this model. A mean square error of 0.1 was employed as a stopping criterion. Table 3 illustrates the results. The AUROC of the trained model in center 1 was 0.8377, while when tested in center 2 dataset, it was 0.5481. This indicated that the model was incapable of properly predicting the risk of REINT on center 2 patients. The Kaplan–Meier curves were plotted for the predictions as shown in Figure 2, and the logrank test was used to determine the significance between the risk groups and also indicate whether the model has correctly classified patients. The p value obtained from the logrank test was equal to 0.33454, meaning that the neural network failed to differentiate between the low- and high-risk groups.

A&M technique used for uncensoring the dataset. A BP neural network model was built using the dataset of center 1 which was processed by the new A&M uncensoring technique to predict the risk of REINT on patients of center 2. The activation function of both input and hidden layers was log-sigmoid and saturating linear transfer function. The number of hidden neurons used was 6. A mean square error of 0.0705 was employed as a stopping criterion. Table 4 illustrates the results. The AUROC of the trained model in center 1 was 0.9498 which is better than that of S&DB (0.8377), when tested in center 2, the AUROC was a better value of 0.666 comparing to the value of 0.5481 achieved using the S&DB technique. The Kaplan–Meier curves were plotted for the uncensored center 1 as shown in Figure 3, and the predictions of center 2 are shown in Figure 4. The logrank tests were used to determine the significance between the two risk groups of the neural network predictions and the p value obtained was equal to 0.00037, meaning that the model has succeeded to differentiate between the low- and high-risk groups.

Table 3. Training and prediction metrics results of the neural network model trained with uncensored center 1 data using the S&DB technique and tested to the data in center 2.

Class	True (%)	False (%)	AUROC
Training with uncensored dataset (center 1)			
High risk (positive)	37.4	62.6	0.8377
Low risk (negative)	97.6	2.4	
Testing with censored dataset (center 2)			
High risk (positive)	23.1	76.9	0.5481
Low risk (negative)	86.5	13.5	

S&DB: Štajduhar and Dalbelo-Bašić; AUROC: area under the receiver operating characteristics curve.

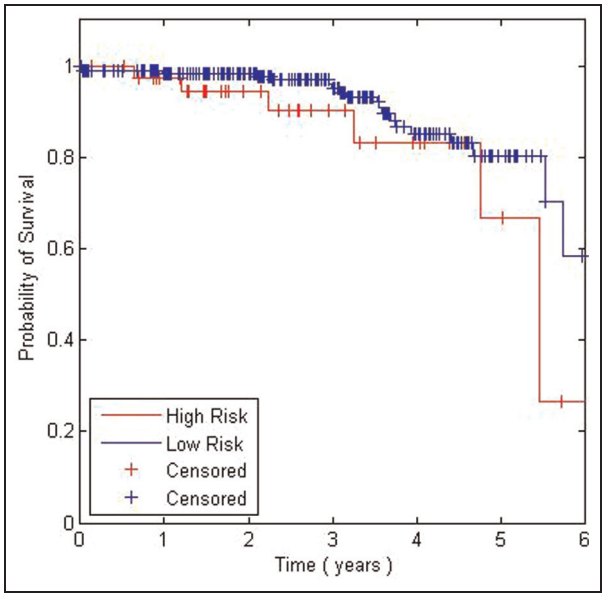

Figure 2. Kaplan–Meier curves' prediction of center 2 EVAR using the neural network model built from uncensored center 1 dataset produced with S&DB technique.

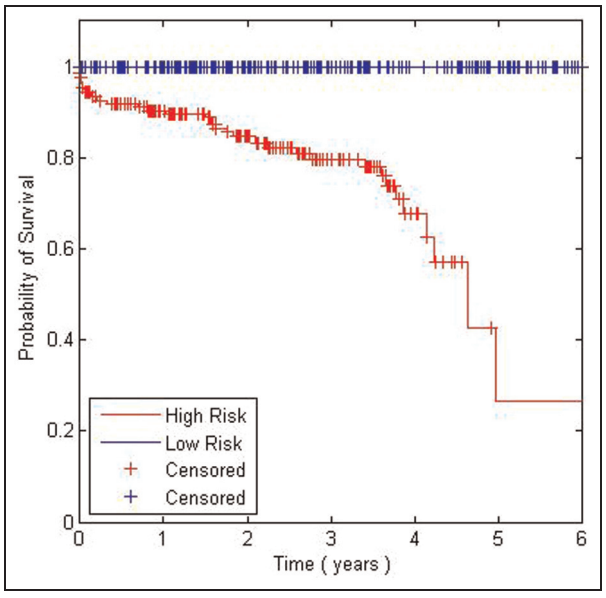

Figure 3. Kaplan–Meier curve for the uncensored center 1 dataset using the new A&M technique.

Discussion

Bayesian network and BP neural network are capable of building a predictive model of REINT for EVAR patients. The Bayesian networks successfully uncensored the clinical datasets. As shown in the Tables 1 and 2, the two BP neural network models that tested the efficiency of the uncensoring algorithms have increased the AUROC from 0.572 to 0.808 and 0.472 to 0.702 for center 1 and center 2, respectively. Additionally, the proposed A&M approach outperforms the S&DB approach in all the model construction processes. The results in Tables 3 and 4 show that the model constructed using the S&DB approach is unable to distinguish between the two risk groups of the censored center 2 data (p value = 0.33454); however, the A&M technique succeeds in differentiating between them (p value = 0.00037).

From the above results, we can see that the neural network model constructed using the proposed A&M technique with center 1 is able to predict the risk of REINT after EVAR of the patients in center 2. The differences in freedom from aortic complications between low- and high-risk patients after 5 years are 100% versus 20% in center 1 and 95% versus 64% in center 2 as shown in Figures 3 and 4 (from the estimated survival functions or probabilities), which means that low-risk patients have higher probabilities to be free from aortic complication than the high-risk patients. The p value of the logrank test was 0 and 0.00037 differentiating the two risk groups of centers 1 and 2, respectively, using the new A&M algorithm. It classified 56% and 62% of the patients as low risk in center 1 and center 2, respectively. Those patients can therefore be monitored at less regular intervals to reduce their exposure to radiation and costs involved with frequent monitoring.

Advantages of this Bayesian neural network approach are as follows: first, usually medical researchers are not aware of data mining and machine-learning techniques. Since Bayesian network is a graphical probabilistic network, it shows the joint probabilities between variables in the form of a graph consisting of nodes representing these variables and arcs showing relations and causality between them. Hence, it is considered an ideal tool that can be used in medicine to show causal influence between variables and their

Table 4. Training and prediction metric results of the neural network model trained with uncensored center 1 data using the new A&M technique and tested to the data in center 2.

Class	True (%)	False (%)	AUROC
Training with uncensored dataset (center 1)			
High risk (positive)	86.3	13.7	0.9498
Low risk (negative)	97.6	2.4	
Testing with censored dataset (center 2)			
High risk (positive)	80.8	19.2	0.666
Low risk (negative)	66.0	34.0	

AUROC: area under the receiver operating characteristics curve.

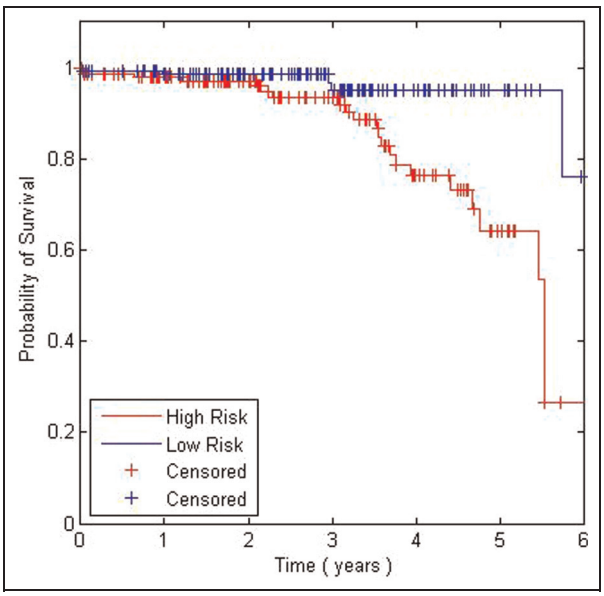

Figure 4. Kaplan–Meier curves' prediction of center 2 using the neural network model built from the uncensored center 1 dataset with the new A&M technique.

probabilistic relations.[39] Moreover, physicians could calculate the condition and marginal probabilities of the network which demonstrate the uncertainty of the studied medical domain.[40] Second, this approach was capable of uncensoring highly censored dataset. This uncensored data can then be used later in any standard supervised machine-learning technique as the time variable was embedded in the uncensoring technique. Third, the neural network model is a powerful classifier as it successfully predicted aortic complications and distinguished between the high-risk and low-risk groups of patients despite the highly censored nature of the datasets.

Conclusion and future work

Machine-learning techniques have been widely used for survival analysis in clinical trials. Censoring is a common problem when dealing with survival data. Censored patients cannot be ignored or discarded, as this may bias the prediction model and affect the prediction results, especially when the data are highly censored. In this article, a new modified Bayesian neural network approach was proposed to deal with the high censoring issue appeared in the EVAR datasets of the two different vascular centers, and its performance was compared to a more traditional approach proposed by S&DB.

In the new A&M approach, the dataset was divided into three groups according to their inherent level of risks: the true high risk (patients who experience the REINT at a time lower than or equal 5 years), the true low risk (patients who did not need the REINT till a time greater than or equal 5 years), and the rest considered as the truly censored group. Two Bayesian networks were established as the high-risk and low-risk networks, and they were used to relocate the truly censored patients to either the low- or high-risk groups. The S&DB approach only considered the high-risk and censored patients, and the censored patients can include the true low-risk patients whose events of interest never happen even after the 5-year survival time. Therefore, simply separating the dataset into high-risk and censored groups will lead to misrepresentation of the data output, further leading to lower prediction accuracies and unsuccessful discriminations in the survival models.

After applying the proposed uncensoring algorithm on the two EVAR datasets, this article presented a successful BP neural network model that was constructed using center 1 dataset and employed to classify patients of center 2 into high- or low-risk categories. The proposed A&M approach has successfully increased the AUROC for both centers. Moreover, it was able to predict the risk of REINT on censored center 2 data. It also succeeded in differentiating between the low- and high-risk groups with a p value of the logrank test equals to 0.00037 which is not the case for the S&DB approach (p value of the logrank test equals 0.33454). The high risk means that the patients will more likely need REINT and therefore frequent monitoring of conditions, whereas less frequent monitoring is needed for low-risk patients. A future observation plan could then be implemented to the new EVAR patients with the proposed Bayesian neural network model.

As for future work, feature selection techniques will be investigated to choose the optimal variables that affect the prediction accuracies the most. Minimizing

the number of input features needed for building the model is beneficial, as the process of measuring and collecting the morphological features from the CT images is exhausting and time-consuming. An optimized and minimized input feature set will improve the efficiency of the data collection process and therefore reduces costs to healthcare authorities.

Acknowledgements

We would like to acknowledge St George's Vascular Institute, London, UK, for supplying us with the two EVAR datasets used in this article and Alan Karthikesalingam at St George's Vascular Institute and Prof. David Lowe at Aston University for their guidance and useful discussions on this research.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Kul S. The use of survival analysis for clinical pathways. *Int J Care Pathw* 2010; 14: 23–26.
2. Leung KM, Elashoff RM and Afifi AA. Censoring issues in survival analysis. *Annu Rev Public Health* 1997; 18: 83–104.
3. Brown SF, Branford AJ and Moran W. On the use of artificial neural networks for the analysis of survival data. *IEEE T Neural Networ* 1997; 8(5): 1071–1076.
4. Singh R and Mukhopadhyay K. Survival analysis in clinical trials: basics and must-know areas. *Perspect Clin Res* 2011; 2: 145–148.
5. Prinja S, Gupta N and Varma R. Censoring in clinical trials: review of survival analysis techniques. *Indian J Community Med* 2010; 35: 217–221.
6. Kaderali L, Zander T, Faigle U, et al. CASPAR: a hierarchical Bayesian approach to predict survival times in cancer from gene expression data. *Bioinformatics* 2006; 22: 1495–1502.
7. Kalderstam J, Edén P, Bendahl PO, et al. Training artificial neural networks directly on the concordance index for censored data using genetic algorithms. *Artif Intell Med* 2013; 58(2): 125–132.
8. Chi CL, Street WH and Wolberg WH. Application of artificial neural network-based survival analysis on two breast cancer datasets. In: *Proceedings of the AMIA annual symposium*, Chicago, IL, 10–14 November 2007, pp.130–134. Chicago: American Medical Informatics Association.
9. Yu CN, Greiner R, Lin HC, et al. Learning patient-specific cancer survival distributions as a sequence of dependent regressors. In: *Advances in Neural Information Processing Systems,* Granada, Spain, 12–15 December 2011, pp.1845–1853. Spain: Neural Information Processing Systems (NIPS).
10. Choi I, Wells BJ, Yu C, et al. An empirical approach to model selection through validation for censored survival data. *J Biomedical Inform* 2011; 44(4): 595–606.
11. Cheng TH, Lan CW, Wei CP, et al. Cost-Sensitive learning for recurrence prediction of breast cancer. In: *Proceedings of the 14th Pacific Asia conference on information systems (PACIS)*, Taipei, Taiwan, 9–12 July 2010. USA: Association for Information Systems.
12. Altman DG and Bland JM. Statistics notes: time to event (survival) data. *BMJ* 1998; 317(7156): 468–469.
13. Collett D. *Modelling survival data in medical research*. Florida, USA: CRC Press, 2003.
14. Zupan B, Demšar J, Kattan MW, et al. Machine learning for survival analysis: a case study on recurrence of prostate cancer. *Artif Intell Med* 2000; 20(1): 59–75.
15. Delen D, Walker G and Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. *Artif Intell Med* 2005; 34(2): 113–127.
16. Moll FL, Powell JT, Fraedrich G, et al. Management of abdominal aortic aneurysms clinical practice guidelines of the European society for vascular surgery. *Eur J Vasc Endovasc Surg* 2011; 41: S1–S58.
17. Hay N, McCracken F, Richardson J, et al. *Endovascular stent-grafts for the treatment of abdominal aortic aneurysms: NICE technology appraisal guidance*. No. 167, February 2009. London: National Institute for Health and Clinical Excellence, pp.1798–1800.
18. Karthikesalingam A, Holt PJ, Vidal-Diez A, et al. Predicting aortic complications after endovascular aneurysm repair. *Br J Surg* 2013; 100(10): 1302–1311.
19. Štajduhar I and Dalbelo-Bašić B. Learning Bayesian networks from survival data using weighting censored instances. *J Biomed Inform* 2010; 43: 613–622.
20. Jerez JM, Molina I, García-Laencina PJ, et al. Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artif Intell Med* 2010; 50: 105–115.
21. Kjearulff UB and Madsen AL. *Probabilistic networks-an introduction to Bayesian networks and influence diagrams*. Denmark: Aalborg University, 2005.
22. Sebastiani P, Nolan VG, Baldwin CT, et al. A network model to predict the risk of death in sickle cell disease. *Blood* 2007; 110: 2727–2735.
23. Steele S, Bilchik A, Eberhardt J, et al. Using machine-learned Bayesian belief networks to predict perioperative risk of clostridium difficile infection following colon surgery. *Interact J Med Res* 2012; 1(2): e(6).
24. Damato B, Eleuteri A, Fisher AC, et al. Artificial neural networks estimating survival probability after treatment of choroidal melanoma. *Ophthalmology* 2008; 115: 1598–1607.
25. Taktak AF, Fisher AC and Damato BE. Modelling survival after treatment of intraocular melanoma using artificial neural networks and Bayes theorem. *Phys Med Biol* 2004; 49: 87–98.
26. Cheng-Mei C, Chien-Yeh H, Hung-Wen C, et al. Prediction of survival in patients with liver cancer using artificial neural networks and classification and regression trees. In: *Proceedings of the seventh international conference on natural computation (ICNC)*, Shanghai, China, 26–28 July 2011, vol. 2, pp.811–815. New York: IEEE.
27. Gohari MR, Biglarian A, Bakhshi E, et al. Artificial neural network to determine the prognostic factors in colorectal cancer patients. *Asian Pac J Cancer Prev* 2011; 12: 1469–1472.
28. Bellaachia A and Guven E. Predicting breast cancer survivability using data mining techniques. In: *Proceedings*

of the ninth workshop on mining scientific and engineering datasets in conjunction with the sixth SIAM international conference on data mining, Bethesda, Maryland, USA, 22 April 2006. USA: Society for Industrial and Applied Mathematics (SIAM).
29. Lisboa PJG, Wong H, Harris P, et al. A Bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer. *Artif Intell Med* 2003; 28: 1–25.
30. Karthikesalingam A, Holt PJ, Hinchliffe RJ, et al. Risk of reintervention after endovascular aortic aneurysm repair. *Br J Surg* 2010; 97(5): 657–663.
31. Patterson BO, Holt PJ, Hinchliffe R, et al. Existing risk prediction methods for elective abdominal aortic aneurysm repair do not predict short-term outcome following endovascular repair. *J Vasc Surg* 2010; 52(1): 25–30.
32. Patterson BO, Hinchliffe RJ, Holt PJ, et al. Importance of aortic morphology in planning aortic interventions. *J Endovasc Ther* 2010; 17(1): 73–77.
33. Štajduhar I and Dalbelo-Bašić B. Uncensoring censored data for machine learning: a likelihood-based approach. *Expert Syst Appl* 2012; 39(8): 7226–7234.
34. Lam W and Bacchus F. Learning Bayesian belief networks: an approach based on the MDL principle. *Comput Intell* 1994; 10: 269–293.
35. Neapolitan RE. *Learning Bayesian networks*. Upper Saddle River, NJ: Prentice Hall, 2003.
36. Kaplan EL and Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958; 53: 457–481.
37. Goel MK, Khanna P and Kishore J. Understanding survival analysis: Kaplan-Meier estimate. *Int J Ayurveda Res* 2010; 1(4): 274–278.
38. Stevenson M. Introduction to survival analysis, http://www.biecek.pl/statystykaMedyczna/Stevenson_survival_analysis_195.721.pdf (2009, accessed 22 July 2013).
39. Štajduhar I, Dalbelo-Bašić B and Bogunovic M. Impact of censoring on learning Bayesian networks in survival modelling. *Artif Intell Med* 2009; 47: 199–217.
40. Blanco R, Inza I, Merino M, et al. Feature selection in Bayesian classifiers for the prognosis. *J Biomed Inform* 2005; 38: 376–388.