Using Multiple Classifiers for Predicting the Risk of Endovascular Aortic Aneurysm Repair Re-intervention through Hybrid Feature Selection

Omneya Attallah [1,2], Alan Karthikesalingam[3], Peter J E Holt[3], Matthew M Thompson[3], Rob Sayers[4], Matthew J Bown[4], Eddie C Choke[4], Xianghong Ma[2, *]

[1] Department of Electronics and Communications, College of Engineering and Technology, Arab Academy for Science and Technology, Alexandria, Egypt

[2] School of Engineering and Applied Science, Aston University, Birmingham, B4 7ET, UK

[3]St George's Vascular Institute, St George's University Hospitals NHS Foundation Trust, Blackshaw Road, London SW17 0QT

[4] NIHR Leicester Cardiovascular Biomedical Research Unit and the Department of Cardiovascular Sciences, University of Leicester

* Corresponding author and address:

Xianghong Ma

School of Engineering and Applied Science,

Aston University, Birmingham, B4 7ET, UK

Phone: +441212043592

Email: x.ma@aston.ac.uk

Abstract

Feature selection (FS) is essential in medical area; however its process becomes complicated with the presence of censoring which is the unique character of survival analysis. Most survival FS methods are based on Cox's proportional hazard model, though machine learning classifiers (MLC) are preferred. They are less employed in survival analysis due to censoring which prevent them from directly being used to survival data. Among the few work that employed MLC, Partial logistic artificial neural network with auto-relevance determination (PLANN-ARD) is a well-known method that deals with censoring and perform FS for survival data. However it depends on data replication to handle censoring which leads to unbalanced and biased prediction results especially in highly censored data. other methods cannot deal with high censoring as well. Therefore, in this paper a new hybrid FS method is proposed which presents a solution to high level censoring. It combines support vector machine, neural network, and K nearest neighbor classifiers using simple majority voting and a new weighted majority voting method based on survival metric to construct a multiple classifier system (MCS). The new hybrid FS process uses MCS as a wrapper method and merges it with iterated feature ranking filter method to further reduce features. Two endovascular aortic repair (EVAR) datasets containing 91 % censored patients collected from two centers were used to construct a multicenter study to evaluate the performance of the proposed approach. The results showed the proposed technique outperformed individual classifiers and variable selection methods based on Cox's model such as Akaike and Bayesian information criterions and Least absolute shrinkage and selector operator in p-values of the log-rank test, sensitivity, and concordance index. This indicates that the proposed classifier is more powerful in correctly predicting the risk of re-intervention enabling doctor in selecting patients' future follow up plan.

Keywords Multiple Classifier System, Hybrid feature selection, Survival analysis; Censoring; Cox's proportional hazard model; Endovascular Aortic Repair

1. Introduction

Feature selection (FS), model Selection (MS), and variable reduction and transformation are important topics in data mining; especially when dealing with real medical datasets of large size. FS methods search for a reduced number of variables that have the ability to improve prediction using a selection criterion. However, feature reduction and transformation convert data into a new domain capable of compressing the necessary information needed for classification in a reduced number of new variables. MS chooses one optimal (or more) model from a number of candidate models formed from either several classifiers or the same one but with different parameters. It can be considered as FS when the purpose is to choose between several subsets of variables generated during MS. Variable reduction and transformation techniques tend to lower the classifier's complexity and speed up the classification task. In addition, they enhance generalization and prevent over-fitting [1]. Clinicians need them to build a reduced predictive model in order to decrease the effort and time needed to measure the unnecessary variables.

FS methods are divided into filter, wrapper, and embedded methods. However, recently, many researches focused on merging two or more techniques to form a new class of FS technique known as hybrid FS. The main reason for doing this is that the hybrid method has the joined advantages of these FS approaches. It also enables the construction of better reduced predictive model.

The literature review revealed that many FS related papers were for standard data. However, this process becomes more complicated for survival data due to the presence of censoring. Censoring is the main characteristic differentiating survival data from standard supervised data. Censoring means that for some patients the event of interest (such as death, recurrence of a disease) did not occur during the study period. The censored patient cannot be ignored in building a predictive model, as this might result in biased predictions especially when there is a large amount of censored patients in the data [2]. Among the work done for censored survival data, most of them were focused on using forward, backward, step wise, penalized and shrinkage variable selection with Cox proportional hazard model, though machine learning classifiers (MLC) are more favored as they consider complex relations and non-linearity existed in the data during the modeling process, which is not the case in statistical methods [3]. However, they are less used in survival analysis due to the fact that censoring makes them less capable to be directly used for survival data [4, 5]. Therefore, the censoring problem should be handled first. MLC that dealt with censoring to improve survival models include Artificial Neural Network (ANN) [6, 7] , naïve Bayes and decision tree [4], and Bayesian networks [8]. However, they were not employed to do FS in survival analysis.

Some work was done for FS in survival analysis using MLC; among them is the well-known partial logistic artificial neural network with auto relevance determination [9] (PLANN-ARD). This method performs FS with Bayesian framework; however, it handles the censoring issue by dividing observation time into time intervals and repeating patients to these intervals. The main drawback of this method is that this repetition will lead to unbalanced model and biased prediction results especially with highly censored data. Moreover, increase in data examples will increase the complexity and the training time of the predictive model, which is not preferred. Therefore, in this paper a hybrid FS is proposed that presents a solution to censoring without data repetition. It can be used with any standard MLC rather than only with neural network as the PLANN-ARD. Others used Cox's model to perform FS, then used MLC to construct predictive models such as SVM [10]. Others wrapped FS around Bayes classifiers [5, 11] or KNN [12]. In [13], the authors use chi-square test to determine the association between variables and survival times of lung cancer and select the most related variables to construct an ANN model. The main drawback of this method is the way to deal with censoring which is using only uncensored patients and ignoring censored cases, or considering censored patients as event free, which is not applicable for high censored datasets like EVAR datasets used in this work.

Recently, the concept of multiple classifiers system (MCS) raises interests among many researchers in the machine learning field. Wolpert has mentioned in [14] that there is no single classifier ideal for all classification tasks; as each one has its area of competency [15, 16]. Therefore, MCS is advantageous. It merges the outputs of multiple classifiers using a fuser in order to improve predictions. Though, care must be taken to prevent generating of unstable models in which predictions are sensitive to any changes in the training data used to build it [16]. Several fusion methods are available in the literature such as bagging, boosting, voting and stacking. Authors in [17-21] applied them to classify Alzheimer disease and fMRI images. They were used in [22-24] in order to predict cardiovascular diseases and protein fold. Moreover, FS techniques were combined with them to predict brain glioma, hepatitis, diabetes, liver disorder, breast cancer, tumors, cardiovascular diseases and protein fold [25-30].

Generally all the above fusion methods produce similar results [31, 32]. Many researchers prefer majority voting fusion algorithm due to its simplicity [33, 34]. Majority voting can be classified into simple and weighted methods. Simple Majority voting approach usually improves predictions results, however it treats all classifiers equally; it does not put attention to classifiers that have higher impact on classification and generalization. Weighted majority voting approach overcomes this drawback by allowing each classifier in the pool to have a weight equivalent to its performance. Higher weights are given to those that have greater contribution to prediction results. The total weights should be equal to one in order to construct a proper weight distribution. In this paper, first simple majority voting was used to construct an MCS. Afterwards, a weighted majority voting method was developed based on survival analysis metric to build the MCS in order to improve the predictions of the simple voting method. This system can be used for censored survival data type.

Endovascular aortic aneurysm repair (EVAR) operation has recently become the preferred surgical route by doctors and patients for handling abdominal aortic aneurysm [35]. Long lasting surveillance is important after EVAR [36]. It is expensive and has low standardizations [37] and its optimization is needed. Several approaches are available with limitations in the techniques used to select the optimal timing or modality [38]. More frequent observations would expose patients to a huge amount of radiations and contrast nephropathy which is unsafe [39]. Moreover, some complications that need to be examined for treatment could be missed between follow up sessions [40]. A re-intervention might be required for some patients after EVAR. Distinguishing between those who have higher probability to surgical re-intervention (high-risk patients) and those who most likely will not need it (low-risk patients) is essential. It will enable doctors to put patients into appropriate future follow up observation plans. High risk patients would be monitored more frequently than low risk ones, leading to the long-lasting effectiveness of the surveillance system.

The aim of this paper is to offer a solution to censoring of high level available in the two EVAR datasets available in this study without deleting, ignoring, or considering censored patients as event free which are common methods to handle censoring. The solution also does not depend on data repetition which increases training data and consequently training time and complexity of the predictive model. It also prevents the construction of unbalance and biased predictive models. The proposed method can be used with any MLC. This solution is used in the hybrid feature selection technique which combines filter and wrapper approaches along with feature reduction and transformation to remove unnecessary variables in the highly censored EVAR datasets in order to produce a final stable predictive model that avoids bias. Moreover, this paper adopts MLC techniques to deal with censorship instead of the traditional statistical models such as Cox's proportional hazard model which is commonly used in the medical area to model survival data and deal with censorship [41]. In addition, this paper uses MCS instead of an individual classifier for cross-center prediction, where a stable predictive model was built with the EVAR data from one medical center to predict the risk of re-intervention on patients in another center. They are equivalent to taking several clinics diagnosis opinions which may result in a more accurate final decision. Two MCSs are constructed, the first used simple majority voting for prediction, and the other used a new weighted majority voting based on a survival metric to be used with censored survival data type. The proposed weighted majority voting method gives different weights to each classifier according to its performance which consequently enhances the prediction results shown later in the results section.

2. Materials and Methodology

2.1 Datasets Description

Patients that had the EVAR surgery in two separate vascular centers located in the UK were monitored from 2004 till 2010. The first center is located in St George hospital in London and the other in Leicester. The morphological variables were collected from computed tomography (CT) images of the thoracic inlet to the level of the common femoral artery bifurcation. Images have slice thickness of 0.625 or 1.25 mm. Morphological features were collected for patients and used in this work as they have greater effect on aortic complications than physiology features. This judgment was reliable with earlier proof that the main factor of endograft failure is patient anatomy rather than co-morbidity [40, 42, 43]. Both datasets contain 45 attributes with 457 and 286 patients, respectively, after removing the ones with missing values. Patient numbers that actually re-experienced the EVAR surgery are 40 and 26 for Center 1 and 2 correspondingly. Details of the datasets can be found in a previous publication [44]. Kaplan Meier (KM) curves were plotted for both centers as shown in Figure 1. More details about KM method can be found in [45].

Figure 1

2.2 Factor Analysis

FA examines the underlying structure of the data. It considers that data attributes are generated from linear combination of unseen (unmeasured) variables called factors. They consist of two parts; unique and common. Unique factor refers to unique variance of one seen (measured) variable, while common factors express common variances between observed ones. Generally, features that are not correlated to any factor could be deleted. These selected observed variables could be used to build a predictive model [46].

2.3 Multiple Classifiers System

An MCS gathers powers of each learning algorithm in order to outperform the performance of each single classifier. In the medical field, it is equivalent to taking the opinion of several doctors to reach a more confident final decision. Sometimes, ensemble classifiers' results are not as good as the performance of the best individual classifier in the pool. However, it prevents the chance of poor decisions that might be taken with a particular inappropriately chosen model [33] .

An MCS has two topologies; serial and parallel. In the serial topology, classifiers are connected in series following some sorting over them. If the first classifier predictions are not accurate enough, the next stronger classifier will be used. Classifiers are added iteratively according to their order until predictions are finally enhanced [47]. On the other hand, in parallel connection, the same variables are used to construct all classifiers in the pool, and the final prediction is determined based on outputs of each single classifier independently. Parallel topology is the most common way used to connect classifiers [48], so it is adopted in this paper.

2.4 The Proposed Algorithm

The algorithm consists of 7 steps. Fig. 2 shows the steps of the algorithm and the three main areas of contribution in the proposed approach highlighted in blue colour (feature selection, uncesoring, and classification) along with their interactions.

- **STEP 1** is FA which is made after both Kaiser-Meyer-Olkin and Bartlett's tests to determine if FA is need for Center 1 or not. The number of factors used for FA was initially determined by performing a scree plot which shows the eigenvalues accompanied with latent factors listed in descending order versus the number of factors. Features not related to any latent factors are deleted using communality value which is part of the variance generated from common variables.

- **STEP 2** is cross validation and permutation. It splits the Center 1 data into five folds, each separate four of which is called outer training folds. They were used for FS process. These folds were shuffled five times.
- **STEP 3** is the first stage feature selection (FSFS) step which is done in two phases, stepwise feature model selection and feature ranking (FR). In the former, each outer training fold uses stepwise searching strategy that swifts between backward and forward searches to reduce the number of features. It eliminates one variable at a time iteratively. Each eliminated variable is inserted in a subset called "visited". It will be given another chance to re-enter the search space. After adding or deleting a variable from every outer training fold, it is shuffled and re-split five times to get the average of the p-value of predictions, which is the criterion for FS. The model with the smallest average p-value is the one chosen. This is repeated until all the variables are visited. Five outer reduced models will be generated at the end of this stage. Usually, in model selection only one model is chosen to win. However, this does not take consideration of the uncertainty in all or some of the candidate models. Therefore, in this paper all variables appeared in the five models were used in the FR phase and ranked according to their frequency distribution.
- **STEP 4** is the uncensoring step in which observation time variable was used to split patients of each training fold into three groups; high risk, low risk, and censored groups. In step 3, low and high risk groups were used to construct two Bayesian networks called low B^{low} and high B^{high} networks after removing the observation time variable. They were used to uncensor every patient of the censored group by comparing him or her to the internal configuration of each network p^{high} and p^{low}

using likelihood information. More details about the uncensoring technique could be found in the researchers' previous work [49].

Each variable V_i represents a node in this network that may be connected to a higher parent node (π) and lower child node. They are directed acyclic graph (DAG) networks given a symbol ξ meaning that nodes are connected in only one direction from parent to children nodes. The Bayesian networks were learned with Hill climbing structure learning algorithm [50]. The scoring function used for choosing the structure of the network was minimum description [51]. Parameter learning was done using maximum likelihood procedure to determine relation between nodes of a network [52].

The likelihood $\ell(x_c / p)$ that each censored patient belongs to which network is calculated using equations (1) and (2) to decide to which group censored patients belong.

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

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

where; $\pi(V_i)$ is the parent node to variable $V_{i,}$, $P^{high}(V_i / \pi(V_i)$, and $P^{low}(V_i / \pi(V_i)$ are the posterior probability of a variable $V_{i,}$ given its parents nodes for high and low Bayesians networks, respectively.

Afterwards, the posterior probability that outcome predictions that patients belong to which network given that they are censored (x_c) $P(O/x_c)$ in equation (5) is calculated using equations (3) and (4).

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

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

$$P(O/x_c) = P(O^{high}/x_c) + P(O^{high}/x_c) = \frac{\hat{P}(O^{high}) * \hat{\ell}(x_c / p^{high}) + \hat{P}(O^{low}) * \hat{\ell}(x_c / p^{low})}{P(x_c)} \tag{5}$$

Equation (5) is then normalized to ignore the effect of probability of a censored instance $P(x_c)$ by dividing equation (5) by $P(O/x_c) * P(x_c)$ to get equation (6).

$$P(O^{high}/x_c) + P(O^{low}/x_c) = 1. \tag{6}$$

Lastly, a threshold is used to decide which risk group each censored patient belongs to. It is called censoring correction threshold P_{Th} . If $P(O^{high}/x_c)$ is greater than P_{Th}, then the patient is considered a high risk to do a re-intervention and vice versa.

- **STEP 5** is iterated nested cross validation. Each shuffled version of step 2 after being uncensored is re-split again into five inner nested folds. Every four inner folds are used for constructing the MCS which is the sixth step while the remaining one is used to test it.

- **STEP 6** is the MCS construction step. The proposed MSC system was constructed using three popular machine learning classifiers; support vector machine (SVM), multiple layer perceptron (MLP) neural network, and K-nearest neighbor (KNN). Both SVM and MLP Neural networks are well known as strong classifiers. Moreover, they can detect the complex and high nonlinearity relations existing in the datasets [33, 53]. They have been widely used in medical applications [25, 28, 54]. KNN is a simple, straightforward and highly efficient classifier even with noisy data [55]. Despite its simplicity, it has shown good performance in medical application [56, 57]. In this paper, classifiers were built using Weka software [58]. Sigmoid function was employed for SVM construction. A three layer MLP ANN was constructed with seven hidden and two output neurons, and sigmoid activation functions with learning rate 0.3 and momentum 0.2. KNN was built using Euclidean distance function and K was set to 3.

Predictions were first combined with simple majority voting which simply gives a final decision to the class which has the majority of the votes. The average of the p-value of the log rank test of the predictions was calculated and chosen as a criterion for feature selection. This procedure is called iterated nested cross validation which produces a stable model and overcome over-fitting that might occur later.

Afterwards, a weighted majority voting based on the p-value of the log –rank test survival metric was developed which can be used for censored survival data type. Prediction of a new instance is made by multiplying the prediction of each classifier by its weights, then adding them to select the class with majority vote using (7), where; $c_{i,j}$ is the class value for the i^{th}

classifier and j^{th} patient, N is the total number of classifiers, w_i is the weight for the i^{th} classifier.

$$Decision = \sum_{i=1}^{N} c_{i,j} \times w_i \tag{7}$$

The issue here is how to determine the weights given to each classifier. Several methods have been proposed to calculate them, which is beyond the focus of this paper, however the most common approach depends on the training errors of each classifier. The weight is usually the reciprocal of this error. Though, in this paper the average of the p-value P_i of the log rank test for the training data was chosen instead due to the censoring nature of the datasets. Since, the average of the p-value for the training sets has a value that is close to zero, their reciprocal will be very large, and therefore, the logarithm of the reciprocal average $Pval$ is usually used to calculate the weight of each classifier in the pool as shown in (8). These weights are then normalized in order that their sum is equal to one

$$w_i = \frac{1}{avg(P_i)} \tag{8}$$

- **STEP 7** is the iterated filter selection (IFR) step that uses the ranking from step 3 to further reduce the number of the features used in the predictive model. The process is similar to the one used in [59]. It starts with the variable of highest score, and then each feature is added iteratively in order to enhance predictions. Both FSFS and IFR steps used the minimum p-value of the log rank test as a criterion for selection. It is commonly used in the medical field to examine if the risk groups predictions were separable and distinguishable. A p-value less than the significance level of 0.05 indicates that the risk groups are significantly different. Steps 3 and 7 are considered as hybrid FS approach. It combines the advantages of filter and wrapper FS methods.

Figure 2

2.5 Classification Models and Evaluation Metrics

The evaluation metrics that were employed to test the performance of the final selected model are discussed below.

- **Sensitivity (True positive rate)** is the portion of patients that were correctly classified as one (high risk of re-intervention) and the number of patients that actually went through re-intervention.
- **Log Rank Test** is a very popular statistical metric in the medical area. It is used to examine if any predictive model was capable of differentiating between the risk groups of patients or separating survival probabilities of patients treated with different medication. It uses chi squared test [60] to determine a score called p-value. P-value less than the significance level of 0.05 means that the two risk groups are separable and discriminative.
- **Concordance Index (**CI) is a discriminative statistical metric that examines if the survival estimates of the predictive model are concordant and distinguished. It calculates the portion of all couples of patients that survival predictions have correct sorting. Then, divide this part by the summation of all pairs of patients in which the event of interest had occurred to at least one of them, and that one must have observation time less than the other [61]. Greater CI values indicate better concordant predictions. The maximum value that could be reached is one.

2.6 Comparative Feature model selection methods

2.6.1 Akaike Information Criterion (AIC)

It was first introduced by Akaike in 1977 to evaluate the quality of candidates' models produced during model selection. AIC measures the distance between each nominated model and the true model (Kullback Leibler distance). Therefore, as the distance decreases, the value of this model increases [62]. The formula shown in equation (8) illustrates how AIC is calculated. It places a penalty to the number of parameters. The final model selected is the one with the minimum AIC.

$$AIC = -2 \cdot \ln(L) + 2 \cdot K \,, \tag{8}$$

where; L is the maximum likelihood of the model given the data and K is the number of parameters in a given model.

2.6.2 Bayesian information criterion (BIC)

It was first introduced by Schwarz in 1978 [63]. BIC evaluates the quality of each candidate model as well. Though, it inserts a penalty not only on the number of parameters, but also on the number of data examples which is not the case in AIC. Therefore, some researches prefer to use it especially when they have models of different sizes. It is calculated using the formula shown in equation (9):

$$BIC = -2\ln(L) + 2 \cdot K \ln(n) \,, \tag{9}$$

where; L is the maximum likelihood of the model given the data and K is the number of parameters in a given model, and n is the number of observations.

2.6.3 Least Absolute Shrinkage and Selection Operator (LASSO)

It was introduced by Robert Tibshirani in 1997 [64]. It is a L_1 penalized estimation method that shrinks the regression coefficients estimates β of Cox regression model towards zero using a tuning parameter λ which gives a penalty on their absolute values. This leads to removing the irrelevant variables from the predictive model. Shrinkage prevents over-fitting that may occur due to collinearity of the variables. The β coefficients of the predictive model are fitted by maximizing penalized partial log likelihood (*PPLL*) for all data with an absolute value LASSO penalty λ on β using equation (10):

$$PPLL_{\lambda}(\beta) = \sum_{i=1}^{n} \delta_i \left[(x_i^T \cdot \beta) - \log\left(\sum_{t_j \geq t_i} \exp(x_j^T \cdot \beta) \right) \right] - \lambda \|\beta\|_1 , \qquad (10)$$

where, δ is the censor indicator for patient i with variables x. $\lambda \geq 0$ and $\|\cdot\|_1$ stands for L_1 norm. λ equal to zero means no shrinkage and infinity means infinity shrinkage. *Penalized* R-software package was used for implementing LASSO. The tuning parameter was selected using likelihood cross validation optimization method.

3. Results of the Proposed MCS Hybrid Feature-Model Selection

3.1 Comparing the Results of the Proposed MCS Hybrid Feature-Model Selection Algorithm with all Features

The common way to select a model with reduced features is to employ the whole dataset. This may consequently lead to overoptimistic results. Resampling techniques such as K-fold cross validation, leave one out cross validation, and bootstrapping are used to overcome this problem and to quantify the quality of the final reduced model on part of the data that were not used in modeling. However, the latter two methods have high computational cost. Therefore, in this paper, five-fold iterated nested cross validation were used for the hybrid feature selection and stable MCS model construction using center 1 data. Center 2 data were used to assess the performance of the final reduced model. The results of the MCS hybrid feature selection based on simple majority voting and weighted majority voting techniques for Center 2 predictions are compared with the full size of the model as shown in Tables 1 and 2.

Table 1

Table 1 shows that the proposed MCS hybrid FS technique based on simple majority voting has reduced the number of features from 45 to 27, 15 and 7 after all steps of proposed approach. Moreover, the concordance index (CI) of the full model is 0.6599 which has increased to 0.6630, 0.6657, and finally 0.6793 after hybrid FS steps. The p-value of the log-rank test has been reduced as well from 0.0331 to 0.0166, 0.0075 and 0.00016 after all steps of the proposed technique, which indicates an enhancement in the performance of the MCS model with the hybrid FS. Finally, the sensitivity was enhanced during all steps of the hybrid approach from 0.423 to finally reach 0.808. Note that, the event of interest in this paper is the risk of re-intervention after the EVAR surgery. Therefore, uncensored patients that experienced EVAR operation have definitely a class value of 1, while the rest are censored (their class value are not guaranteed to be 1 or 0). For this reason, the sensitivity metric was employed for comparing proposed predictive models. It indicates the ability of the proposed techniques to correctly classify the event of interest which is the minority class. CI is used as well, as it is a survival metric used for measuring survival model performance. Both metrics were used together as a predictive model with both higher CI and sensitivity rates indicate better ability to predict the risk of re-intervention and discriminate between risk groups.

Table 2

Table 2 shows that the proposed MCS hybrid FS approach based on weighted majority voting has reduced the number of features from 45 to 27, 17 and 6 after all steps of proposed approach. Moreover, the CI of the full model is 0.6710 which has increased to 0.6762, 0.6793, and finally 0.6808 after the hybrid FS steps, which are greater than that of the unweighted majority voting in

Table 1 (0.6599, 0.6630, 0.6657, and 0.6793). The p-value of the log-rank test has been reduced as well from 0.014 to 0.001, 0.0008 and 0.000038 after all steps of the proposed technique, which indicates an enhancement in the performance of the MCS model based on weighted voting with the hybrid FS compared to unweighted majority voting which has reached a final p-value of 0.00016. In addition, the sensitivity has increased from 0.423 to reach 0.7308.

3.2 Comparing the Results of the Proposed MCS Hybrid Algorithm with the Performance of the Individual Classifiers

In this section, the performances of the MCS hybrid FS algorithm and individual classifiers used to construct it are compared. As shown in Table 3, the MCS based on simple majority voting, weighted majority voting , and single classifiers have reduced the feature space to 7,6,5,5,6 for MCS based on simple majority voting, MCS based on weighted majority voting, and individual SVM, MLP, and KNN models, respectively. Predictions of Center 2 are used for comparison as it was not used in constructing and training the predictive model. The MCS based on weighted majority voting has outperformed the unweighted majority voting in both CI (0.6808 vs. 0.6793) and p-value of the log rank test (0.000038 vs. 0.00016); however, the later has higher sensitivity (0.808 vs. 0.7308). Moreover, the MCS hybrid FS approach using unweighted and weighted majority voting methods outperformed the other individual classifiers in p-value (0.00016 and 0.000038 vs. 0.00085, 0.00073, and 0.0011). However, the MLP's CI (0.6813) is better than MCS, SVM, and KNN (0.6793 and 0.6808, 0.6776, and 0.6411).

Table 3

3.3 Comparing the Results of the Proposed MCS Hybrid Algorithm with Performance of Cox's Model Using AIC, BIC and LASSO

In this section, the results of the MCS hybrid feature selection based on simple and weighted majority voting are compared with the state of art variable selection methods based on the Cox's regression model which are AIC, BIC and LASSO penalized methods. It is well known that the Cox's output is continuous. In order to translate this output to binary representing the risk group, the estimated parameters of the final reduced model are multiplied by each variable to generate a risk score. A value above the threshold indicates high risk (class value of 1) and vice versa. The one used for LASSO is 6.7 which is equivalent to mean of the risk score; while for other methods they are 2.4 and 3.1. The same threshold is applied to Center 2 data.

As shown in Table 4. The number of features of the final MCS model is seven for simple majority voting and six for weighted voting which are better than 14 for AIC and BIC, but equal or smaller than seven of LASSO. For Center 1 prediction, the CI of MCS based on weighted majority voting (0.7881), which is higher than simple majority voting (0.7521), BIC (0.7624) and LASSO (0.738), but smaller than AIC (0.7898). All models have p-value lower than 0.0001, which indicates that they are all capable of separating the two risk groups of Center 1. The sensitivity of MCS model using unweighted majority voting (0.84) and weighted majority voting (0.87) are greater than that of the other methods (0.69, 0.38, and 0.714). Moreover, for Center 2 predictions, the proposed MCS technique beats the other techniques in both the p-value of the log rank test (0.00016 and 0.000038 vs. 0.034, 0.029, and 0.0068) and the CI (0.6793 and 0.6808 vs. 0.6103, 0.630, and 0.6153). The main advantage in the MCS hybrid FS algorithm appears in the sensitivity results (0.808 and 0.7308 vs. 0.35, 0.23, and 0.5), which indicates that it can correctly classify more patients than did the re-intervention (the event of interest in this study). Thus, it is favored than the other methods.

Table 4

Figures 3 and 4 show the KM curves for the two risk groups predictions of both centers using the MCS hybrid FS technique based on simple and weighted majority voting compared with KM curves for the two risk groups predictions of both centers with AIC (Figure 5), BIC (Figure 6) and LASSO (Figure 7) Cox's models. Figure 3 indicates that the MCS model based on unweighted model classified 163 and 126 of Center 1 (upper) and Center 2 (lower) patients as high risk, which is equivalent to 36% and 44% of total Center 1 and Center 2 patients. Moreover, Figure 4 shows that the MCS model based on simple majority voting model classified 177 and 101 of Center 1 (upper) and

Center 2 (lower) patients as high risk, which is equivalent to 38 % and 35% of total Center 1 and Center 2 patients. The classification of the MCS model is better than the prediction of the AIC model (104 high risk patients equivalent to 23%) for Center 1 (Figure 5 upper) and (41 high risk patients equivalent to 14%) for Center 2 (Figure 5 lower), the BIC model in (58 high risk patients equivalent to 13%) for Center 1 (Figure 6 upper) and (25 high risk patients equivalent to 9%) for Center 2(Figure 6 lower), and the LASSO model (196 high risk patients equivalent to 43%) for Center 1 (Figure 7 upper), and (76 high risk patients equivalent to 26%) for Center 2 (Figure 7 lower).

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7

4. Discussion

Features that were selected using simple (unweighted) majority voting are the total aneurysm neck volume, maximum aneurysm neck diameter, diameter of the left common iliac artery 1 and 5 mm below internal iliac ostium, maximum iliac tortuosity index, diameter of the right common iliac artery 1mm below Internal iliac ostium, and right common iliac artery non luminal volume. Moreover, features resulted from weighted voting are the maximum common iliac aneurysm area, aneurysm neck diameter 10 mm below lowest renal, aneurysm neck length, common Iliac artery diameter 1 and 5 mm proximal to internal iliac origin, and right iliac tortuosity index. These features were reviewed by the clinical investigators. They confirmed that these variables have good face validity in terms of predicting technically difficult or challenging morphology for endografts currently available. It is well known that hostile sealing zones both proximally (at the aortic neck) or distally (at the common iliac artery) pose considerable technical challenges for durable endograft seal, and therefore it is plausible that the features selected (aortic neck area; and various aspects of iliac morphology) might be predictive of poor long-term clinical performance. Predictions using these features are clinically feasible and make excellent sense. However, weighted majority makes more sense as it includes neck length which is often thought of by surgeons planning the case [65-67]. Moreover, the concordance index and sensitivity rates are very promising and would have clinical importance if used prospectively. Also, the assignment of most patients to a low risk group counts well with clinical practice in which less patients will have re-intervention over five years [68].

5. Conclusion

Two datasets (743 patients) were collected from patients undergoing endovascular aortic surgery over the observation period from 2004 to 2010 in two separate vascular centers located in the UK (St George and Leicester hospitals). They were capable of building and validating a multiple classifier predictive model to predict the long-term risk of aortic complications after EVAR. The paper has offered a successful solution to the high level of censoring. This solution was used with the proposed hybrid feature model selection approach to reduce the number of features needed to construct it with censored survival data type. Moreover, the predictive model may be used for cross-centers prediction as well, as it was constructed and evaluated by patients of two different centers. The model will enable doctors to take decisions about future follow up observation plan for each patient. High risk patients will have to undergo more regular surveillance than low risk patients.

In the proposed technique, the instability that might occur during FS, MS and MCS construction was reduced using iterated nested cross validation. The uncensoring issue was solved using Bayesian networks. Two MCS models were constructed using three popular machine learning classifiers (SVM, MLP and KNN) combined with simple and weighted majority voting based on survival analysis metric. Machine learning techniques cannot be used directly with censored survival data. Therefore, the proposed approach make these MCSs constructed using machine learning techniques have the ability to be used with censored survival data. The MCSs constructed were capable of predicting the risk of re-intervention after EVAR. Their performances were compared with both individual classifiers and the statistical Cox's model. Three well-known model selection techniques called AIC, BIC and LASSO were used with Cox's regression model for comparison with the MCS hybrid feature selection approach. The same searching strategy was used for the selection in AIC and BIC.

The results have shown that MCS using simple and weighted voting outperformed both individual classifiers and Cox's model selection methods in both p-values and CI expect for the CI of MLP for Center 2. It successively separated between the risks groups for both centers as the p-value of the log rank test was less than 0.0001 for Center 1 and 0.00016 and 0.000038 for Center 2 using simple and weighted voting, In addition, the CI has increased from 0.6559 and 0.6710 to finally reach 0.6793 and 0.6808 with sensitivity of 0.808 and 0.7308 which allows it to be used for cross-center prediction. Moreover, the proposed technique has a higher sensitivity as compared to other techniques which make it stronger than the other ones in classifying the long term risk of aortic complications after EVAR for new patients. Therefore, it can be used by doctors to facilitate the future follow up plan decision. Patients with high risk prediction will be more monitored than other ones which prevent low risk patients to be exposed to excess harmful radiations.

6. Acknowledgments

We would like to acknowledge Prof. David Lowe at Aston University for his guidance and useful discussions on this research.

7. Conflict of interests

None of the authors has conflict of interests to disclose or competing interests to declare.

8. References

[1] M. Blachnik, "Comparison of various feature selection methods in application to prototype best rules," in *Computer Recognition Systems 3*. vol. 57, ed New York: Springer, 2009, pp. 257-264.

[2] S. Kul, "The use of survival analysis for clinical pathways," *International Journal of Care Pathways,* vol. 14, pp. 23-26, 2010.

[3] K. M. Leung, R. M. Elashoff, and A. A. Afifi, "Censoring issues in survival analysis," *Annual Review of Public Health,* vol. 18, pp. 83-104, 1997.

[4] B. Zupan, J. DemšAr, M. W. Kattan, J. R. Beck, and I. Bratko, "Machine learning for survival analysis: a case study on recurrence of prostate cancer," *Artificial intelligence in medicine,* vol. 20, pp. 59-75, 2000.

[5] Y. Liu, U. Aickelin, J. Feyereisl, and L. G. Durrant, "Wavelet feature extraction and genetic algorithm for biomarker detection in colorectal cancer data," *Knowledge-Based Systems,* vol. 37, pp. 502-514, 2013.

[6] A. Eleuteri, Tagliaferri, R,Milano, L, Sansone, G, D'Agostino, D, De Placido, S, De Laurentiis, M, "Survival analysis and neural networks," presented at the Proceedings of the International Joint Conference on Neural Networks, Portland, OR, 2003.

[7] E. Biganzoli, P. Boracchi, L. Mariani, and E. Marubini, "Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach," *Statistics in Medicine,* vol. 17, pp. 1169-1186, 1998.

[8] I. Štajduhar and B. Dalbelo-Bašić, "Uncensoring censored data for machine learning: A likelihood-based approach," *Expert Systems with Applications,* vol. 39, pp. 7226-7234, 2012.

[9] P. J. Lisboa, H. Wong, P. Harris, and R. Swindell, "A Bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer," *Artificial Intelligence in Medicine,* vol. 28, pp. 1-25, 2003.

[10] Q. Tan, M. Thomassen, K. M. Jochumsen, O. Mogensen, K. Christensen, and T. A. Kruse, "Gene selection for predicting survival outcomes of cancer patients in microarray studies," in *Advances in Computer and Information Sciences and Engineering*, ed Netherlands: Springer, 2008, pp. 405-409.

[11] R. Blanco, I. Inza, M. Merino, J. Quiroga, and P. Larrañaga, "Feature selection in Bayesian classifiers for the prognosis of survival of cirrhotic patients treated with TIPS," *Journal of Biomedical Informatics,* vol. 38, pp. 376-388, 2005.

[12] H. Neuvirth, Michal Ozery-Flato, Jianying Hu, Jonathan Laserson, Martin S. Kohn, Shahram Ebadollahi, and Michal Rosen-Zvi, "Toward personalized care management

of patients at risk: the diabetes case study," presented at the 17th ACM SIGKDD San Diego, CA,USA, 2011.

[13] Y.-C. Chen, W.-C. Ke, and H.-W. Chiu, "Risk classification of cancer survival using ANN with gene expression data from multiple laboratories," *Computers in Biology and Medicine,* vol. 48, pp. 1-7, 2014.

[14] D. H. Wolpert, "The supervised learning no-free-lunch theorems," presented at the Proceedings of the 6th Online World Conference on Soft Computing and Industry, 2001.

[15] M. Woźniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Information Fusion*, vol. 16, pp. 3-17, 2014.

[16] K. M. Ting, & Quek, R. J. Y., "Model stability: a key factor in determining whether an algorithm produces an optimal model from a matching distribution," in *Third IEEE International Conference on Data Mining, ICDM 2003.* , Melbourne, Florida, USA., 2003, pp. 653 - 656.

[17] M. Termenon and M. Graña, "A two stage sequential ensemble applied to the classification of Alzheimer's disease based on mri features," *Neural processing letters,* vol. 35, pp. 1-12, 2012.

[18] A. Savio, M. T. García-Sebastián, D. Chyzyk, C. Hernández, M. Graña, A. Sistiaga, *et al.*, "Neurocognitive disorder detection based on feature vectors extracted from VBM analysis of structural MRI," *Computers in biology and medicine,* vol. 41, pp. 600-610, 2011.

[19] D. Chyzhyk, M. Graña, A. Savio, and J. Maiora, "Hybrid dendritic computing with kernel-LICA applied to Alzheimer's disease detection in MRI," *Neurocomputing,* vol. 75, pp. 72-77, 2012.

[20] C. O. Plumpton, L. I. Kuncheva, N. N. Oosterhof, and S. J. Johnston, "Naive random subspace ensemble with linear classifiers for real-time classification of fMRI data," *Pattern Recognition,* vol. 45, pp. 2101-2108, 2012.

[21] C. Cabral, M. Silveira, and P. Figueiredo, "Decoding visual brain states from fMRI using an ensemble of classifiers," *Pattern Recognition*, vol. 45, pp. 2064-2074, 2012.

[22] W. G. Baxt, "Improving the accuracy of an artificial neural network using multiple differently trained networks," *Neural Computation,* vol. 4, pp. 772-780, 1992.

[23] L. Nanni, "Ensemble of classifiers for protein fold recognition," *Neurocomputing,* vol. 69, pp. 850-853, 2006.

[24] T. Yang, V. Kecman, L. Cao, C. Zhang, and J. Z. Huang, "Margin-based ensemble classifier for protein fold recognition," *Expert Systems with Applications,* vol. 38, pp. 12348-12355, 2011.

[25] G.-Z. Li, T.-Y. Liu, and V. S. Cheng, "Classification of brain glioma by using SVMs bagging with feature selection," in *Data Mining for Biomedical Applications*, ed: Springer, 2006, pp. 124-130.

[26] J.-H. Eom, S.-C. Kim, and B.-T. Zhang, "AptaCDSS-E: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction," *Expert Systems with Applications,* vol. 34, pp. 2465-2479, 2008.

[27] R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," *Expert systems with applications,* vol. 36, pp. 7675-7680, 2009.

[28] R. Das, I. Turkoglu, and A. Sengur, "Diagnosis of valvular heart disease through neural networks ensembles," *computer methods and programs in biomedicine,* vol. 93, pp. 185-191, 2009.

[29] S. D. Bay, "Nearest neighbor classification from multiple feature subsets," *Intelligent data analysis,* vol. 3, pp. 191-209, 1999.

[30] M. A. Mazurowski, P. A. Habas, J. M. Zurada, J. Y. Lo, J. A. Baker, and G. D. Tourassi, "Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance," *Neural networks,* vol. 21, pp. 427-436, 2008.

[31] B. Twala, "Multiple classifier application to credit risk assessment," *Expert Systems with Applications,* vol. 37, pp. 3326-3336, 2010.

[32] G. Wang, J. Hao, J. Ma, and H. Jiang, "A comparative assessment of ensemble learning for credit scoring," *Expert Systems with Applications,* vol. 38, pp. 223-230, 2011.

[33] M. Sattlecker, R. Baker, N. Stone, and C. Bessant, "Support vector machine ensembles for breast cancer type prediction from mid-FTIR micro-calcification spectra," *Chemometrics and Intelligent Laboratory Systems,* vol. 107, pp. 363-370, 2011.

[34] C.-F. Tsai, Y.-F. Hsu, and D. C. Yen, "A comparative study of classifier ensembles for bankruptcy prediction," *Applied Soft Computing,* vol. 24, pp. 977-984, 2014.

[35] F. L. Moll, J. Powell, G. Fraedrich, F. Verzini, S. Haulon, M. Waltham*, et al.*, "Management of abdominal aortic aneurysms clinical practice guidelines of the European society for vascular surgery," *European Journal of Vascular and Endovascular Surgery,* vol. 41, pp. S1-S58, 2011.

[36] L. C. Brown, J. T. Powell, S. G. Thompson, D. M. Epstein, M. J. Sculpher, and R. M. Greenhalgh, "The UK EndoVascular Aneurysm Repair (EVAR) trials: design, methodology and progress," *European Journal of Vascular and Endovascular Surgery,* vol. 27, pp. 372-381, 2004.

[37] N. Hay, F. McCracken, J. Richardson, E. George, and D. Barnett, "Endovascular stent-grafts for the treatment of abdominal aortic aneurysms: NICE technology appraisal guidance," *Heart,* vol. 95, pp. 1798-800, Nov 2009.

[38] A. Karthikesalingam, A. A. Page, C. Pettengell, R. J. Hinchliffe, I. M. Loftus, M. M. Thompson*, et al.*, "Heterogeneity in surveillance after endovascular aneurysm repair in the UK," *Europian Journal of Vascular and Endovascular Surgery,* vol. 42, pp. 585-90, Nov 2011.

[39] R. A. Weerakkody, S. R. Walsh, C. Cousins, K. E. Goldstone, T. Y. Tang, and M. E. Gaunt, "Radiation exposure during endovascular aneurysm repair," *British Journal of Surgery,* vol. 95, pp. 699-702, Jun 2008.

[40] A. Karthikesalingam, P. J. Holt, R. J. Hinchliffe, I. M. Nordon, I. M. Loftus, and M. M. Thompson, "Risk of reintervention after endovascular aortic aneurysm repair," *British Journal of Surgery,* vol. 97, pp. 657-63, May 2010.

[41] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society. Series B,* vol. 34, pp. 187–220, 1972.

[42] B. O. Patterson, R. J. Hinchliffe, P. J. Holt, I. M. Loftus, and M. M. Thompson, "Importance of aortic morphology in planning aortic interventions," *Journal of Endovascular Therapy,* vol. 17, pp. 73-7, Feb 2010.

[43] B. O. Patterson, P. J. Holt, R. Hinchliffe, I. M. Nordon, I. M. Loftus, and M. M. Thompson, "Existing risk prediction methods for elective abdominal aortic aneurysm repair do not predict short-term outcome following endovascular repair," *Journal of Vascular Surgery,* vol. 52, pp. 25-30, Jul 2010.

[44] T. Ghatwary, Karthikesalingam, A., Patterson, B., Hinchliffe, R., Morgan, R., Loftus, I., Salem, A., Thompson, M. M., Holt, P. J., "St George's Vascular Institute Protocol:

an accurate and reproducible methodology to enable comprehensive characterization of infrarenal abdominal aortic aneurysm morphology in clinical and research applications," *Journal of Endovascular Therapy,* vol. 19, pp. 400-14, Jun 2012.

[45] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *Journal of the American statistical association,* vol. 53, pp. 457-481, 1958.

[46] H.-J. Kim, "Common factor analysis versus principal component analysis: choice for symptom cluster research," *Asian Nursing Research,* vol. 2, pp. 17-24, 2008.

[47] L. Lam, "Classifier combinations: implementations and theoretical issues," in *Proceedings of the First International Workshop on Multiple Classifier Systems, MCS '00,* Cagliari, Italy, 2000, pp. 77–86.

[48] A. Karthikesalingam, Attallah, Omneya, Ma, Xianghong, Bahia, Sandeep Singh, Thompson, Luke, Vidal-Diez, Alberto, Choke, Edward C, Bown, Matt J, Sayers, Robert D, Thompson, Matt M, Peter J. Holt, "An Artificial Neural Network Stratifies the Risks of Reintervention and Mortality after Endovascular Aneurysm Repair; a Retrospective Observational study," *PloS one,* vol. 10, p. e0129024, 2015.

[49] O. Attallah and X. Ma, "Bayesian neural network approach for determining the risk of re-intervention after endovascular aortic aneurysm repair," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine,* vol. 228, pp. 857-66, Sep 2014.

[50] B. B. Perry, "A genetic algorithm for learning Bayesian network adjacency matrices from data," PhD, Kansas State University, 2003.

[51] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning,* vol. 29, pp. 131-163, 1997.

[52] A. M. Hassan, "A probabilistic relaxation framework for learning Bayesian network structures from data," Master of Science Faculty of Engineering, Cairo University 2007.

[53] J. Peláez, J. Doña, J. Fornari, and G. Serra, "Ischemia classification via ECG using MLP neural networks," *International Journal of Computational Intelligence Systems,* vol. 7, pp. 344-352, 2014.

[54] J. Xie, J. Lei, W. Xie, Y. Shi, and X. Liu, "Two-stage hybrid feature selection algorithms for diagnosing erythemato-squamous diseases," *Health Information Science and Systems,* vol. 1, 2013.

[55] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy k-nearest neighbor algorithm," *Systems, Man and Cybernetics, IEEE Transactions on,* pp. 580-585, 1985.

[56] D. Lowsky, Ding, Y, Lee, DKK, McCulloch, CE, Ross, LF, Thistlethwaite, JR, Zenios, SA, "AK-nearest neighbors survival probability prediction method," *Statistics in Medicine,* vol. 32, pp. 2062-2069, 2013.

[57] H. Kim and M. Bredel, "Feature selection and survival modeling in The Cancer Genome Atlas," *International Journal of Nanomedicine,* vol. 8, p. 57, 2013.

[58] I. H. Witten and E. Frank, *Data mining : practical machine learning tools and techniques,* 2nd ed. Amsterdam ; Boston, MA: Morgan Kaufman, 2005.

[59] I. Choi, B. J. Wells, C. Yu, and M. W. Kattan, "An empirical approach to model selection through validation for censored survival data," *Journal of Biomedical Informatics,* vol. 44, pp. 595-606, 2011.

[60] S. R. Rao and D. A. Schoenfeld, "Survival methods," *Circulation,* vol. 115, pp. 109-113, 2007.

[61] F. E. Harrell, Jr., K. L. Lee, and D. B. Mark, "Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors," *Statistics in Medicine,* vol. 15, pp. 361-87, Feb 28 1996.

[62] K. P. Burnham, D. R. Anderson, and K. P. Burnham, *Model selection and multimodel inference : a practical information-theoretic approach*, 2nd ed. New York: Springer, 2002.

[63] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics,* vol. 6, pp. 461-464, 1978.

[64] R. Tibshirani, "The lasso method for variable selection in the Cox model," *Statistics in Medicine,* vol. 16, pp. 385-395, 1997.

[65] W. D. Jordan, Ouriel, Kenneth, Mehta, Manish, Varnagy, David, Moore, William M, Arko, Frank R, Joye, James, de Vries, Jean-Paul PM, "Outcome-based anatomic criteria for defining the hostile aortic neck," *Journal of Vascular Surgery,* 2015.

[66] P. W. Stather, J. B. Wild, R. D. Sayers, M. J. Bown, and E. Choke, "Endovascular aortic aneurysm repair in patients with hostile neck anatomy," *Journal of Endovascular Therapy,* vol. 20, pp. 623-637, 2013.

[67] G. A. Antoniou, G. S. Georgiadis, S. A. Antoniou, G. Kuhan, and D. Murray, "A meta-analysis of outcomes of endovascular abdominal aortic aneurysm repair in patients with hostile and friendly neck anatomy," *Journal of Vascular Surgery,* vol. 57, pp. 527-538, 2013.

[68] A. Karthikesalingam, S. Markar, P. Holt, and R. Praseedom, "Meta-analysis of randomized controlled trials comparing laparoscopic with open mesh repair of recurrent inguinal hernia," *British Journal of Surgery,* vol. 97, pp. 4-11, 2010.

Table Captions

Table 1: Results of the proposed MCS using Simple Majority Voting on the testing set (center 2) after the two steps of hybrid feature selection.

Table 2: Results of the proposed MCS using Weighted Majority Voting on the testing set (center 2) after the two steps of hybrid feature selection.

Table 3: Performance of the proposed MCS on the testing dataset (center 2) compared with individual classifiers after hybrid feature selection.

Table 4: Results of the proposed MCS after hybrid feature selection compared with Cox's model using AIC, BIC, and LASSO

Figure Captions

Figure 1. Kaplan Meier curves for center 1 (Upper) and center 2 (Lower)

Figure 2. Flow chart of the proposed algorithm

Figure 3. Kaplan Meier curves for the two risk groups predictions of (upper)center1 and (lower) center 2 using the MCS hybrid FS technique based on simple majority voting.

Figure 4. Kaplan Meier curves for the two risk groups predictions of (upper)center1 and (lower) center 2 using the MCS hybrid FS technique based on weighted majority voting

Figure 5. Kaplan Meier curves of the predictions of the risk groups for center 1 (Upper) and center2 (Lower) using Cox's model with AIC

Figure 6. Kaplan Meier curves of the predictions of the risk groups for center 1 (Upper) and center 2 (Lower) using Cox's model with BIC

Figure 7. Kaplan Meier curves of the predictions of the risk groups for center 1 (Upper) and center 2 (Lower) using Cox's model with LASSO

Table 1: Results of the proposed MCS using Simple Majority Voting on the testing set (center 2) after the two steps of hybrid feature selection

Proposed algorithm	Number of features	p-value (Log rank test)	CI (Standard Deviation SD)	Sensitivity
MCS All Features	45	0.0331	0.6599 (0.0634)	0.423
MCS FA step	27	0.0166	0.6630 (0.0571)	0.461
MCS FSFS step	15	0.0075	0.6657 (0.0732)	0.654
MCS IFR step	7	0.00016	0.6793 (0.0556)	0.808

Table 2: Results of the proposed MCS using Weighted Majority Voting on the testing set (center 2) after the two steps of hybrid feature selection

Proposed algorithm	Number of features	p-value (Log rank test)	CI(SD)	Sensitivity
MCS All Features	45	0.014	0.6710 (0.0572)	0.423
MCS FA step	27	0.0010	0.6762 (0.0643)	0.539
MCS FSFS step	17	0.0008	0.6793(0.0573)	0.615
MCS IFR step	6	0.000038	0.6808 (0.0528)	0.7308

Table 3: Performance of the proposed MCS on the testing dataset (center 2) compared with individual classifiers after hybrid feature selection

Classifier	Number of final features	p-value (Log rank test)	CI (SD)	Sensitivity
MCS Simple Majority Voting	7	0.00016	0.6793 (0.0556)	0.808
MCS Weighted Majority Voting	6	0.000038	0.6808 (0.0528)	0.7308
SVM	5	0.00039	0.6776 (0.0499)	0.7308
MLP	5	0.00073	0.6817 (0.0804)	0.7308
KNN	6	0.0011	0.6411 (0.0628)	0.6538

Table 4: Results of the proposed MCS after hybrid feature selection compared with Cox's model using AIC, BIC, and LASSO

Technique	**Model Size**	**p-value (Log rank test)**		**CI (SD)**		**Sensitivity**	
		Center 1	**Center 2**	**Center 1**	**Center 2**	**Center 1**	**Center 2**
Simple Majority Voting MCS Hybrid FS	7	<0.0001	0.00016	0.7521 (0.0332)	0.6793 (0.0556)	0.84	0.808
Weighted Majority Voting MCS Hybrid FS	6	<0.0001	0.000038	0.7881 (0.0337)	0.6808 (0.0528)	0.87	0.7308
AIC Cox FS	14	<0.0001	0.034	0.7898 (0.0408)	0.6103 (0.0725)	0.69	0.35
BIC Cox FS	14	<0.0001	0.029	0.7624 (0.0465)	0.630 (0.0685)	0.38	0.23
LASSO Cox FS	7	<0.0001	0.0068	0.7382 (0.0426)	0.6153 (0.0864)	0.714	0.50

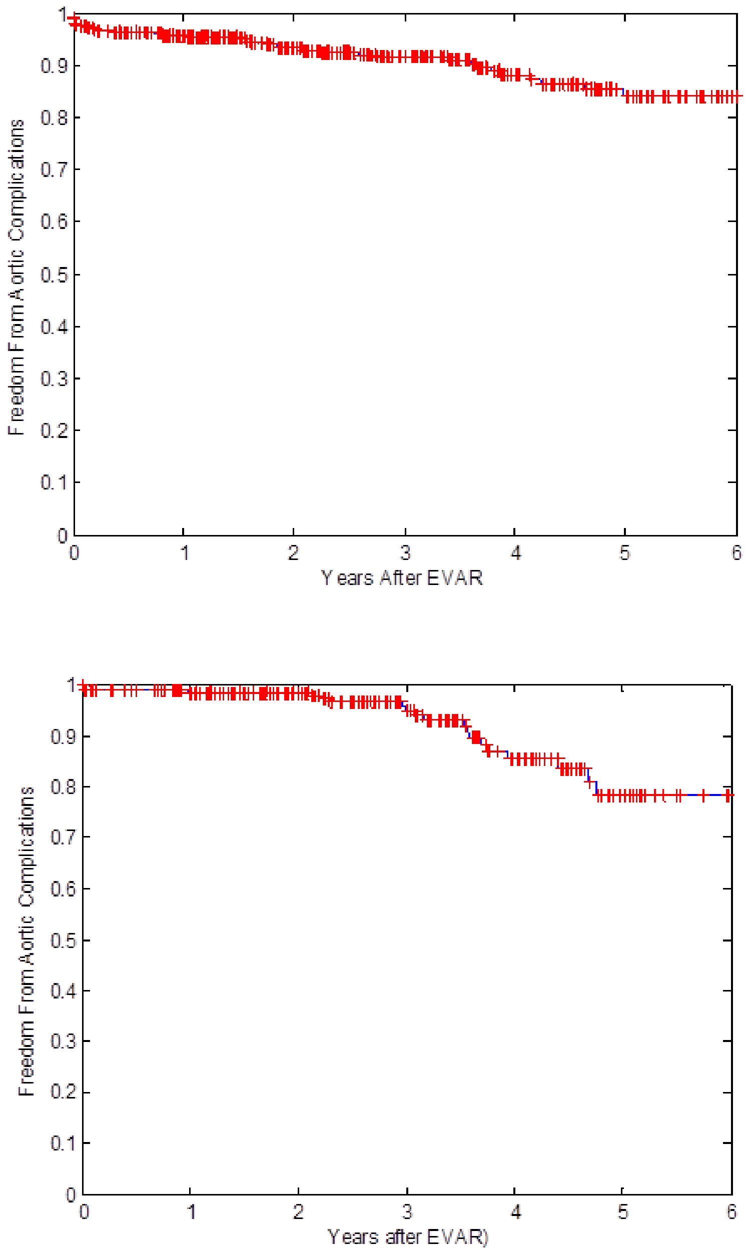


Figure 1 (Attallah, O.) Kaplan Meier curves for center 1 (Upper) and center 2 (Lower)

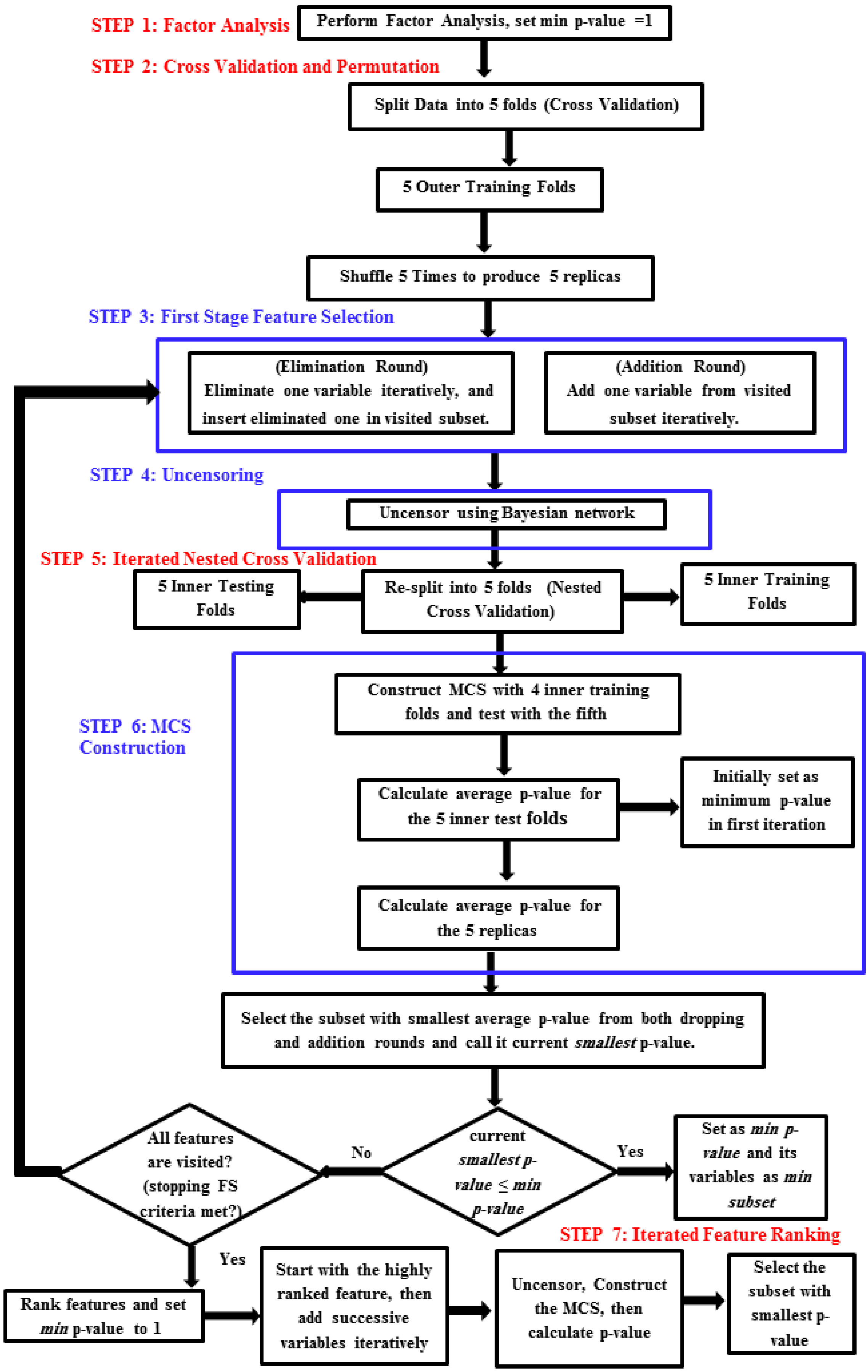
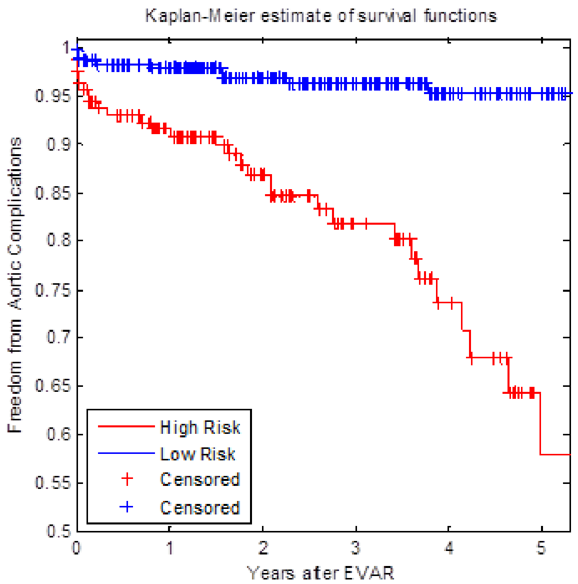
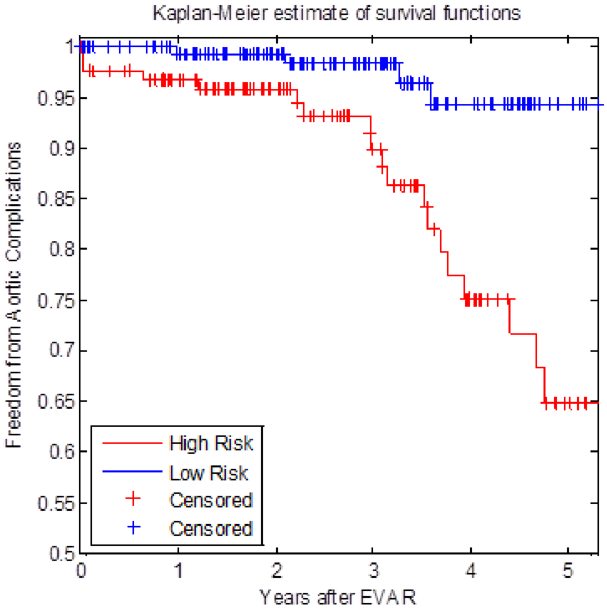


Figure 2 (Attallah, O.) Flowchart of the proposed algorithm

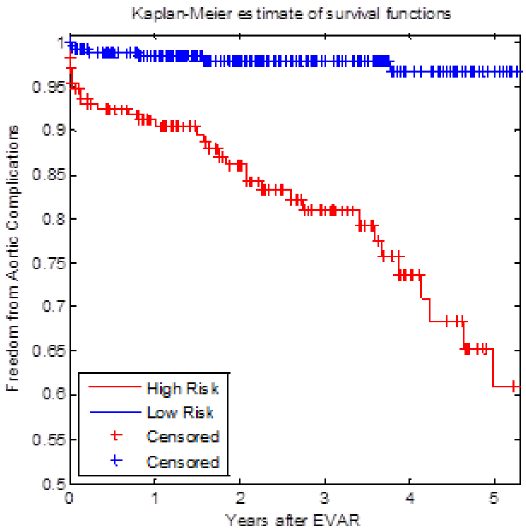


Year		0	1	2	3	4	5
Freedom from Aortic Complications	Low-risk	-	97.9%	97%	96.3%	95.3%	95.3%
	High-Risk	-	90.8%	87%	81.8%	73.6%	51.5%
Number at Risk	Low-risk	294	243	181	125	78	43
	High-Risk	163	118	81	51	28	9

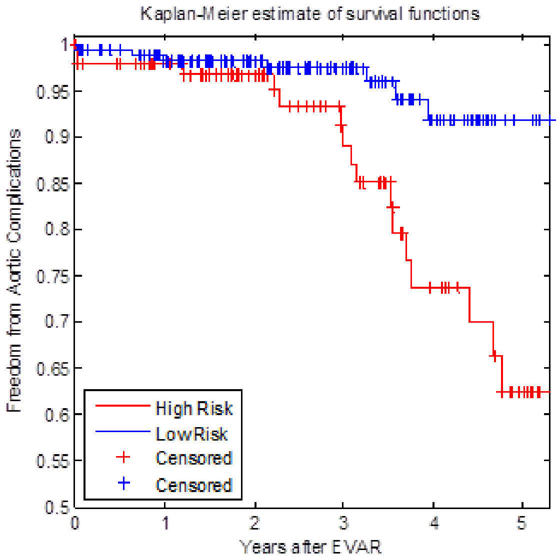


Year		0	1	2	3	4	5
Freedom from Aortic Complications	Low-risk	-	99.3%	99.3%	98.3%	94.3%	94.3%
	High-Risk	-	96.7%	95.8%	89.9%	75.2%	64.9%
Number at Risk	Low-risk	160	140	106	66	33	5
	High-Risk	126	108	78	54	29	11

Figure 3 (Attallah, O.) Kaplan Meier curves for the two risk groups predictions of (upper)center1 and (lower) center 2 using the MCS hybrid FS technique based on simple majority voting

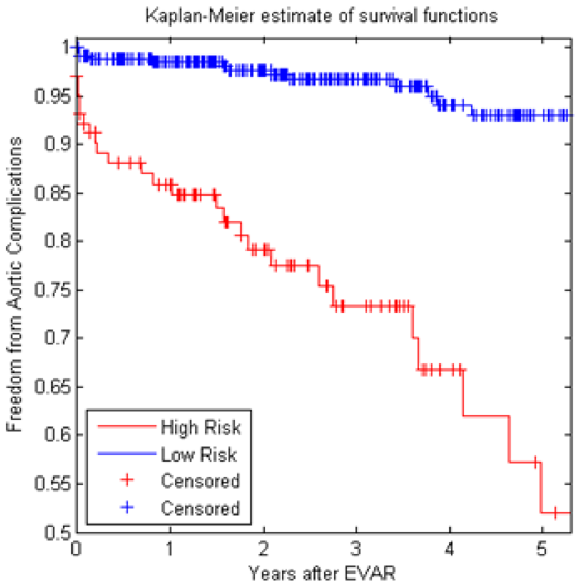


Year		0	1	2	3	4	5
Freedom from Aortic Complications	Low-risk	-	98.5%	97.8%	97.8%	96.8%	96.8%
	High-Risk	-	90.5%	86.2%	81%	73.6%	60.9%
Number at Risk	Low-risk	280	232	169	117	74	39
	High-Risk	177	129	93	59	32	14

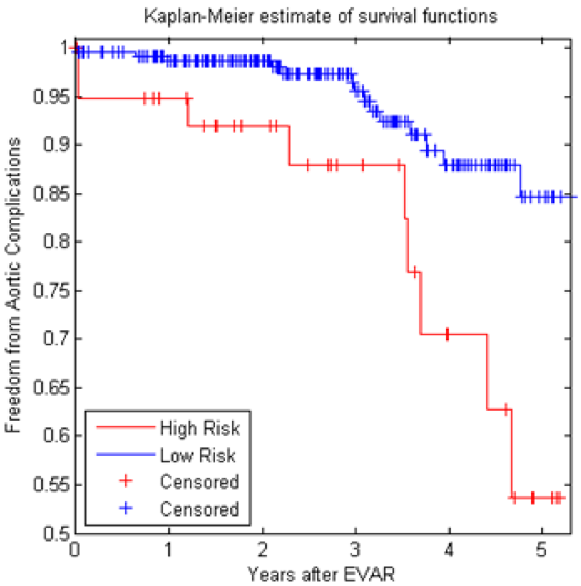


Year		0	1	2	3	4	5
Freedom from Aortic Complications	Low-risk	-	98.3%	98.3%	97.4%	91.8%	91.8%
	High-Risk	-	98%	96.8%	87.2%	73%	62.4%
Number at Risk	Low-risk	185	163	122	77	38	11
	High-Risk	101	85	62	43	24	6

Figure 4 (O.Attallah) Kaplan Meier curves for the two risk groups predictions of (upper)center1 and (lower) center 2 using the MCS hybrid FS technique based on weighted majority voting

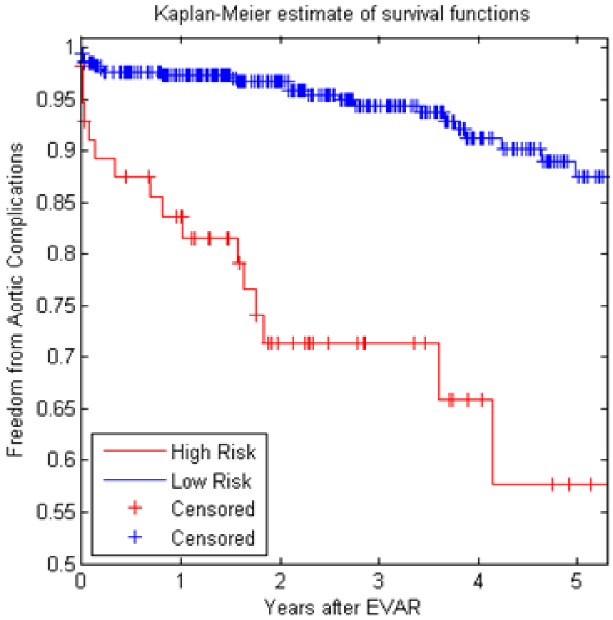


Year		0	1	2	3	4	5
Freedom from Aortic Complications	Low-risk	-	98.5%	97.7%	96.7%	94%	93%
	High-Risk	-	86%	79%	73.4%	66.73%	52%
Number at Risk	Low-risk	353	284	211	144	90	42
	High-Risk	104	77	51	32	16	9

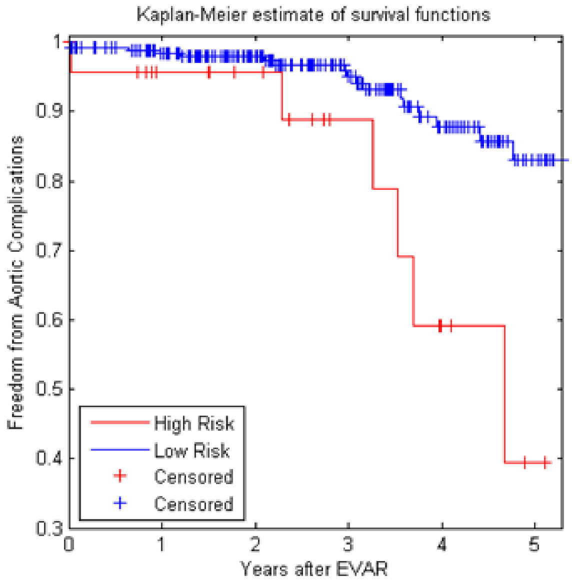


Year		0	1	2	3	4	5
Freedom from Aortic Complications	Low-risk	-	98.7%	98.7%	95.5%	88%	84.6%
	High-Risk	-	94.8%	92%	88%	70.5%	53.7%
Number at Risk	Low-risk	245	214	159	102	53	15
	High-Risk	41	34	25	18	10	1

Figure 5 (Attallah, O.) Kaplan Meier curves of the predictions of the risk groups for center 1 (Upper) and center2 (Lower) using Cox's model with AIC

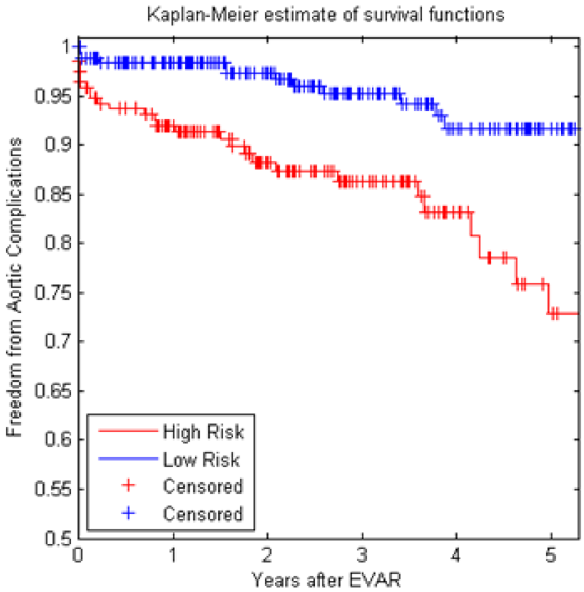


Year		0	1	2	3	4	5
Freedom from Aortic Complications	Low-risk	-	97.4%	96.7%	94.3%	91.2%	87.4%
	High-Risk	-	83.6%	71.4%	71.4%	66%	57.6%
Number at Risk	Low-risk	399	318	238	161	97	48
	High-Risk	58	43	24	16	9	5

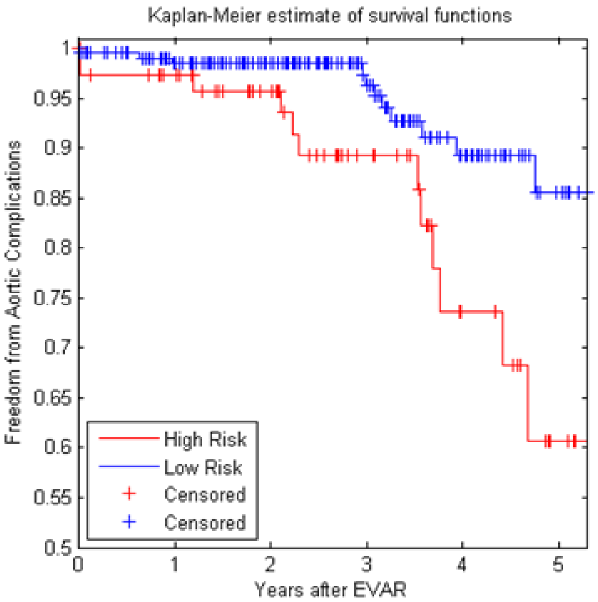


Year		0	1	2	3	4	5
Freedom from Aortic Complications	Low-risk	-	98.4%	98%	95%	87.8%	83%
	High-Risk	-	95.6%	95.6%	88.5%	89.2%	39.5%
Number at Risk	Low-risk	261	230	169	111	58	15
	High-Risk	25	19	15	9	4	1

Figure 6 (Attallah, O.) Kaplan Meier curves of the predictions of the risk groups for center 1 (Upper) and center 2 (Lower) using Cox's model with BIC



Year		0	1	2	3	4	5
Freedom from Aortic Complications	Low-risk	-	98.4%	97.3%	95.2%	91.7%	91.7%
	High-Risk	-	92%	88%	86.3%	83%	73%
Number at Risk	Low-risk	261	210	159	106	67	30
	High-Risk	196	151	103	70	39	23



Year		0	1	2	3	4	5
Freedom from Aortic Complications	Low-risk	-	98.5%	98.5%	96.3%	89.2%	85.5%
	High-Risk	-	97.3%	95.7%	89.2%	73.5%	60.7%
Number at Risk	Low-risk	210	182	134	89	47	15
	High-Risk	76	66	50	31	15	3

Figure 7 (Attallah, O.) Kaplan Meier curves of the predictions of the risk groups for center 1 (Upper) and center 2 (Lower) using Cox's model with LASSO