JCC
Journal of
CENTRUM
Cathedra

CENTRUM
CENTRO DE NEGOCIOS

An Analytical Approach to Predict the Performance of Thoracic Transplantations

Asil Oztekin
University of Massachusetts Lowell, MA, USA

Abstract

Predicting the performance of planned organ transplantation has proved to be a critical problem to solve. The purpose of this study is to present a data mining-based model for variable filtering and selection in order to predict the performance of thoracic transplantation via the graft survivability after the transplant. To this end, 10-fold cross-validated information fusion-based sensitivity analyses on machine learning models are conducted to receive an unbiased predictor variable ranking to be used in a subsequent Cox survival analysis. The study is unique in that it provides a mathematical means for medical experts to deal with thoracic recipients more efficiently and effectively.

Keywords: Prediction model, machine learning, United Network for Organ Sharing (UNOS)

JEL Classification codes: C3, C8

As the economies in developed countries are shifting from a manufacturing base toward a service orientation, the role of the service industry has become more important (Lee, Fiedler, & Smith, 2008). The healthcare sector is one of the most critical sectors in the service industry since it is life-crucial, and mistakes can cause inevitable and incurable results (Kaplan, 1987). Improper resource allocation has been one of the perennial problems in the healthcare service industry (Lee, Ng, & Zhang, 2007). Particularly, the allocation of scarce organs for organ transplantation has been one of the most critical problems faced in the healthcare service. Although organ transplantation is the sole viable therapy for various end-stage diseases, the number of donor organs unfortunately does not often meet the needs (Shechter et al., 2005). Therefore, the patients awaiting an organ transplant are lined up in waiting lists whereas some of the donor organs are wasted due to suboptimal matching between the donor and the recipient.

Long organ transplant waiting lists result mainly from two factors. First, since the success rate of organ transplantations has increased due to the advancements in the medical field, there is a corresponding increase in the number of patients asking for a transplant. In addition, the drastic increase of transplant centers throughout the United States has made them much more accessible to patients. While there were only four transplant centers in the late 20th century, as of 2008, there are 249 centers in the United States (UNOS, 2012). Second, the increase in the number of donated organs has never reached the level of the increase in demand, which results in a shortage of donor organs. This increasing gap between the number of patients waiting for an organ transplant and donor supplies has increased waiting times, which in turn has led to the death of patients on the waiting list (Abouna, 2003).

Organ transplantation is a vital treatment for the chronic failure of major organs. Survival analysis, which is defined as the surviving time after a patient receives transplantation surgery, has been the primary evaluation method for the effectiveness of such an operation. The primary objective of this study is to develop an integrated data mining methodology to predict accurately the survivability and to analyze the prognostic factors for different risk groups of transplant patients in order to discover novel patterns to augment clinical and biological studies. In doing so, we propose to use very large data sets with hundreds of determinative variables regarding the donors, the potential recipients, and transplantation procedures. While the main research goal can be summarized as to improve the effectiveness and efficiency of the organ transplantation procedures, the specific objectives of this study can be listed as follows:

1. To develop an integrated data mining methodology to build accurate predictive models for survivability and use these models to investigate the fundamental relationship between predictor variables and survivability in order to identify the factors that have the most significant impact on survivability;
2. To create a comprehensive prognostic index (PI) related to thoracic transplantation, determine risk groups of patients based on their survivability quantified using the developed PI, and identify the optimal setting so as to achieve better survivability.

Survivability prediction is becoming increasingly more important in medicine. When a resource is scarce, the need for accurate prediction becomes acute (Sheppard et al., 1999). In particular, the prediction of survival time and the prognosis prediction of medical treatments are clinically important and challenging problems (Lin, Horn, Hurdle, & Goldfarb-Rumyantzev, 2008). Organ scarceness requires the development of effective and efficient procedures to select the most optimal organ receiver since the demand for organs might not be satisfied. Hence, the first step is to develop a data mining-based method to reveal the knowledge underlying the huge amount of data collected and stored from organ transplantation procedures.

The main objective of this data mining-based method is to find patterns to maximize the survival time and to optimize the prognosis for the organ recipients. By processing large volumes of transplantation data, one may identify the important factors and their relationships to the survival of the graft and the patient. Thereafter, a PI is devised to classify the patients into different risk groups (Cox & Oakes, 1984). Predicting the thoracic survivability and classifying the patients (potential thoracic organ receivers) by means of a validated PI into different classes of risks would help decision makers in determining which patients should receive the scarce resources. Such a PI would enhance the effectiveness of existing procedures to identify the patient risk groups and prioritize them for receiving organs.

A large body of research exists for data-driven analytics in various organ transplantation cases. Kusiak, Dixon, and Shah (2005) conducted a study which compared two rule-based data mining techniques, namely decision trees and rough sets, for predicting the survival time of kidney dialysis patients. Their study presented not very high but considerable prediction accuracy rates. The main limitation of the study was the utilization of a small dataset with 188 patients in total, and many patient-related parameters were ignored. Hong et al. (2006) presented a survival analysis of liver transplant patients in Canada by considering only some of the determinative factors such as age, blood type, donor type (cadaveric or alive), race, and gender of recipients and donors. In addition to the variables limitation, they also admitted that the clinical information used in the study lacked many details.

Focusing specifically on thoracic transplantation, Jenkins et al. (2000) and Fernández-Yánez et al. (2005) had a rich pool of dependent variables for survivability prediction. They employed the Kaplan-Meier method of survival analysis with the Mantel-Haenszel log-rank test, which are fundamental statistical survival analysis techniques. These studies, however, have two major limitations. First, they lack an enhanced data-mining perspective which would utilize machine learning and artificial intelligence tools (which are independent of the nonlinearity and multicollinearity assumptions of traditional linear modeling techniques) to reveal previously unknown but potentially useful patterns. Second, the variables selection was based on the experiences and intuitions of the analysts who conducted the study.

Another study carried out by Tjang, Van der Heijden, Tenderich, Grobbee, and Korfer (2008) has the same drawbacks: based on their experience, they adopted some newer explanatory variables such as body mass index, waiting time on the list, and previous cardiac surgery to determine the survivability in heart transplantation. However, similarly to the aforementioned studies, they also utilized only statistical techniques such as the chi-square test, the Fisher's test, the nonparametric Kruskal-Wallis rank test, and the Kaplan-Meier survivorship

function. Similar limitations also exist in some other studies related to lung transplantation (Agüero et al., 2007; Cope et al., 2001; Lin et al., 1998), which therefore cannot be considered to be detailed data-mining studies.

A PI provides compact prognosis information regarding a specific patient based on the results of a Cox proportional hazards model (Parmar & Machin, 1995). The Cox proportional hazards model helps identify variables of prognostic importance; hence, the PI can be used to define groups of individuals at different risk categories. Some existing studies related to devising a PI in the transplant domain are summarized as follows.

In the study conducted by Christensen, Gunson, and Neuberger (1999), it is mentioned that primary biliary cirrhosis requires a liver transplantation operation at the end stage. However, a very critical issue is the timing for transplantation, which must be carried out neither too early nor too late. A prognosis analysis with and without transplantation would make it easier to decide whether or not the transplantation is required, and if it is necessary, when would be the most suitable time. To achieve this goal, corresponding PI's and thence probabilities of surviving are computed for transplantation and nontransplantation cases. Using these, a Cox regression model was created for 6-month survival, confirming some variables previously listed, but their model brings significant new variables into play. As a result, they learned that the gain from transplantation starts to become positive around 8 months prior to death, when the PI = 2.5.

The gain of transplantation is defined as the difference between survival probability with transplantation and without transplantation. If it gives a negative value, transplantation should not be performed and vice versa. The predicted gain from transplantation starts to become clinically important when the PI reaches about 2.5, corresponding to a predicted 6-month survival of about 0.85. The consequence of this is the following: If PI $\geq$ 2.5, transplantation should be performed within the following 6 months. Yoo, Galabova, Edwin, and Thuluvath (2002) developed a similar index and sought to answer whether or not socioeconomic status affects the survivability in liver transplantation, for both patients and grafts. The study revealed that socioeconomic status does not influence patient or graft survival of liver transplantation at their institute.

Deng, De Meester, Smits, Heinecke, and Scheld (2000) conducted a study with a national dataset in Germany; their objective was to discover the effect of receiving a heart transplant on patients in a waiting list. The results indicated that cardiac transplant is associated with a survival benefit only for patients with a predicted high risk of dying on the waiting list. Ghobrial et al. (2005) performed a study to determine the prognostic factors for overall survival in 107 adult patients with posttransplantation lymphoproliferative disorders (PTLDs). They found that in differentiating the patients with low and high scores, the proposed prognostic scoring significantly performs better than the international PI for the subset of the patients (56 out of 107) with lactate dehydrogenase.

The common limitation in all of these studies is similar to the limitations of the studies summarized when reviewing data mining models: researchers directly devise a PI without determining whether the variables used in the PI-devising phase are necessary and sufficient. This limitation motivates a machine learning-based initial step in the variable selection procedure because, if the critical predictive factors are not captured effectively due to the intuition- and experience-based selection, the resulting PI's developed based on the selected variables would be inaccurate and, in turn, related risk groups of patients would be deviated from the real classes. This consequence might lead decision makers to make errors in devising organ transplantation policies.

Proposed Methodology

As shown in the previous section, existing studies of organ transplantation procedures have employed traditional statistical approaches along with intuitively selected variables in order to predict survivability. However, a large number of variables are collected and stored in organ transplantation procedures. Disregarding some of the important variables might potentially cause the inaccurate classification of patient risk groups after the transplant has taken place. Therefore, in this study, machine learning techniques as well as statistical methods are used to determine the most critical factors affecting the survivability of thoracic transplant patients. After data understanding and preparation, artificial neural networks (ANNs) and support vector machines (SVMs) are used to develop prediction models for the graft survivability, which is a binary categorical variable.

These data mining-based prediction models also help extract and rank order the most important variables. As the next step, consolidated candidate sets of critical predictor variables are determined via the prediction models, based on the literature and the domain expertise. Then, a Cox survival analysis is conducted using the abovementioned consolidated sets of critical predictor variables, and PIs are also devised. Within the last

step, organ recipients are categorized into risk groups by a clustering algorithm, and the validation of this categorization is realized using the Kaplan-Meier survival curves. The specifics of these steps will be further detailed in subsequent sections.

UNOS Thoracic Dataset

The data source that was used to validate the methodology was provided by the UNOS. These data files are named UNOS Standard Transplant Analysis and Research (STAR) files for thoracic transplants. Each STAR file consists of information on all thoracic transplants that have been performed in the United States and reported to the Organ Procurement and Transplantation Network (OPTN) since October 1, 1987. It includes both deceased- and living-donor transplants. None of the files include any specific patient or transplant hospital identifiers due to the privacy and security issues. However, there is a patient identification number, unique to each patient, which allows patient tracking. Considering these features, UNOS is perceived to be the most comprehensive dataset available in any single field of medicine and for organ transplantation in the United States (Cupples & Ohler, 2002).

The full dataset consists of 443 variables and 61391 records. These variables include the socio-demographic and health-related factors regarding both the donors and the recipients and procedure-related factors. To assign as an output (dependent variable), there are four possible variables which are called *pstatus, ptime, gstatus,* and *gtime*. These variables have the following meanings: *pstatus* refers to whether the patient died after transplantation occurred (dead = 1 and alive = 0), and *gstatus* refers to whether the graft has failed (1 = failed and 0 = succeeded). The variable *ptime* denotes patient follow-up time (in days) from transplant to death/last follow up time, and *gtime* is the graft lifespan from transplant to death/last follow up time. For most of the records, *gtime* and *ptime* had the same value, and so did *gstatus* and *pstatus*.

Since the goal of this study was to develop models to predict the performance of thoracic transplant (i.e., survivability), the dependent variable was assigned as *gstatus,* to distinguish the patients who died solely due to the thoracic graft incompatibility from the ones who died from any other reason. Therefore, the other potential dependent variables (*pstatus, ptime,* and *gtime*) were eliminated from the dataset to avoid biasing the results considering the fact that they would obviously have a very strong association with *gstatus*, which might cause overwhelming the rest of the variables' impacts.

As *gstatus* was the categorical dependent variable, the records of the patients whose corresponding values for *gstatus* were not entered were removed from the dataset. There were also some identification variables (e.g., *Donor ID*) which only help tracking the patient and the recipient anonymously and the thoracic transplant procedure but do not have any effect on the prediction capability of the models. These types of variables were also excluded from the analysis. Moreover, the name of the dataset was recorded as *Dataset,* which had one value (*TH* referring to thoracic) and *Date of Run* which also had one value for all records. These were for UNOS internal use only for record-matching purposes and were not recorded for survivability prediction. Hence, another type of problematic variables was those having only one possible value for each record, and those were also eliminated since they would have no effect on the predictive modeling.

This dataset had excessive missing values, which render most of the records and variables seemingly insignificant. Following the regular convention for the row deletion, our rule of thumb was constructed based on the Pareto rule (Pareto, 1971): 80% of the effects come from 20% of the causes. Therefore, if the row (record) had missing values for more than 80% of the independent variables, it was deleted. However, in data mining studies, one should be very reluctant to remove the candidate predictor variables while trying to avoid an artificial data imputation procedure. There is an obvious trade-off in this case: as a rule of thumb for column (variable) deletion, we were cautious to remove any variable from the analysis and assumed that if a variable had more than 95% missing values, it did not carry much significant information, and it was thus deleted.

The next step was handling the missing values: the categorical variables were filled with some heuristic values such as *E* (referring to empty) and *NR* (referring to not reported). The continuous variables were imputed with the average of the existing records. After adopting these data preparation strategies, the final dataset consisted of 372 initial sets of cleansed variables and one dependent variable (*gstatus*) with 36438 records. As for the test design, the dataset was randomly partitioned into two subsets, namely training and testing datasets; this splitting is further explained in the *k*-Fold Cross-validation subsection.

Data Mining Models

In this study, two popular classification models from the machine learning domain were adopted, namely neural networks (NNs) and SVMs. The preliminary studies conducted to determine which models perform better in terms of classification accuracy had indicated these two models. The following sections provide a small description of these classification models.

Neural networks:

NNs have been utilized to model complex relationships among the predictor variables and the dependent variable such as nonlinear functions and multicollinearity (Mitchell, 1997). Formally defined, NNs are highly sophisticated analytic techniques capable of predicting new observations (on specific variables) from other observations (on the same or other variables) after executing a process of so-called "learning" from existing data (Haykin, 1998).

Support vector machines:

SVMs are supervised learning methods that generate input-output mapping functions from a set of labeled training data. They belong to a family of generalized linear models which achieve a classification or regression decision based on the value of the linear combination of features. They are also said to belong to the kernel methods (Cristianini & Shawe-Taylor, 2000). The mapping function in SVMs can be either a classification function (used to categorize the data) or a regression function (used to estimate the numerical value of the desired output).

Nonlinear kernel functions are often used to transport the input data (inherently representing highly complex nonlinear relationships) to a high dimensional feature space in which the input data become more separable (i.e., linearly separable) compared to the original input space. Then, maximum-margin hyperplanes are constructed to optimally separate the classes in the training data. Two parallel hyperplanes are constructed on each side of the hyperplane that separates the data by maximizing the distance between the two parallel hyperplanes. An assumption is made that the larger the margin or distance between these parallel hyperplanes, the better the generalization error of the prediction would be.

k-Fold Cross-validation

In data mining, in order to minimize the bias associated with the random sampling of the training and holdout (testing) data samples in comparing the predictive accuracy of two or more methods, researchers tend to use *k*-fold cross-validation (Kohavi, 1995). In *k*-fold cross-validation, also called rotation estimation, the complete dataset (*D*) is randomly split into *k* mutually exclusive subsets (the folds: $D_1, D_2, ..., D_k$) of approximately equal size. The classification model is trained and tested *k* times. Each time ($t \in \{1, 2, ..., k\}$), it is trained on all but one fold (D_t) and tested on the remaining single fold (D_t).

The cross-validation estimate of the overall performance criteria is calculated as simply the average of the *k* individual performance measures as follows,

$$CV = \frac{1}{k}\sum_{i=1}^{k} PM_i \tag{1}$$

where *CV* stands for the cross-validation, *k* is the number of folds used, and *PM* is the performance measure for each fold (Olson & Delen, 2008). In this study, a stratified 10-fold cross-validation approach was used to estimate the performance of classifiers. Empirical studies have shown that 10 is the optimal number of folds that optimizes the time it takes to complete the test while minimizing the bias and variance associated with the validation process (Kohavi, 1995). In 10-fold cross-validation, the entire dataset is divided into 10 mutually exclusive subsets (or folds) with approximately the same class distribution as the original dataset (stratified). Each fold is used once to test the performance of the classifier that is generated from the combined data of the remaining nine folds, leading to 10 independent performance estimates.

Performance Measures of Data Mining Models via a Confusion Matrix

A confusion matrix (as shown in Figure 1) is a matrix representation of the classification results. In a two-class classification problem (as in our case), the upper left cell denotes the number of samples classified as true while they were true in the actual classification (also called true positives–TP), and the lower right cell denotes the number of samples classified as false while they were actually false (also called true negatives–TN). The upper right cell represents the number of samples classified as false while they were actually true (also called false negatives–FN), and the lower left cell represents the number of samples classified as true while they were actually false (also called false positives–FP).

		Model Classification	
		Positive	Negative
Actual Classification	Positive	TP	FN
	Negative	FP	TN

Figure 1. A confusion matrix representation for a two-class classification problem.

To compare the classification models, four performance criteria were adopted in this study; they are given by the following equations:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{3}$$

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

where *TP, TN, FP, FN* denote true positive, true negative, false positive, and false negative, respectively. Accuracy, shown by Equation 2, measures the proportion of correctly classified test examples, therefore predicting the overall probability of the correct classification. Sensitivity and specificity, shown by Equations 3 and 4 respectively, measure the model's ability to recognize the patients of a certain group. For example, if grafts are concerned, sensitivity is a probability that a graft which has failed in reality is also classified as failed, and specificity is a probability that a succeeding graft is classified as succeeding (Demsar et al., 2001). Precision, shown by Equation 5, is the proportion of the observations assigned to the class by the model that are really class members (Lewis, 1995).

Information Fusion-based Sensitivity Analysis

In prediction modeling, there is no universally accepted "best model" that works for any problem. The best model depends on the scenario being analyzed and the data set being used and can only be obtained through trial-and-error experimentation (Ruiz & Nieto, 2000). Just as there is no single best model, there is also no single best implementation of different model types. Researchers are developing new ways to improve the accuracy and efficiency of prediction models. Therefore, it would be desirable to combine the results developed

by multiple models (Batchelor & Dua, 1995). Use of multiple models should make the forecasts more accurate and more efficient.

Information fusion is the process of intelligently combining the information created and provided by two or more information sources (i.e., models). While there is an ongoing debate about the sophistication level of the fusion methods to be employed, there is a general consensus that fusion (combining forecasts and/or predictions) produces more useful information (Armstrong, 2001). Combining forecasts can improve accuracy, completeness, and robustness of information, while reducing the uncertainty and bias associated with individual models (Chase, 2000). This multi-model fusion algorithm can be mathematically illustrated as follows (Delen, Sharda, & Kumar, 2007).

Given the expected response variable (*y*) and the decision variables $(x_1, x_2, ..., x_n)$ the formulation for any prediction model can be written as Equation 6:

$$\hat{y} = f(x_1, x_2, ..., x_n) \tag{6}$$

The prediction model *f* can take many forms. For instance, a linear regression model can be written as Equation 7:

$$f(x_1, x_2, ..., x_n) = \beta + \sum_{i=1}^{n} a_i x_i \tag{7}$$

where β is the intercept, and $a_i's$ are the coefficients for $x_i's$. For a NN model, for a single neuron, it may be written as Equation 8:

$$f(x_1, x_2, ..., x_n) = \phi(w_0 + \sum_{j=1}^{n} w_j x_j) \tag{8}$$

where ϕ is the transfer function, and $w_i's$ are the weights for $x_i's$. Given that we use *m* number of prediction models, the fusion model can be written as

$$\hat{y}_{fused} = \psi(\hat{y}_{individual,i}) = \psi(f_1(x), f_2(x), ..., f_m(x)) \tag{9}$$

If ψ is a linear function, which is the case in this study, then we can write Equation 10 as

$$\hat{y}_{fused} = \sum_{i=1}^{m} \omega_i f_i(x) = \omega_1 f_1(x) + \omega_2 f_2(x) + ... + \omega_m f_m(x) \tag{10}$$

where

$$\sum_{i=1}^{n} \omega_i = 1$$

The values for $\omega's$ are derived from the up-to-now prediction accuracy measure of the individual predictors. That is, the higher the accuracy of a predictor on independent test cases, the larger the weight that is assigned to that predictor type (Delen et al., 2007).

In addition, the rank order for the importance of the independent variables needs to be determined. In ANNs, the sensitivity analysis is a method for extracting the cause and effect relationship between the inputs and outputs of a trained ANN model (Davis, 1989). In the process of performing a sensitivity analysis after the model is trained, the ANN learning is disabled so that the network weights are not affected. The fundamental idea is that the sensitivity analysis measures the predictor variables based on the change in modeling performance that occurs if a predictor variable is not included in the model.

Hence, the measure of sensitivity of a specific predictor variable is the ratio of the error of the NN model without the predictor variable to the error of the model that includes this predictor variable (Principe, Euliano, & Lefebvre, 1999). The more sensitive the network is to a particular variable, the greater the performance decrease would be in the absence of that variable, and therefore the greater the ratio of importance. The same method is followed in SVMs to rank the variables in terms of their importance according to the sensitivity measure defined as in Equation 11 (Saltelli, 2002).

$$S_i = \frac{V_i}{V(F_t)} = \frac{V(E(F_t \mid X_i))}{V(F_t)} \tag{11}$$

where $V(F_t)$ is the unconditional output variance. In the numerator, the expectation operator E calls for an integral over X_{-i} that is, over all input variables but X_i, and the variance operator V implies a further integral over X_i. Variable importance is then computed as the normalized sensitivity. Saltelli, Tarantola, Campolongo, and Ratto (2004) showed that Equation 4 is the proper measure of sensitivity to rank the predictors in order of importance for any combination of interaction and nonorthogonality among predictors.

Considering Equations 10 and 11 simultaneously, the sensitivity measure of the variable *n* with information fused by *m* prediction models can then be given by Equation 12

$$S_{n(fused)} = \sum_{i=1}^{m} \omega_i S_{in} = \omega_1 S_{1n} + \omega_2 S_{2n} + \ldots + \omega_m S_{mn} \tag{12}$$

where $\omega's$ refer to the normalized accuracy value of each prediction model with *m* models in total, and S_{in} is the sensitivity measure of the n^{th} variable in the i^{th} model.

Effective Predictor Variables in Thoracic Transplant Performance

The insignificant variables need to be eliminated, which, by optimizing the predictor variables list, would improve the accuracy of the model. The potential input variables to this step consist of three candidate sets of predictor variables. The first set is composed of predictive model-selected variables. The predictive models explained in the Data Mining Models subsection can rank the predictor variables based on their importance level with regard to performance measurement via the *gstatus*. However, powerful machine learning tools such as NNs hardly eliminate the input variables. Instead, they tend to use all of them and try to capture the relations amongst all. Therefore, a subsequent step is to employ a Pareto analysis-like procedure to select the "vital few" variables from the "trivial many" (Pareto, 1971). By applying a pseudo-Pareto analysis to each predictive model separately, a union set of predictive variables would be constructed, named as the first set of predictive variables.

The second set of predictive variables is obtained by considering the common-sense domain knowledge. This set includes variables which are logically related to thoracic transplantation such as donors' history of smoking. The third set of predictive variables is compiled from the related literature. This set essentially consists of the variables which have been commonly and repeatedly used in previous studies in the organ transplantation area. The second and third sets of predictive variables can be referred as the expert input to the variable determination stage of the proposed method. These sets (Set 2 and Set 3) provide one more chance to the next step–the Cox regression model–to evaluate the variables that might have importance in the survival analysis although they were determined to be insignificant by the predictive models.

Cox Survival Analysis

All the aforementioned three sets of predictive variables are fed into the Cox survival model to model the graft survivability while filtering out the candidate predictive variables which do not have a significant effect on the performance measurement of the transplant. Hence, the final critical predictive variables are determined by the Cox survival analysis. The Cox model also enables us to devise a PI to categorize the patients into differing risk groups such as low, medium, and high.

The Cox regression model is a semi-parametric model which is extensively used in survival analysis (Cox

& Oakes, 1984; Ohno-Machado, 2001). The survival time of each patient is assumed to follow the hazard function (h_i) given by Equation 13 as follows:

$$h_i = h_0 \exp(x_i \beta) \tag{13}$$

where h_0 is the baseline hazard function, and x_i is the vector of predictor variables for the i^{th} patient. β is the vector of regression coefficients for the predictor variables and is assumed to be the same for all patients (Grambsch & Therneau, 1994).

One important application of the Cox regression model is to help identify variables which may be of prognostic importance (Parmar & Machin, 1995). Once the variables are identified, knowledge from these variables may be combined and used to define a PI, which in turn defines groups of organ recipients at differing risk. To use the PI, key patient characteristics are recorded, and a score is derived from these. This score gives an indication of whether, for example, a particular patient has a good, intermediate, or bad prognosis for the disease (Parmar & Machin, 1995). Remembering Equation 13, the PI for each patient can be calculated by Equation 14:

$$PI = x_1\beta_1 + x_2\beta_2 + ... + x_n\beta_n \tag{14}$$

where x_1 to x_n are the patient's values for the variables in the Cox model, and β_1 to β_n are the corresponding regression coefficients determined by the Cox regression model (Christensen, 1987).

It should be noted that the PI in Equation 14 represents the exponent portion in Equation 13. Therefore, the smaller the PI, the smaller the hazard function value, and hence the smaller the risk associated with a particular recipient.

Categorization of Thoracic Recipients

A *k*-means clustering algorithm is used to provide an answer to the question of how many risk groups should patients be classified into. As a statistical and/or pictorial verification mechanism for the number of groups determined by the *k*-means clustering algorithm, the Kaplan-Meier survival analysis (Kaplan & Meier, 1958) is adopted and corresponding survival curves are generated.

The *k*-means method is an extensively used and arguably the most popular clustering algorithm that searches for a nearly optimal partition with a fixed number of clusters represented by the parameter *k* (MacQueen, 1967). It proceeds by assigning *k* initial centroids to the multi-dimensional datasets. Each record in the dataset is allocated to the centroid which is nearest and hence forms a cluster. Each cluster centroid is then updated to be the center of its members, followed by a new assignment of records to the nearest centroids to reconstruct the clusters. The algorithm converges when there is no further change in the allocation of members to clusters, or some predefined time-based stopping criteria is satisfied (Krishna & Murty, 1999).

The Kaplan-Meier analysis is a nonparametric technique which is used to test the statistical significance of differences between the survival curves associated with two different circumstances (Kaplan & Meier, 1958). The analysis expresses the distribution of patient survival times in terms of the proportion of patients still alive up to a given time. The Kaplan-Meier survival curves plot the proportion of patients surviving against time which has a characteristic decline. In biostatistics, a typical application of the Kaplan-Meier survival curves involves grouping patients into risk categories such as low risk, medium risk, and high risk.

Results and Discussion

Following the methodology proposed in the previous section, the preliminary analysis showed that NNs and SVMs gave satisfactorily high prediction accuracy results in terms of performance measures. Therefore, these two machine learning models were kept as a modeling technique, and other statistical binary classifier models (such as logistic regression, discriminant analysis, and decision trees) were eliminated since their accuracy rates were not observed to be satisfactory in our preliminary trials. The cutoff value for success was to adopt a general rule of thumb (Hair, Anderson, Tatham, & Black, 1998) which claimed that the model should be able to predict the classes 25% better than random chance. In our study, with 38% and 62% of each class of dependent variables, a "good enough" model should exceed the random chance of 47% and 77%, respectively.

Hence, NNs and SVMs were kept to sort out the first set of candidate predictor factors, as further explained in the next subsection. These two models were employed for classification on the dependent variable *gstatus*. Table 1 shows the confusion matrices for both models. Based on the confusion matrix, the accuracy, sensitivity, specificity, and precision of each fold were calculated via the methods presented in the *k*-Fold Cross-validation subsection and in the Performance Measures of Data Mining Models via Confusion Matrix subsection. It was observed in our preliminary runs that a multilayer Perceptron (MLP) architecture with a single hidden layer performs better than other NN architectures such as radial basis function (RBF) and dynamic networks; hence, it was implemented in this study.

As for the SVM, based on the favorable prediction results obtained from the preliminary runs in this study, we chose to use the RBF-based kernel method. Table 1 shows that the RBF SVM model outperformed NNs with an 89% overall accuracy, 85% sensitivity, 93% specificity, and 88% precision. These results are better than those of any other study reported in the existing literature reviewed at the beginning of this article. For example, the recent studies conducted by Kusiak et al. (2005) and Lin et al. (2008) could achieve an average accuracy rate of 75% at most whereas our study could reach 89% accuracy. Moreover, none of the reported studies handled thoracic transplant procedure with such a voluminous dataset and in a data-mining fashion.

Table 1
Comparison of the Two Predictive Models

Fold No.	NN (MLP)						SVM (RBF)					
	Confusion Matrix		*Accuracy*	*Sensitivity*	*Specificity*	*Precision*	Confusion Matrix		*Accuracy*	*Sensitivity*	*Specificity*	*Precision*
1	895	248	0.7956	0.6430	0.8899	0.7830	1 216	194	0.8985	0.8736	0.9139	0.8624
	497	2 004					176	2 058				
2	925	245	0.8046	0.6645	0.8912	0.7906	1 175	119	0.9078	0.8441	0.9472	0.9080
	467	2 007					217	2 133				
3	919	246	0.8027	0.6602	0.8908	0.7888	1 169	125	0.9045	0.8398	0.9445	0.9034
	473	2 006					223	2 127				
4	898	237	0.7994	0.6451	0.8948	0.7912	1 243	222	0.8982	0.8930	0.9014	0.8485
	494	2 015					149	2 030				
5	847	229	0.7876	0.6085	0.8983	0.7872	1 130	119	0.8954	0.8118	0.9472	0.9047
	545	2 023					262	2 133				
6	829	179	0.7966	0.5960	0.9205	0.8224	1 229	221	0.8949	0.8835	0.9019	0.8476
	562	2 073					162	2 031				
7	904	218	0.8063	0.6494	0.9032	0.8057	1 257	191	0.9105	0.9030	0.9152	0.8681
	488	2 034					135	2 061				
8	880	243	0.7930	0.6326	0.8921	0.7836	1 098	107	0.8902	0.7894	0.9525	0.9112
	511	2 009					293	2 145				
9	902	246	0.7980	0.6480	0.8908	0.7857	1 219	195	0.8990	0.8757	0.9134	0.8621
	490	2 006					173	2 057				
10	851	238	0.7862	0.6114	0.8943	0.7815	1 138	109	0.9004	0.8175	0.9516	0.9126
	541	2 014					254	2 143				
Mean			**0.7970**	0.6359	0.8966	0.7920			**0.8999**	0.8531	0.9289	0.8829
St. Dev.			0.0067	0.0232	0.0094	0.0127			0.0062	0.0383	0.0214	0.0273

Note. Standard Deviation (St. Dev.).

Relative Importance of Predictor Variables and Pseudo-Pareto Analyses

The information fusion-based sensitivity analysis procedure described previously was employed on 10 folds of testing data to determine the importance of the predictor variables effective on the performance measurement of thoracic transplants. The next step in the method proposes to apply pseudo-Pareto analysis to the prediction-models-extracted variables to focus on only the vital few rather than the trivial many. The Pareto rule basically claims that 80% of the problems stem from 20% of the causes. Therefore, instead of dealing with all root causes on hand, it suggests handling 20% of them that would hypothetically help solve 80% of the problems. The relative importance of the predictor variables were extracted by a pseudo-Pareto analysis but with a small difference: instead of applying the 80/20 rule, we adopted a threshold where the Pareto charts tail off. In other words, the cutoff points were where the addition of one more variable does not bring as much significance to the model, and the trend of the line becomes parallel to the x-axis. The corresponding Pareto charts are illustrated in Figures 2 and 3, in which the cutoff points are shown with the dashed lines.

These cutoffs correspond to 4 variables for SVM (out of 21) and 32 variables for MLP (out of 292) with the cumulative percentages of 91% and 87% of the total importance value, respectively. Therefore, our Pareto rules can be summarized as 19% (4/21) to 91% for SVM and 11% (32/292) to 87% for MLP, as opposed to the fixed 80/20 ratio.

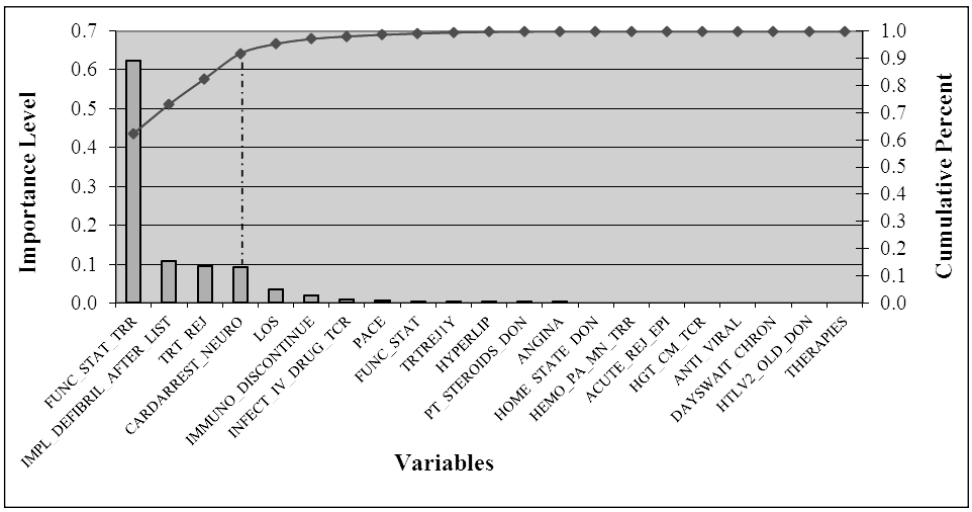

Figure 2. Pareto chart for variable importance ranking in SVM.

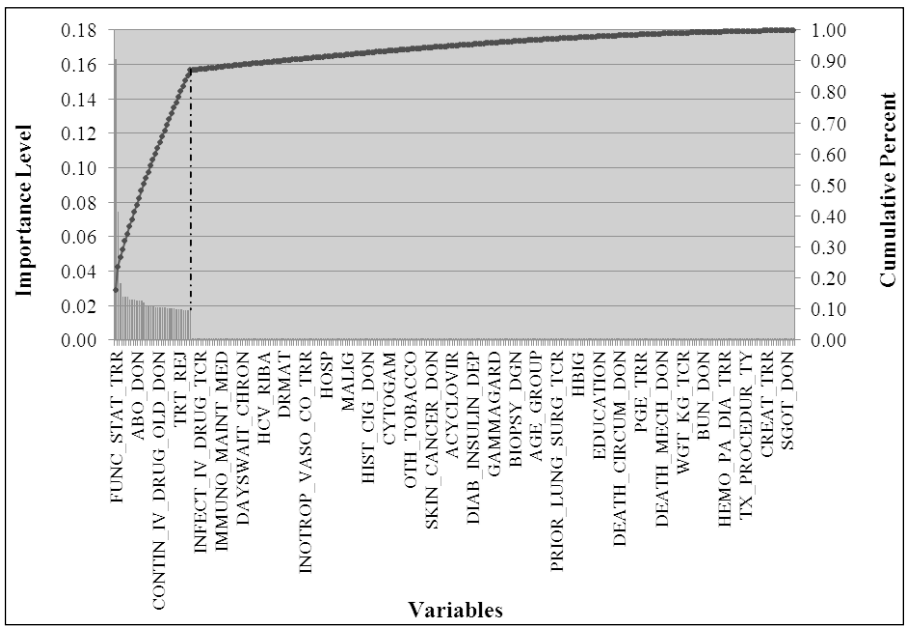

Figure 3. Pareto chart for sensitivity analysis of variables in MLP.

A clear quantitative plan to help the medical experts use the methodology proposed in this study could be summarized and explained as follows:

1. Develop a Pareto chart based on the model used.
2. Via eyeballing, determine a cut-off value regarding which variables to include in subsequent steps.
3. Based on the cutoff determined above at Step 2, fit two separate linear regression models using the data that lie before and after the cutoff values.
4. If the linear trend (the significance of the linear trend line slope) proves to be significant, use the cutoff determined at Step 2. If not, switch to and try another cutoff point on the right or left of the cutoff tried at Step 2.

In this study, for example, if we fit two linear regression models for Figure 3, we obtain the following results, shown in Table 2:

Table 2
Linear Regression Results with regard to Figure 3

For Figure 3	*Coefficients*	*Significance test*
slope for the 1st part	0.02049	2.41386E-33
slope for the 2nd part	0.00005	3.11

This table shows that where there is a clear significant linear trend in the left side of the line drawn in Figure 3, there is no significant slope on the right side of it. Therefore, we can include variables determined with regard to the dashed line drawn on Figure 3. Similar results are received when Figure 2 is analyzed, as seen in Table 3.

Table 3
Linear Regression Results with regard to Figure 2

For Figure 2	*Coefficients*	*Significance test*
slope for the 1st part	0.09828	0.00062
slope for the 2nd part	0.00191	0.47401

Selecting Predictor Variables for the Cox Survival Analysis

The proposed method involves providing three different sets of candidate covariates to be used in the Cox model. The first set is the union set of the predictor variables that were extracted by the pseudo-Pareto analyses detailed in the previous subsection. The list of these variables is presented in Table 4.

Table 4
First Set of Predictor Variables Selected by Data Mining Models

Variables from NN model	Explanation
Func_stat_trr	Recipient functional status at transplant
Func_stat_tcr	Recipient functional status at registration
Therapies	Other therapies
Cardarrest_neuro	Deceased donor –cardiac arrest post brain death
HTLV2_old_don	Deceased donor –antibody to HTLV II result
Physical_capacity	Physical capacity
Secondary_pay_tcr	Recipient secondary projected pay source at listing
HBV_core_don	Donor HBV core antibody
Secondary_pay_trr	Recipient secondary projected pay source at transplant
ABO_don	Donor blood type
Trtrej1y	Treated for rejection within 1 year
Anti_viral	Antiviral therapy
Life_sup_tcr	Recipient life support at registration
Impl_defibril_after_list	Implantable defibrillator inserted between listing and transplant
Pretreat_med_don_old	Deceased donor prerecovery medication(s) from brain death to 24 hours prior
Hlamat	HLA match level
Vad_tah_tcr	Recipient on life support –ventilator at registration (1 = yes, 0 = no)
Hlamis	HLA mismatch level
Contin_IV_drug_old_don	Deceased donor –history of iv drug use + recent 6 months of use
Pst_pacemaker	Events prior to discharge: permanent pacemaker
Diabetes_don	Deceased donor –history of diabetes (yes, no)
Infect_IV_drug_tcr	Infection requiring iv drug therapy (within 2 weeks prior to listing)
VDRL_don	Deceased donor –RPR– VDRL result
O_2*_req*	Lung graft status –oxygen requirement (l/min) at rest
Pace	Heart graft status –permanent pacemaker inserted since last follow-up
ABO	Recipient blood group
Trt_rej	Recipient treated for rejection during follow-up period
Hep_C_anti_don	Deceased donor –antibody to hepatitis C virus result
Dantiarr_old	Deceased donor given antiarrythmics 24 hours prior to cross clamp
Pst_surgical	Events prior to discharge: other surgical procedures
Pt_T4_don	Deceased donor –thyroxin brain death within 24 hours of procurement
Contin_cig_don	Deceased donor –history of cigarettes in past and > 20 pack years + recent 6 months of use

First set of candidate covariates generated from the data mining models

Variables from SVM model	Explanation
Func_stat_trr	
Impl_defibril_after_list	As described above
Trt_rej	
Cardarrest_neuro	

Note. (a) Human T-cell Lymphotropic Viruses Type II (HTLV II), Hepatitis B Virus (HBV), Human Leukocyte Antigen (HLA), Rapid Plasma Reagent (RPR), Venereal Disease Research Laboratory (VDRL). (b) 20 pack years: whether the individual has consumed 20 packets of cigarettes within the past years.

The second set of predictor variables are the ones which are not shown in the literature although they might have potential importance in thoracic transplants. These variables were selected through brainstorming sessions with healthcare experts in the field. The second set of candidate covariates is shown in Table 5.

Table 5
Second Set of Candidate Covariates

Variables	Explanation
Antiarry	Heart medical factors: antiarrythmics at registration
Contin_Alcohol_Old_Don	Deceased donor history of alcohol dependency + recent 6 months of use
Contin_Cig_Don	Deceased donor history of cigarettes in past and > 20 pack years + recent 6 months of use
Contin_IV_Drug_Old_Don	Deceased donor history of iv drug use + recent 6 months of use
Contin_Oth_Drug_Don	Deceased donor history of other drugs in past + recent 6 months of use
Hist_Alcohol_Old_Don	Deceased donor history of alcohol dependency
Hist_Cancer_Don	Deceased donor history of cancer (yes/no)
Hist_Cig_Don	Deceased donor history of cigarette use in past and > 20 pack years
Hist_Cocaine_Don	Deceased donor history of cocaine use in past
Hist_Diabetes_Don	Deceased donor history of diabetes, including duration of disease
Hist_Hypertens_Don	Deceased donor history of hypertension
LOS	Recipient length of posttransplant stay
Oth_Tobacco	Other tobacco use
Pack_Yrs	If history of cigarette use, number of pack years

Note. 20 pack years: whether the individual has consumed 20 packets of cigarettes within the past years.

The third set of candidate covariates was determined through the literature research presented at the beginning of this paper. This set includes the variables which have been commonly used in the previously published studies related to organ transplantation. The third set of candidate covariates is shown in Table 6.

Table 6
Third Set of Candidate Covariates

Variables	Explanation
ABO	Recipient blood group at registration
ABO_Don	Donor blood type
ABO_Mat	Donor-recipient ABO match level
Age	Recipient age (years)
Age_Don	Donor age (years)
Dayswait_Chron	Active days on waiting list
Don_TY	Donor type –deceased/living
Ethcat	Recipient ethnicity category
Ethcat_Don	Donor ethnicity category
Gender	Recipient gender

Gender_Don	Donor gender
Hbsab_Don	Deceased donor HBsAb test result
Ischtime	Ischemic time in hours
Med_Cond_Tcr	Recipient medical condition at registration
Med_Cond_Trr	Recipient medical condition pretransplant at transplant
Wgt_kg_Don	Donor weight (kg)
Wgt_kg_Tcr	Recipient weight (kg) at registration

Note. Alliance of Blood Operators (ABO), Hepatitis B surface Antibody (HBsAb).

Cox Survival Analysis Implementation

All the candidate covariates determined were then assigned to the Cox regression model. The predictor variables determined to be significant by the Cox regression model are listed along with their corresponding statistics in Table 7.

Table 7
Variables Kept in the Cox Regression Model

Variable	SE	Wald Test	DF	Sig.	exp(*β*)	95% CI for exp(*β*)	
						Lower	Upper
Trt_rej	0.006	6 323.384	1	0.000	1.582	1.564	1.600
Physical_capacity	0.000	804.551	1	0.000	1.001	1.001	1.001
O2_req	0.005	1 091.459	1	0.000	1.171	1.160	1.182
Pace	0.020	41.254	1	0.000	0.878	0.844	0.914
Med_Cond_Trr	0.008	10.970	1	0.001	0.974	0.958	0.989
Anti_Viral	0.014	15.072	1	0.000	1.056	1.027	1.085
Therapies	0.011	72.310	1	0.000	1.099	1.076	1.124
Secondary_Pay_Trr	0.002	82.613	1	0.000	0.986	0.983	0.989
Gender	0.013	15.837	1	0.000	0.950	0.927	0.974
Med_Cond_Tcr	0.008	8.682	1	0.003	1.024	1.008	1.040
Oth_Tobacco	0.076	11.200	1	0.001	0.775	0.667	0.900
Hlamat	0.005	40.101	1	0.000	0.970	0.961	0.979
Trtrej1y	0.004	9.110	1	0.003	0.987	0.979	0.995
Dayswait_Chron	0.000	7.988	1	0.005	1.000	1.000	1.000
Hist_Cocaine_Don	0.023	22.095	1	0.000	1.115	1.066	1.167
Age_Don	0.000	177.951	1	0.000	1.006	1.005	1.007
HBV_Core_Don	0.008	7.918	1	0.005	0.978	0.964	0.993
ABO_Don	0.003	29.437	1	0.000	0.984	0.978	0.989
Hep_C_Anti_Don	0.014	195.853	1	0.000	1.225	1.190	1.260
HTLV2_Old_Don	0.006	51.764	1	0.000	0.955	0.943	0.967
Pretreat_Med_Old_Don	0.022	17.704	1	0.000	1.098	1.051	1.147
Contin_Oth_Drug_Don	0.009	19.658	1	0.000	0.962	0.945	0.978

Hist_Alcohol_Old_Don	0.014	12.818	1	0.000	0.952	0.926	0.978
Hist_Cancer_Don	0.034	20.320	1	0.000	1.164	1.090	1.244
Hist_Diabetes_Don	0.000	20.530	1	0.000	1.000	1.000	1.001
Diabetes_Don	0.037	83.167	1	0.000	1.405	1.306	1.511
Age	0.000	45.003	1	0.000	1.003	1.002	1.003
Ischtime	0.005	8.227	1	0.004	0.987	0.098	0.996
LOS	0.000	28.606	1	0.000	1.002	1.001	1.003
Infect_IV_Drug_Tcr	0.030	76.255	1	0.000	1.296	1.223	1.374
Impl_Defibril_After_List	0.010	28.953	1	0.000	1.054	1.034	1.075
Cardarrest_Neuro	0.029	33.927	1	0.000	1.187	1.120	1.257

Note. Standard Error (SE), Degrees of Freedom (DF), Significance (Sig.), Exponentiated Coefficient–exp(β), Confidence Interval (CI).

The other variables (which were in Tables 4, 5, or 6 but not in Table 7) were eliminated since they were found to be insignificant by the Cox regression model. As listed in Table 7, 32 of the variables had prognostic value, which means they are determined by the Cox model as significant and kept in the Cox equation. Therefore, they were used to calculate the PIs by means of Equation 14. The PI values received here were ranging between 0.001 and 4.45.

k-Means Algorithm Implementation

Once the PIs for each recipient were calculated, the next step was to cluster the recipients through these PIs. However, how to define these clusters, decide which value to cut off, and hence categorize the recipients was not known a priori. Therefore, we adopted a *k*-means clustering algorithm to determine these clusters. The algorithm was run by changing the value for *k* (the number of clusters to be formed). We tried two, three, four, and five clusters because we considered that having more than five clusters would not provide logical risk groups to categorize and would probably not be easy to interpret medically. The results for each run are represented in Table 8.

The performance of the *k*-means algorithm with a different number of clusters ($k = 2, 3, 4, 5$) was compared using *intraclass inertia*. It is a performance measure which shows how compact each cluster is. Intraclass inertia is the average of the distances between the means and the observations in each cluster. Equation 15 indicates this value for a given *k* number of clusters (Michaud, 1997).

$$F(k) = \frac{1}{n} \sum_{k} \sum_{i \in C_k} \sum_{P=1}^{m} (X_{iP} - \mu_{kP}) \tag{15}$$

where n is the number of total observations, C_k is the set of k^{th} cluster, X_{iP} is the value of the attribute P for observation i and μ_{kP} is the mean of the attribute P in the k^{th} cluster. In the present case, there is only one attribute which is PI, and hence $m = 1$.

Deployment of Kaplan-Meier Survival Curves

Kaplan-Meier survival curves were constructed for each possible number of cluster formations to validate the established PIs and hence the various risk groups. The main objective was to compare the survivor functions for different risk groups of thoracic recipients. If the survivor function for one risk group is always higher than the survivor function for another risk group, then the first group clearly lives longer than the first one. The less the survivor functions cross, the better the discrimination of the patients would be by the PIs derived. Figure 4 shows this clear distinction in all possible number of clusters.

In order to show statistically that there is a significant difference among the risk groups, the test of equality over strata was conducted for each of the survivor function in Figure 4. The test of equality over strata contains rank and likelihood-based statistics for testing the homogeneity of survivor functions across strata. The rank tests with the log-rank test and the Wilcoxon test indicate a significant difference between the risk groups. These results are also supported by likelihood-based statistics. These statistical test results and intra-class inertia values for each possible number of clusters are also summarized in Table 8. Table 8 shows that in the present case, PIs were clustered best with $k = 4$.

Table 8
Tests of Equality over Risk Groups for each Potential Cluster

No. of risk groups	Test	Chi-Square	DF	Pr > Chi-Square	Intraclass Inertia
with 2 groups	Log-Rank	4 711.9167	1	< 0.0001	0.0946
	Wilcoxon	4 623.9311	1	< 0.0001	
	-2Log(LR)	2 881.6489	1	< 0.0001	
with 3 groups	Log-Rank	5 735.5048	2	< 0.0001	0.1064
	Wilcoxon	5 733.897	2	< 0.0001	
	-2Log(LR)	3 197.8192	2	< 0.0001	
with 4 groups	Log-Rank	9 277.4524	3	< 0.0001	0.0618*
	Wilcoxon	8 969.3411	3	< 0.0001	
	-2Log(LR)	5 520.4709	3	< 0.0001	
with 5 groups	Log-Rank	10 740.8635	4	< 0.0001	2.2421
	Wilcoxon	10 312.6482	4	< 0.0001	
	-2Log(LR)	5 986.9124	4	< 0.0001	

Note. Degrees of Freedom (DF), Logarithmic Likelihood Ratio–Log(LR).

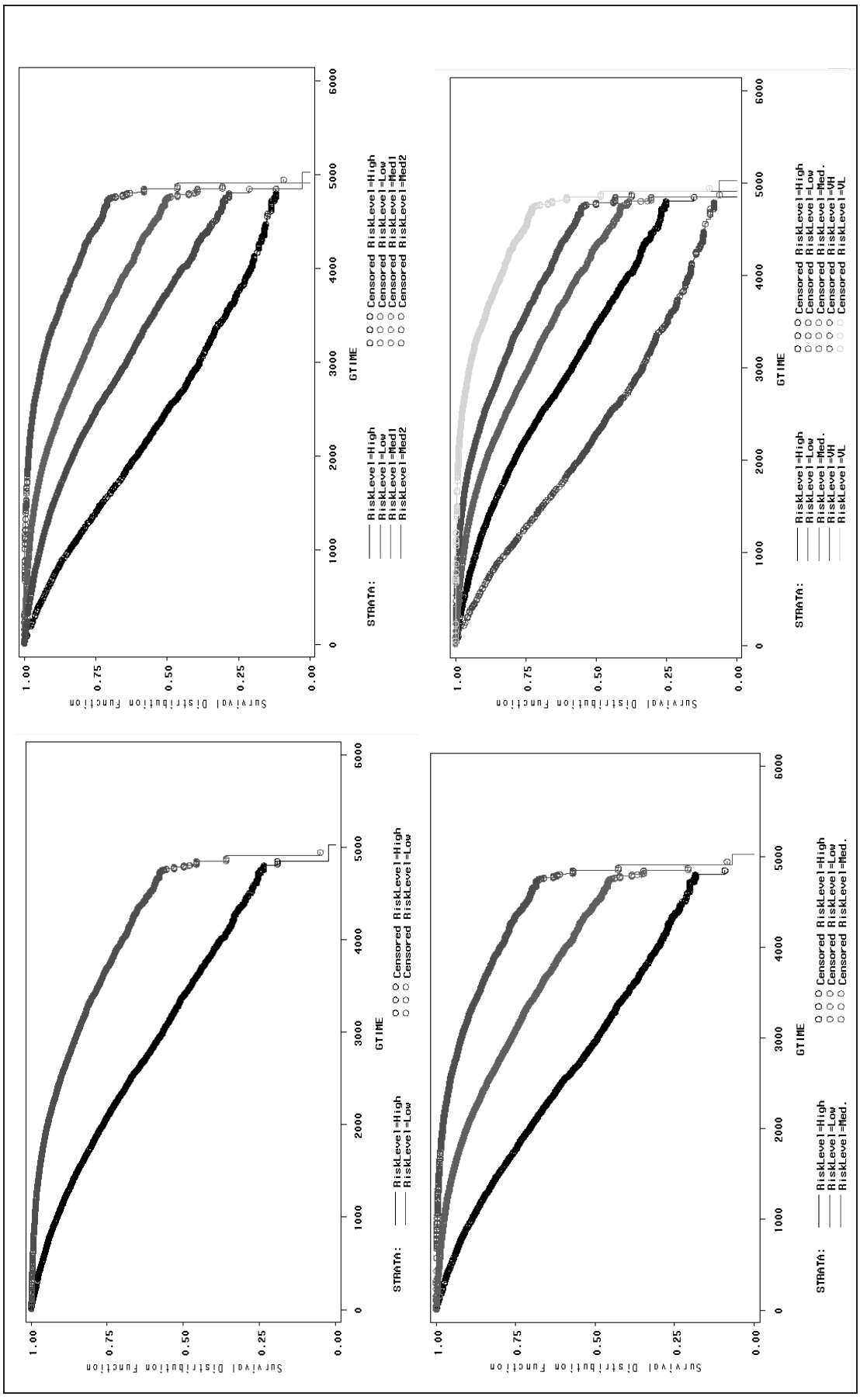

Figure 4. Kaplan-Meier survival curves for each potential number of risk groups.

Conclusions

This study proposed a data mining-based prediction methodology for thoracic transplants. It revealed that if the performance measure of the transplant is assigned to be *gstatus,* the thoracic organ recipients should be categorized into four risk groups, namely low, low-medium, medium-high, and high. This finding is different from the results of conventional discrimination commonly used in this field of study where there have been only three risk groups: low, medium, and high. This is the point where the medical professionals should be consulted to find out whether or not this categorization is feasible and might potentially bring any medical value to the thoracic transplant study area. An example of application could be the scheduling of follow-up strategies regarding the frequency of posttransplant doctor visits and checking for each risk group of patients.

The research presented also showed that in addition to the list of predictors used in the existing literature, arising out of the medical experts' intuitions and experiences, other variables (e.g., donor history in terms of alcohol dependency, cancer, and diabetes) should also be used in predictive modeling for an improved performance measurement. The determination of these new variables was performed by the 10-fold cross-validated information fusion-based sensitivity analyses followed by pseudo-Pareto analyses. The high prediction accuracy levels received via NNs and SVM models supported the validity of the methodology proposed in the study.

Although data mining methods are highly capable of extracting novel patterns and relationships hidden deep in organ transplantation large datasets, without the cooperation and feedback from the medical professionals, the findings would have little value. The patterns found via data mining methods should be evaluated by medical professionals who have years of experience in the problem domain to decide whether they are logical, actionable, and novel to foster new biological and clinical research directions. In short, data mining is not to replace the medical professionals and researchers, but to complement their invaluable research efforts to solve unsolvable medical problems and hence save more human lives.

The findings of data mining studies are only as good as the data provided to them. The quality and the quantity of the data determine the correctness and usefulness measures of the findings. The popular expression "garbage-in-garbage-out" applies to data mining more so than to any other field. One should also choose appropriate methods to analyze the data. As of now, data mining is far from being a standardized field of study where a recipe type procedure leads to optimal results. Rather, it is still an art driven by rule-of-thumb, experiences, and extensive experimentations. Different application domains with different types of data require researchers to devise a different, rather unique solution to the problem on hand. In that sense, this study has proposed one of many potential methodologies to effectively and efficiently deal with the prognostic analysis of thoracic organ transplant recipients. Others may use different data (or different representations of similar data) with a different set of methods to produce different results, as useful or even more useful than the ones reported in this study, which would contribute to the body of knowledge regarding data mining.

References

Abouna, G. M. (2003). Ethical issues in organ transplantation. *Medical Principle and Practice, 12,* 54-69. doi:10.1159/000068158

Agüero, J., Almenar, L., Martínez-Dolz, L., Moro, J., Izquierdo, M. T., Cano, O., & Salvador, A. (2007). Differences in clinical profile and survival after heart transplantation according to prior heart disease. *Transplantation Proceedings, 39*(7), 2350-2352. doi:10.1016/j.transproceed.2007.06.068

Armstrong, J. S. (2001). Combining forecasts. In J. S. Armstrong (Ed.), *Principles of forecasting*: *A handbook for researchers and practitioners* (pp. 417-439). Norwell, MA: Kluwer Academic.

Batchelor, R., & Dua, P. (1995). Forecaster diversity and the benefits of combining forecasts. *Management Science, 41*(1), 68-75. doi:10.1287/mnsc.41.1.68

Chase, C. W., Jr. (2000). Composite forecasting: Combining forecasts for improved accuracy. *Journal of Business Forecasting Methods & Systems, 19*(2), 2-6.

Christensen, E. (1987). Multivariate survival analysis using Cox's regression model. *Hepatology, 7*(6), 1346-1358. doi:10.1002/hep.1840070628

Christensen, E., Gunson, B., & Neuberger, J. (1999). Optimal timing of liver transplantation for patients with primary biliary cirrhosis: Use of prognostic modelling. *Journal of Hepatology, 30*(2), 285-292. doi:10.1016/S0168-8278(99)80075-0

Cope, J. T., Kaza, A. K., Reade, C. C., Shockey, K. S., Kern, J. A., Tribble, C. G., & Kron, I. L. (2001). A cost comparison of heart transplantation versus alternative operations for cardiomyopathy. *Annals of Thoracic Surgery*, *72*(4), 1298-1305. doi:10.1016/S0003-4975(01)02997-6

Cox, D. R., & Oakes, D. (1984). *Analysis of survival data*. Boca Raton, FL: Chapman & Hall/CRC.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods.* Cambridge, UK: Cambridge University Press.

Cupples, S. A., & Ohler, L. (2002). *Transplantation nursing secrets*. Philadelphia, PA: Hanley & Belfus.

Davis, G. W. (1989). Sensitivity analysis in neural net solutions. *IEEE Transactions on Systems, Man and Cybernetics, 19*(5), 1078-1082. doi:10.1109/21.44023

Delen, D., Sharda, R., & Kumar, P. (2007). Movie forecast guru: A web-based DSS for Hollywood managers. *Decision Support Systems, 43*(4), 1151-1170. doi:10.1016/j.dss.2005.07.005

Demsar, J., Zupan, B., Aoki, N., Wall, M. J., Granchi, T. H., & Beck, J. R. (2001). Feature mining and predictive model construction from severe trauma patient's data. *International Journal of Medical Informatics, 63*(1-2), 41-50. doi:10.1016/S1386-5056(01)00170-8

Deng, M. C., De Meester, J. M. J., Smits, J. M. A., Heinecke, J., & Scheld, H. H. (2000). Effect of receiving a heart transplant: Analysis of a national cohort entered on to a waiting list, stratified by heart failure severity. *British Medical Journal, 321*(7260), 540-545. doi:10.1136/bmj.321.7260.540

Fernández-Yánez, J., Palomo, J., Torrecilla, E. G., Pascual, D., Garrido, G., De Diego, J. J. G., Domínguez, M., & Almendral, J. (2005). Prognosis of heart transplant candidates stabilized on medical therapy. *Revista Española de Cardiología, 58*(10), 1162-1170. doi:10.1016/S1885-5857(06)60395-2

Ghobrial, I. M., Habermann, T. M., Maurer, M. J., Geyer, S. M., Ristow, K. M., Larson, T. S., Walker, R. C., Ansell, S. M., Macon, W. R., Gores, G. G., Stegall, M. D., & McGregor, C. G. (2005). Prognostic analysis for survival in adult solid organ transplant recipients with post-transplantation Lymphoproliferative disorders. *Journal of Clinical Oncology, 23*(30), 7574-7582. doi:10.1200/JCO.2005.01.0934

Grambsch, P. M., & Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika, 81*(3), 515-526. doi:10.1093/biomet/81.3.515

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. (1998). *Multivariate data analysis* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Haykin, S. (1998). *Neural networks*: *A comprehensive foundation* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Hong, Z., Wu, J., Smart, G., Kaita, K., Wen, S. W., Paton, S., & Dawood, M. (2006). Survival analysis of liver transplant patients in Canada 1997-2002. *Transplantation Proceedings*, *38*(9), 2951-2956. doi:10.1016/j.transproceed.2006.08.180

Jenkins, P. C., Flanagan, M. F., Jenkins, K. J., Sargent, J. D., Canter, C. E., Chinnock, R. E., Vincent, R. N., Tosteson, A. N., & O'Connor, G. T. (2000). Survival analysis and risk factors for mortality in transplantation and staged surgery for hypoplastic left heart syndrome. *Journal of the American College of Cardiology, 36*(4), 1178-1185. doi:10.1016/S0735-1097(00)00855-X

Kaplan, B. (1987). The medical computing "lag": Perceptions of barriers to the application of computers to medicine. *International Journal of Technology Assessment in Health Care, 3*(1), 123-136. doi:10.1017/S026646230001179X

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53*(282), 457-481. doi:10.2307/2281868

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In C. S. Mellish (Ed.), *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*: *Vol. 2* (pp. 1137-1143). San Francisco, CA: Morgan Kaufman.

Krishna, K., & Murty, M. N. (1999). Genetic k-means algorithm. *IEEE Transactions on Systems, Man, and Cybernetics-Part B*: *Cybernetics, 29*(3), 433-439. doi:10.1109/3477.764879

Kusiak, A., Dixon, B., & Shah, S. (2005). Predicting survival time for kidney dialysis patients: A data mining approach. *Computers in Biology and Medicine, 35*(4), 311-327. doi:10.1016/j.compbiomed.2004.02.004

Lee, L. S., Fiedler, K. D., & Smith, J. S. (2008). Radio frequency identification (RFID) implementation in the service sector: A customer-facing diffusion model. *International Journal of Production Economics, 112*(2), 587-600. doi:10.1016/j.ijpe.2007.05.008

Lee, S-H., Ng, A. W., & Zhang, K. (2007). The quest to improve Chinese healthcare: Some fundamental issues. *International Journal of Health Care Quality Assurance, 20*(5), 416-428. doi:10.1108/09526860710763334

Lewis, D. D. (1995). Evaluating and optimizing autonomous text classification systems. In E. A. Fox, P. Ingwersen & R. Fidel (Eds.), *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 246-254). New York, NY: ACM Press. doi:10.1145/215206.215366

Lin, H. M., Kauffman, H. M., McBride, M. A., Davies, D. B., Rosendale, J. D., Smith, C. M., Edwards, E. B., Daily, O. P., Kirklin, J., Shield, C. F., & Hunsicker, L. G. (1998). Center-specific graft and patient survival rates: 1997 UNOS report. *The Journal of the American Medical Association (JAMA), 280*(13), 1153-1160. doi:10.1001/jama.280.13.1153

Lin, R. S., Horn, S. D., Hurdle, J. F., & Goldfarb-Rumyantzev, A. S. (2008). Single and multiple time-point prediction models in kidney transplant outcomes. *Journal of Biomedical Informatics, 41*(6), 944-952. doi:10.1016/j.jbi.2008.03.005

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*: *Vol. 1* (pp. 281-297). Berkeley, CA: University of California Press.

Michaud, P. (1997). Clustering techniques. *Future Generation Computer Systems, 13*(2-3), 135-147. doi:10.1016/S0167-739X(97)00017-4

Mitchell, T. (1997). *Machine learning*. New York, NY: McGraw-Hill.

Ohno-Machado, L. (2001). Modeling medical prognosis: Survival analysis techniques. *Journal of Biomedical Informatics, 34*(6), 428-439. doi:10.1006/jbin.2002.1038

Olson, D. L., & Delen, D. (2008). *Advanced data mining techniques.* Berlin, Germany: Springer-Verlag.

Pareto, V. (1971). *Manual of political economy* (A. S. Schwier, Trans.). New York, NY: Augustus M. Kelley.

Parmar, M., & Machin, D. (1995). *Survival analysis*: *A practical approach.* Chichester, UK: John Wiley & Sons.

Principe, J. C., Euliano, N. R., & Lefebvre, W. C. (1999). *Neural and adaptive systems*: *Fundamentals through simulations*. New York, NY: John Wiley & Sons.

Ruiz, E., & Nieto, F. H. (2000). A note on linear combination of predictors. *Statistics & Probability Letters, 47*(4), 351-356. doi:10.1016/S0167-7152(99)00177-7

Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications, 145*(2), 280-297. doi:10.1016/S0010-4655(02)00280-1

Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice*: *A guide to assessing scientific models*. Chichester, UK: John Wiley & Sons.

Shechter, S. M., Bryce, C. L., Alagoz, O., Kreke, J. E., Stahl, J. E., Schaefer, A. J., Angus, D. C., & Roberts, M. S. (2005). A clinically based discrete-event simulation of end-stage liver disease and the organ allocation process. *Medical Decision Making, 25*(2), 199-209. doi:10.1177/0272989X04268956

Sheppard, D., McPhee, D., Darke, C., Shrethra, B., Moore, R., Jurewitz, A., & Gray, A. (1999). Predicting cytomegalovirus disease after renal transplantation: An artificial neural network approach. *International Journal of Medical Informatics, 54*(1), 55-76. doi:10.1016/S1386-5056(98)00169-5

Tjang, Y. S., Van der Heijden, G. J. M. G., Tenderich, G., Grobbee, D. E., & Korfer, R. (2008). Survival analysis in heart transplantation: Results from an analysis of 1290 cases in a single center. *European Journal of Cardio-Thoracic Surgery, 33*(5), 856-861. doi:10.1016/j.ejcts.2008.02.014

United Network for Organ Sharing. (2012). Chapter V: OPTN/STRN Annual Report. Retrieved from http://www.unos.org/

Yoo, H. Y., Galabova, V., Edwin, D., & Thuluvath, P. J. (2002). Socioeconomic status does not affect the outcome of liver transplantation. *Liver Transplantation, 8*(12), 1133-1137. doi:10.1053/jlts.2002.37000

Author Note

Asil Oztekin, Manning School of Business, University of Massachusetts Lowell, Lowell, MA 01854, USA.

Correspondence concerning this article should be addressed to Asil Oztekin, Email: Asil_Oztekin@uml.edu

The author is thankful to the Editor-in-Chief Vincent Charles, Special Issue Guest Editors, Joe Zhu and Rolf Färe, for processing this manuscript in a timely manner. Also, he is thankful to the anonymous reviewers for their invaluable suggestions to improve this paper.