An analytic approach to better understanding and management of coronary surgeries

Dursun Delen a,⁎, Asil Oztekin b, Leman Tomak c

a Department of Management Science and Information Systems, Spears School of Business, Oklahoma State University, USA
b Department of Operations and Information Systems, Robert J. Manning School of Business, University of Massachusetts Lowell, USA
c Department of Biostatistics, College of Medicine, Ondokuz Mayis University, Samsun, Turkey

ARTICLE INFO

Article history:
Received 14 March 2011
Received in revised form 29 October 2011
Accepted 4 November 2011
Available online 11 November 2011

Keywords:
Heart disease
Coronary artery bypass surgery (CABG)
Clinical decision support systems
Survival prediction
Data mining
Machine learning
Sensitivity analysis

ABSTRACT

Demand for high-quality, affordable healthcare services increasing with the aging population in the US. In order to cope with this situation, decision makers in healthcare (managerial, administrative and/or clinical) need to be increasingly more effective and efficient at what they do. Along with expertise, information and knowledge are the other key sources for better decisions. Data mining techniques are becoming a popular tool for extracting information/knowledge hidden deep into large healthcare databases. In this study, using a large, feature-rich, nationwide inpatient databases along with four popular machine learning techniques, we developed predictive models and using an information fusion based sensitivity analysis on these models, we explained the surgical outcome of a patient undergoing a coronary artery bypass grafting. In this study, support vector machines produced the best prediction results (87.74%) followed by decision trees and neural networks. Studies like this illustrate the fact that accurate prediction and better understanding of such complex medical interventions can potentially lead to more favorable outcomes and optimal use of limited healthcare resources.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

In recent years, healthcare has become one of the most spoken issues that have a direct impact on quality of life in the US and abroad. While the demand for healthcare services is increasing with the aging population, the supply side is having serious problems to keep up with the needed level and quality of service. In order to close the gap, healthcare systems ought to significantly improve their operational efficacy (i.e., effectiveness and efficiency). Effectiveness (doing the right thing, e.g., diagnosing and treating accurately) and efficiency (doing it the right way, e.g., using least amount of resources and using the least amount of time) are the two fundamental pillars upon which the healthcare system can be revived [16]. One promising way to improve the healthcare efficacy is to take advantage of advanced modeling techniques and large and feature rich data sources (true reflections of medical and healthcare experiences) to support accurate and timely decision making [16].

Decisions in healthcare can roughly be classified as either managerial or clinical. Managerial decisions involve optimal allocation of resources to various demand centers (which at the more detail level involves forecasting, capacity planning, scheduling, etc.). Medical decisions involve making the right diagnosis and implementing the right treatment. Computer systems that help in making medical decisions are often called medical decision support systems or clinical decision support systems. Clinical decision support systems (CDSS) are interactive computer programs, which are designed to assist physicians and other healthcare professionals with decision making tasks. The main goal of a CDSS is to link health observations (i.e., data) with health knowledge (i.e., expertise) to guide health choices by clinicians and managers for an improved healthcare. These CDSS systems encompass knowledge to guide the decision maker in optimal analysis of patient’s condition, and subsequently leading to accurate diagnosis and treatment. The knowledge component of these systems may come from the experts (via expert systems) and/or it can be extracted from historical data sources (via data mining).

According to the American Heart Association, cardiovascular disease (CVD) was the underlying cause for 20.67% of deaths in the US [14]. Since 1900, CVD has been the number one killer every year except 1918, which was the year of the great flu pandemic. CVD kills more people than the next four leading causes of deaths combined; cancer, chronic lower respiratory disease, accidents and diabetes mellitus. Out of all CVD deaths, more than half of them are attributed to coronary diseases. Not only does CVD take a huge toll on the personal health and well-being of the population, it is also a great drain on the healthcare resources in the US and elsewhere in the world. The direct and indirect costs associated with CVD for a year is estimated to in excess of $500 billion [14]. Even though the cost of a coronary artery bypass surgery depends on the patient and service provider related factors [6], the average rate is between $50,000 and $100,000 in the US [6].

As is the case in many complex problem situations, the outcome, if predicted accurately beforehand, could lead to better decision making.
Especially the prediction of survival time is a clinically important and challenging problem [18,32]. In the case of cardiac surgery, which in itself poses great risk to the patient, this becomes a matter of life and death. Today coronary artery bypass surgery (CABG) is a common surgical procedure thousands being performed every year in the US and more around the world. CABG surgery is advised for selected groups of patients with significant narrowing and blockage of the heart arteries. CABG surgery creates new routes around narrowed and blocked arteries, allowing sufficient blood flow to deliver oxygen and nutrients to the heart muscles [1,15]. The coronary artery disease happens with hardening of the arteries or plaque buildup on the walls of the arteries. Smoking can increase this plaque, as well as high blood pressure, high cholesterol, diabetes, among many others. Age also increases the risk, as well as similar family history. When the arteries narrow, the blood supply to the heart is not enough to meet increased oxygen demand. The heart muscle becomes starved of oxygen, and can cause chest pain or death of heart tissue [15].

In this paper, we report on a study where we employed various data mining methods to predict the outcome of coronary artery bypass surgery (CABG), and applied an information fusion based sensitivity analysis on the trained models to better understand the importance of the prognostic factors. Our main goal was to illustrate that predictive and explanatory analysis of large and future rich data sets provides invaluable information to make more efficient and effective decisions in healthcare. Machine learning-based prediction models integrated into a CDSS that provides accurate prediction of the procedure and help practitioners better understand the causal effects of the patient prognostic factors can improve the timeliness of the procedure and optimal allocation of resource, and hence, can improve the chances of success/survival for a significantly more number of patients. The rest of the paper is organized as follows. The next section provides a review of the relevant literature of this medical prediction domain. Section 3 describes the methodology (i.e., data, prediction model types and the evaluation methods used in the study) followed by Section 4 which provides the prediction and sensitivity analysis results. Finally, Section 5 summarizes the study, discusses the findings, and identifies the limitations and future research directions.

2. Literature review

In the last couple of decades, there have been numerous studies on the risk of death with CABG, but only recently a number of large single-center and multi-center cardiac surgical databases were established for use in risk stratification models. These databases typically contain patient and disease characteristics, of which inclusion (or exclusion) in a statistical method or algorithm could potentially influence the predictive power of the model. The committee led by K.A. Eagle to revise the 1991 Guidelines for Coronary Artery Bypass Graft Surgery revealed that variables relating to the urgency of operation, age, and prior coronary bypass surgery have the greatest predictive power, while variables describing coronary anatomy have the least predictive power [12]. Age has consistently predicted mortality after CABG, with advancing age being associated with higher mortality. Sex is also a predictor, with females facing an increased risk from CABG. A second CABG is of higher risk than the first and if the re-operation was carried out within 1 year of the primary operation, it poses further increased risk [12].

Various risk assessment models have been developed for cardiac surgery, and the first to become popular was the Parsonnet risk stratification system which was developed in the USA in the 1980s [26]. The model allocates additive predicted mortality percentage points to 14 patient risk factors to give a “Parsonnet score”; which is supposed to be indicative of the percent mortality for each patient, and has been used to categorize patients into good risk (predicted mortality 0% to 4%), fair risk (5% to 9%), poor risk (10% to 14%), high risk (15% to 19%), and extremely high risk (> 20%). However, it has been criticized for the nature of its statistical derivation and related restrictive assumptions. Furthermore, it is shown to constantly overestimate mortality, particularly for high risk patients, and its scoring system is fairly subjective [4].

Several prediction models have been developed to predict mortality following CABG. These models predict operative mortality, which is defined as death temporally or causally related to surgery (death within 30 days of operation or in the same hospital admission as operation, regardless of cause). To exemplify, Bridgewater et al. [4] compared multivariate statistical data analysis models using cross-validation, with a confidence level of 95% and reporting the predictive ability of these models in terms of the area under the Receiver Operating Characteristic (ROC) curve. The overall operative mortality was between 2.8% and 3.7%, whereas the mean overall predicted mortality varied between 1.07% and 4.7% for various models [4]. They also argued that the models developed for North American patients did not work well in the UK. Dudley et al. [11] searched the effect of clinical factors on the cost of CABG surgery. They utilized three statistical models, namely ordinary least squares (OLS) linear regression, logistic regression with Cox proportional hazards model, and logistic regression with Weibull distribution. The logistic regression with Cox proportional hazards model gave considerably better results showing that lower ejection fraction and older age are independent clinical predictors of increased cost of CABG [11].

As data mining techniques have gained a lot of attention in the healthcare and medical informatics due to their high prediction accuracies [9,24,25], they have started to replace (or at least complement) the conventional statistical data analyses in CABG surgeries. Warner and Misra [33] developed and compared a neural networks model against linear regression to predict the mortality 30 days after a CABG surgery. Their neural network model produced as good as they are compared to the logistic regression model and even better calibration. Therefore, they concluded that neural networks, which are non-linear regression methods, enable the modeler to reveal complex forms of functions, leading to better prediction accuracy. Mobley et al. [20] also made use of a neural network-model based prediction tool for distinguishing patients who need coronary angiography from those who do not. This study, besides producing promising results, revealed the importance of ranked variables (identifying and including the most important ones) in predicting the outcomes of coronary artery stenosis. Nilsson et al. [22] developed a neural networks-based prediction method to select risk variables and predict mortality after cardiac surgery. The rank order of risk factors contributing to the mortality prediction was identified. Their ROC-curve value was higher than the conventional logistic European System for Cardiac Operative Risk Evaluation model (0.81 compared to 0.79). Rowan et al. [29] developed a neural networks-based software to successfully predict and stratify the length of stay (LOS) of cardiac patients (such as CABG and valve repair/replacement surgery) based on preoperative and postoperative factors. Chi et al. [5] developed a decision support system using calibrated support vector machines, which predicts the survival outcome of hospitals after CABG surgeries. In this study, calibrated support vectors significantly overcame the linear regression model to predict CABG survivability.

Even though a number of previous studies had some level of success in developing models that predicts and/or explains the outcome of a CABG surgery, our study stands out in two respect: first, we used a much larger, feature-rich, nationwide dataset along with proven machine learning techniques provided us with more generalizable and reliable information about the prognostic nature of this medical procedure. The four prediction model types are chosen based on their predictive abilities reported in the recent data mining literature,
and confirmed by our preliminary experiments with the same dataset. Our preliminary experiments showed that more traditional algorithms like logistic regression, discriminant analysis and Naive Bayes were not comparable in their predictive ability to the machine learning techniques used in this study. Additionally, and perhaps more importantly, we have conducted information fusion based sensitivity analyses on the trained prediction models to identify the significant prognostic factors and better understand the causal relationships.

3. Research method

At the highest level, the methodology that we followed in this study can be depicted as a 4-step process as illustrated in Fig. 1. Step 1 involves data identification, data acquisition, and data preprocessing tasks. The outcome of this step is a pre-processed dataset which can be used in the rest of the steps. Step 2 involves the model building and model calibration tasks. Here we used four different types of models (artificial neural networks, support vector machines, and two types of decision trees, namely C5 and CART), and went through a large number of experimental runs to calibrate the modeling parameters for each model type. The outcome of this step is a list of trained prediction models which only used the training step. Step 3 involves the comparative testing of the prediction models. Specifically, it involves applying and recording the prediction ability (as measured by accuracy, sensitivity and specificity) of each of the four model types using the test dataset. Step 4 is the final step, where trained models are exposed to a sensitivity analysis procedure where the comparative contribution of the variables to the prediction is measured.

3.1. Data description and preprocessing

One problem that has hindered data discovery efforts in healthcare, like many other industries, is the fact that available databases have been developed for reasons other than analytic modeling and/or knowledge discovery efforts. The data is often “dirty”; containing missing or erroneous values and has a large amount of redundancy [16]. This is, to some degree, because data comes from multiple sources that use different data items, categories, and specifications. However, many academic centers and governmental agencies are establishing and developing clinical data repositories (CDRs) that are databases containing a variety of comprehensive clinical and administrative data on large patient cohorts over time [21]. Another issue that is unique to medical data discovery efforts is the need to maintain privacy of medical information. In the US, this is legally required under the Health Insurance, Portability, and Accountability Act (HIPAA), which provides penalties up to $10,000 for each incidence (record) unlawfully disclosed.

For this project, we used the Nationwide Inpatient Sample (NIS) databases, which are a part of the Healthcare Cost and Utilization Project (HCUP), sponsored by the Agency for Healthcare Research and Quality (AHRQ). It is a database that stores the specifics about hospital inpatient stays. The NIS is the largest all-payer inpatient care database that is publicly available in the US for a small fee. The data includes approximately 39 types of data elements in the core tables. There are an additional 32 severity measures and 25 hospital related elements in additional two tables. The severity measures are primarily co-morbidities; chronic illnesses that could affect the outcome of medical treatment. The entire database represents approximately 8 million discharges from over 1000 hospitals in 37 states for 2004. The data files are in ASCII format and the core file was nearly 4 gigabytes in size. Because of the sheer size of the dataset, and the need for multi-table merging, the data was imported into MS SQL server for integration, and subsequent processing.

The first step in data evaluation was to narrow the database to the population of patients undergoing CABG surgery. This was done based on the diagnostic related group (DRG) code assigned at discharge of the patients. DRGs, standardized codes for medical procedures, were introduced by Medicare in 1983 as part of its prospective payment system to hospitals. Each patient’s hospitalization is assigned a DRG code based on the patient’s diagnosis, procedures done during the hospitalization and patient demographic information. After consulting to the domain experts, the following DRG codes found to be directly related to CABG surgery.

106 Coronary Bypass with Angioplasty.
107 Coronary Bypass with Cardiac Catheterization.
109 Coronary Bypass without Cardiac Catheterization or Angioplasty.

The integrated, complete dataset is filtered for these three DRG codes. That is, only the records containing these DRG codes are included in the dataset, and all other records are excluded. After eliminating unrelated records and variables, the resulted dataset included over 50 individual variables and over 60 thousand separate admission records. Using Microsoft Excel® (Excel 2010) and Statistica Data Miner® (Statistica Data Miner v9.1), each of these variables was evaluated for quality and completeness of data values. It was found that 61 records lacked usable data for the dependent variable. These were deleted from the data set. There were three records with a negative integer in the age field, which were deleted. An additional 16 records that lacked valid information for the patient’s gender were also deleted.

When considering the admission source, it was felt that the most important distinction was between emergent and routine admission status. There were 29,140 records showing routine admission, 11,638 showing admission through the emergency department, 9908 from other sources and 205 records without data for this element. The other sources included other hospitals, other healthcare facilities (including long-term care) and court/law enforcement. It was not possible to determine whether the patients coming from other sources were of an emergency nature. This group and those without data totaled 10,113; nearly as many as the known emergent group. Another data element specified whether the admission was elective or non-elective. Examination of the elective variable revealed that 158 records did not have valid data, and were removed from the data set.

Initially it was decided to include the top five admission diagnoses for each admission. Upon inspection of the data, it was found that complete data was available for only the first admission diagnosis. Subsequent diagnoses had a significant number of missing or invalid data elements. Furthermore, the first diagnosis included many different forms of the same general diagnosis. For example, there were more than 10 different types of heart attack diagnoses. It was then considered using another variable from the original data set that categorized the diagnoses. However, this too would result in too
many different categories to be useful to our classification study. Therefore, this field was removed from the data set.

Upon further discussion with the domain experts regarding the variables for length of stay and number of procedures during the hospitalization, concerns were raised that these two elements could not be known prior to the decision for surgery and therefore were not appropriate predictors of outcome. They were removed from consideration. The significance of the payer source on outcome was determined to be related to the possibility that patients who lacked insurance or other third-party payment assistance may delay seeking care and result in increased risk of mortality. For this reason, this categorical variable was converted into a binary (i.e., self-pay versus any payer source) and included in the independent variable list.

The patient location variable included four categories. These were redefined/aggregated into a binary variable; urban versus rural. After experimental analysis, it was determined that the potential impact on mortality was better defined in this way. The variable representing the median household income for the patient’s zip code (based on 1999 demographics) contained four levels of income and was left as is. All other variables were included as they were. Except as noted above, any missing or invalid data elements were simply left blank. Table 1 shows the list of variables involved in the study along with our reasoning for inclusion.

After the data cleaning tasks, the complete data set used further processing consisted of 40 individual variables and 50,656 records. The output variable (i.e., “died”) had unbalanced representation of positives and negatives. Our preliminary runs with this highly skewed data indicated that all model types are favoring the accurate prediction of the most representative class (i.e., Negative) in the expense of the other class. In prediction-type problems (such as the one in this study) it is advisable to balance the dataset for better and more representative prediction results [34]. Therefore, we balanced the data so that we would end up with a roughly equal representation of the two class labels. Specifically, we included the entire set of the less represented class along with a randomly selected, roughly equal number of records of the more representative class in constructing the final dataset. This pre-processed dataset is then partitioned (using a stratified random sampling technique) and fed into the model building, model testing and sensitivity analysis tasks.

### 3.2. Modeling and analyses

As the objective in this study was to predict the survival after a CABG surgery, all four modeling techniques used were of classification types (even though three out of the four is also capable of modeling regression type prediction problems). These four modeling techniques are chosen based on (1) their popularity and superior prediction ability for similar problems published on the data mining literature [23], and (2) our preliminary experiments with the same dataset, where they have produced significantly better predictive ability compared to their more traditional counterparts including logistic regression, discriminant analysis and Naïve Bayes. What follows is a brief description of the four classification-type prediction methods used in this study:

**Artificial neural networks (ANNs)** have been utilized to model complex relationships (such as highly-nonlinear relational functions) among the predictor variables and the dependent variable [19]. ANNs 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 [13]. ANNs have been one of the most popular artificial intelligence-based data modeling algorithms used in recent medical informatics studies due to their pleasing predictive performance [2]. In this study, we used the multi-layer perceptron (MLP) (an ANN topology/architecture) with back-propagation learning mechanism. Our experimental runs showed that, for this type of classification problem, MLP performs better than other ANN architectures such as radial basis function (RBF), recurrent neural network (RNN), and self-organizing map (SOM). In fact, Hornik et al. [17] empirically showed that given the right size and structure, MLP is capable of learning arbitrarily complex nonlinear functions to arbitrary accuracy levels.

**Support vector machines (SVMs)** belong to a family of generalized linear models which achieves a classification or regression decision based on the value of the linear combination of features (i.e., independent variables). The mapping function in SVMs can be either a classification function (used to categorize the data, as is the case in this study) or a regression function (used to estimate the numerical value of the desired output). For classification, nonlinear kernel functions are often used to transform the input data (inherently representing highly complex nonlinear relationships) to a high dimensional feature space in which the input data becomes more separable (i.e., linearly separable) compared to the original input space. Then, the maximum-margin hyperplanes are constructed to optimally separate the classes in the training data. Two parallel

<table>
<thead>
<tr>
<th>Variable</th>
<th>Inclusion criteria</th>
<th>Type of variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Previous research shows this to be a significant predictor</td>
<td>Continuous</td>
</tr>
<tr>
<td>Admission source</td>
<td>It was felt that patients who had an emergent admission may have an increased chance of mortality</td>
<td>Categorical</td>
</tr>
<tr>
<td>Admission on weekend</td>
<td>It was wondered whether admission on the weekend could cause a delay in surgery that affected outcome</td>
<td>Binary</td>
</tr>
<tr>
<td>Died</td>
<td>Dependent variable</td>
<td>Binary</td>
</tr>
<tr>
<td>DRC18</td>
<td>Used for sorting/partitioning only</td>
<td>Categorical</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Initially included were the first five admission diagnoses</td>
<td>Categorical</td>
</tr>
<tr>
<td>Elective</td>
<td>Emergent vs elective admission status could impact outcome</td>
<td>Binary</td>
</tr>
<tr>
<td>Female</td>
<td>Gender is a well-recognized risk factor</td>
<td>Binary</td>
</tr>
<tr>
<td>Hospital state</td>
<td>Used for sorting/partitioning only</td>
<td>Categorical</td>
</tr>
<tr>
<td>Key</td>
<td>Used to link data tables</td>
<td>Unique Identifier</td>
</tr>
<tr>
<td>Length of stay</td>
<td>It was anticipated that patients with very short or very long LOSs would have increased mortality</td>
<td>Continuous range (0–160 days)</td>
</tr>
<tr>
<td>Number of procedures</td>
<td>It was felt that patients who required a higher number of procedures during the hospitalization were at higher risk</td>
<td>Categorical</td>
</tr>
<tr>
<td>Pay source</td>
<td>It was questioned if self-pay patients were less likely to seek timely care and thus were at increased risk</td>
<td>Categorical</td>
</tr>
<tr>
<td>Patient’s home location</td>
<td>Whether the patient lives in an urban vs rural location may affect timeliness of treatment</td>
<td>Categorical</td>
</tr>
<tr>
<td>Race</td>
<td>It was questioned whether race was a significant factor in survival</td>
<td>Categorical</td>
</tr>
<tr>
<td>Patient’s home zip code income level</td>
<td>It was questioned whether patients from more affluent areas had different outcomes from others</td>
<td>Categorical</td>
</tr>
<tr>
<td>Co-morbidity measures</td>
<td>All 29 items were included (see Appendix A for a complete list of these co-morbidity measures)</td>
<td>Binary</td>
</tr>
<tr>
<td>Hospital location</td>
<td>Urban vs rural location of the hospital could affect outcome</td>
<td>Categorical</td>
</tr>
<tr>
<td>Hospital teaching status</td>
<td>Teaching status of the hospital could have an effect on outcome</td>
<td>Binary</td>
</tr>
</tbody>
</table>
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 classifier will be [7]. SVMs are gaining popularity of being an excellent alternative to ANNs for prediction type problems in medical informatics.

Decision trees (DTs) are another viable and increasingly more popular alternative to ANNs and SVMs in prediction type problems. Compared to other machine learning methods (such as ANNs and SVMs), decision trees have the advantage of not being a black box modeling technique, having the capability to explain the inner structure of the model in the form of a graphically represented inverse tree or a collection of condition-action rules. This advantage has made them an increasingly more popular alternative method in medical informatics [10]. Popular decision tree algorithms include Quinlan’s ID3, C4.5, C5 [27,28] and Breiman et al.’s CART (Classification and Regression Trees) [3]. Based on the favorable prediction results we have obtained from the preliminary runs, in this study we chose to use both CART and C5 (an improved version of C4.5 and ID3) algorithms as our decision tree-based prediction method.

A process picture that shows training, testing and sensitivity analysis of all four model types is shown in Fig. 2. As can be seen, the pre-processed data is presented in an Excel file. This complete dataset is then randomly split into mutually exclusive training and testing partitions (i.e., 2/3 training dataset and 1/3 testing data set). These exact partitioned dataset are then fed into the model building and model testing procedures of all model types. Once the models are trained and tested, the trained models are also used in sensitivity analysis procedure. The results coming from testing procedures are collected and tabulated into confusion matrices, of which are then used to calculate the three performance metrics (i.e., accuracy, sensitivity and specificity). The sensitivity analysis results of all four model types are collected, normalized and fused into a single table, which is then used to create a bar chart for visual representation of the ranked importance of the variables.

4. Results

4.1. Prediction results

To compare the classification models, three performance criteria are adopted as follows:

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)
\]

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \quad (2)
\]

\[
\text{Specificity} = \frac{TN}{TN + FP} \quad (3)
\]

where \(TP, TN, FP, FN\) denote true positive, true negative, false positive, and false negative, respectively. Accuracy, shown by Eq. (1), measures the proportion of correctly classified test examples, therefore predicting the overall probability of the correct classification. Sensitivity and
specificity, shown by Eqs. (2) and (3) respectively, measure the model’s ability to predict the individual groups within itself (i.e., how accurately model predicts the positives and negatives). Table 2 shows all of the validation results (using only the test dataset) for all four model types in a tabular format.

As the results indicate, SVMs produced the best performance with a prediction accuracy of 87.74% on the holdout sample (also produced the best sensitivity and specificity results with 89.48% and 86.01% respectively). The second best model was the C5 decision tree algorithm with a prediction accuracy of 79.62% on the holdout sample (and sensitivity and specificity results of 80.29% and 78.96% respectively). The other two methods, ANNs and CART decision tree algorithms took the third and the last places with prediction accuracy of 74.72% and 71.15% on hold-out sample respectively.

4.2. Information fusion-based sensitivity analysis results

In machine learning, sensitivity analysis is a method for extracting the cause and effect relationship between the inputs and outputs of a trained prediction model [8]. Sensitivity analysis is similar to feature selection in that they both try to find the relative importance of the independent variables (features) as they relate to the output variable. The main difference is when they are used. Feature selection is often used before starting to build prediction models in order to identify which independent variables should be included in the model building phase. On the other hand, sensitivity analysis is used after the models are built and sufficiently accurate prediction results are obtained. Sensitivity analysis provides the user with information about the comparative contribution that each independent variable provides in terms of their importance according to the sensitivity measure deﬁned in such a way that the sensitivity measure of the variable $i$ with information fused by $m$ prediction models can be represented as Eq. (5).

$$S_{i/m} = \sum_{n=1}^{m} \omega_n S_{i,n} = \omega_1 S_{i,1} + \omega_2 S_{i,2} + \ldots + \omega_m S_{i,m}$$

where $\omega$ refers to the normalized weight values (i.e., importance of models, proportional to their predictive power) of each prediction model with $m$ models in total (where $m$ is equal to 4 in this study) and $S_{i,n}$ is the sensitivity measure of the $n$th variable in the $i$th model. Table 3 shows a bar chart representation of the information fused sensitivity analysis results.

As the results indicate, the most important variables (in determining the CABG outcome) are AGE (the age of the patient), CM_HTN_C (Hypertension), CM_RANFAIL (Renal failure) and PAY1 (Payment type). In all four model types, the variable AGE came out to be the most important variable, which has also been found to be important, but not necessarily the most important factor in predicting the outcome of a CABG surgery in previously reported studies. Even though the rest of the variables were also ranked similarly in four model types, apart from the AGE variable, there was not a unanimous consensus among all models on the ranking of the factors.

5. Discussion and conclusions

Data mining, as a knowledge discovery tool, is becoming a popular enabler for improving the decision performance and hence the effectiveness and efficiency in healthcare. Better information and knowledge provided to the right person at the right time in a healthcare setting will undoubtedly lead to better decisions and more favorable outcomes. Such data/fact driven information/knowledge resources packaged in a user-friendly decision support system can be an invaluable aid to both managerial as well as clinical decision makers.

In this study, we showed the power of data mining (i.e., machine learning techniques) in predicting the outcome and in analyzing the prognostic factors of complex medical procedures such as CABG surgery. It is shown herein that using a number of methods (as opposed to only one) in a competitive experimental setting has the potential to produce better predictive as well as explanatory results. Among the four methods that we used, SVMs produced the best prediction results with prediction accuracy of 88% on the hold-out sample. The information fusion-based sensitivity analysis results revealed the ranked importance of the independent variables. Some
of the top variables identified in this analysis having to overlap with the most important variables identified in previously conducted clinical and biological studies confirm the validity and effectiveness of the proposed data mining methodology.

From the managerial standpoint, clinical decision support systems that uses the outcome of data mining studies (such as the ones presented in this paper) are not meant to replace healthcare managers and/or medical professionals. Rather, they meant to support them in making accurate and timely decision to optimally allocate resources in order to increase the quantity and quality of medical services. There still is a long way to go before we can see these decision aids being used extensively in healthcare practices. Among others, there are behavioral, ethical, and political reasons for this resistance to adoption. May be the need and the government incentives for better healthcare systems will expedite the adoption.

Although data mining methods are capable of extracting patterns and relationships hidden deep into large medical datasets, without the cooperation and feedback from the medical professional, their results would have very limited use. The patterns found with 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 fuel new biological and clinical research directions, and become a part of medical decision support systems.

Data mining methods rely on the existence and the correctness of data. If the data is not present or it is not accurate, the data mining results would be worthless; exemplifying the famous GIGO (garbage in garbage out) syndrome. In order to reap the benefits of data mining, healthcare organizations should pay extra attention to collecting, storing and making available the proper data sources while complying with the privacy related laws and regulations (e.g., HIPPA, Sarbanes Oxley and etc.).

In summary, as the results indicated, machine learning techniques are capable of predicting the outcome of complex medical procedures (such as CABG surgery, which was the case used in this study), providing invaluable information for decision making. A good prediction model does more than just predict the outcome; it also provides information on the ranked importance of the prognostic factors. Having a good understanding of the medical situation helps improve the effectiveness and the efficiency of clinical decision making and hence optimal use of healthcare resources; not to mention it would save more lives, which unarguably is the most important benefit one would look for.
References


[34] D. Sheppard, D. McPhee, C. Darke, B. Shretha, R. Moore, A. Jurewitz, A. Gray, Predictive data mining and quality control techniques in clinical chemistry.

Dr. Dursun Delen is the William S. Spears Chair in Business Administration and Associate Professor of Management Information Systems in the Spears School of Business at Oklahoma State University (OSU). He received his Ph.D. in Industrial Engineering and Management from OSU in 1997. Prior to his appointment as an Assistant Professor at OSU in 2001, he worked for a privately-owned research and consultancy company, Knowledge Based Systems Inc., in College Station, Texas, as a research scientist for five years, during which he led a number of decision support and other information systems related research projects funded by federal agencies, including DoD, NASA, NIST and DOE. His research has appeared in major journals including Decision Support Systems, Communications of the ACM, Computers and Operations Research, Computers in Industry, Journal of Production Operations Management, Artificial Intelligence in Medicine, Expert Systems with Applications, among others. He recently published three books: Advanced Data Mining Techniques with Springer, 2008; Decision Support and Business Intelligence Systems with Prentice Hall, 2010; and Business Intelligence: A Managerial Approach, with Prentice Hall, 2010. He is often invited to national and international conferences for keynote addresses on topics related to Data/Text Mining, Business Intelligence, Decision Support Systems, and Knowledge Management. He served as the general co-chair for the 4th International Conference on Network Computing and Advanced Information Management (September 2–4, 2008 in Seoul, South Korea), and regularly chairs tracks and mini-tracks at various international conferences. He is the associate editor-in-chief for International Journal of Experimental Algorithms, associate editor for International Journal of RF Technologies, and is on editorial boards of five other technical journals. His research and teaching interests are in data and text mining, decision support systems, knowledge management, business intelligence and enterprise modeling.

Dr. Asil Oztekin received the B.S. degree from Yildiz Technical University, Istanbul, Turkey, in 2004, and the M.S. degree from Fatih University, Istanbul, Turkey, in 2006, both in industrial engineering. He also worked for a privately-owned research and development company in the Industrial Engineering and Management at Oklahoma State University (OSU), Stillwater, in 2010. He joined the Robert J. Manning School of Business at University of Massachusetts Lowell in July 2011. Prior to joining UMass Lowell, he worked as a Visiting Assistant Professor in the Department of Statistics at Oklahoma State University. His research interests include medical informatics, healthcare engineering, decision analysis, multivariate data analysis, data mining, quality engineering, and human–computer interaction. His work has been published in leading journals such as Decision Support Systems, International Journal of Production Research, Production Planning & Control, International Journal of Medical Informatics, Artificial Intelligence in Medicine, International Journal of Industrial Ergonomics, and etc. He edited a special issue “Intelligent Computational Techniques in Science, Engineering, and Business” in Expert Systems with Applications journal. Dr. Oztekin has reviewed manuscripts for IIE Transactions, Decision Support Systems, International Journal of Production Research, Production Planning & Control, Computers in Industry, Journal of Systems and Software, Intelligent Information Management, and Applied Clinical Informatics. Dr. Oztekin is a member of AQR, IIE, and INFORMATS and was the recipient of the Alpha Pi Mu Outstanding Industrial Engineering and Management Research Assistant Award from OSU in 2009.

Dr. Leman Tomak is an Associate Professor of Biostatistics in the College of Medicine at Ondokuz Mayis University, Samsun, Turkey. She received her M.D. degree from the College of Medicine in Ondokuz Mayis University in 1995. After obtaining her degree, she worked at several hospitals and medical centers for three years on Public Health Care and Medicine under the supervision of the College of Medicine at Ondokuz Mayis University until 2003. She then obtained a Master of Biostatistics degree from the College of Medicine at Ondokuz Mayis University in 2008. As part of her graduate study, she has completed two theses. The subject of the first thesis is the nutrition in adolescents and the relationship between nutrition and development. The subject of the second was quality control and quality control techniques in clinic chemistry laboratory. She is currently working in Biostatistics Department as a full-time faculty member in the College of Medicine at Ondokuz Mayis University. She has published in leading medical informatics and biostatistics journals. Her research interests include analytical and statistical techniques in medicine and healthcare and quality control techniques in clinical chemistry.