Clinical Information System in improving Surgery Patient Care

Nan-Chen Hsieh[1], Jui-Fa Chen[2], Hsin-Che Tsai[3]

Department of Information Management,
National Taipei University of Nursing and Health Sciences, Taiwan[1]
Department of Computer Science and Information Engineering,
Tamkang University, Taiwan[2,3]
nchsieh@gmail.com[1]
alpha@mail.tku.edu.tw[2]
shin@live.com[3]
Corresponding Author: nchsieh@gmail.com

Abstract

A computerized clinical information system can make clinical care more efficient while improving patient care by providing surgeons with performance feedback. This study developed a clinical information system for cardiovascular disease to facilitate surgery treatment tracking. The system tracks hundreds of variables of data through surgical stages and converts these data into computerized registries, and generates clinical interpretive reports in a timely manner. Surgeons can use the system to identify surgical-related data and interventional cardiovascular procedure risks based on specific patient characteristics, and it has increased the quality and efficiency of patient care. An intelligent data analysis (IDA) tool based on the Weka library that seamlessly integrates with the clinical information system has helped provide prognosis models for clinical research. The IDA tool includes an ensemble model, classification modules, model evaluation, and data visualization. The probabilities of complication calculated by the model of complications and a receiver operating characteristic (ROC) curve were used to evaluate the accuracy of postoperative morbidity prediction.

Keywords: Clinical information system, cardiovascular disease, intelligent data analysis, machine learning

1. Introduction

Cardiovascular disease is a type of disease that involves problems with the heart or blood vessels. Major cardiac surgical interventions include coronary artery bypass grafting (CABG), repair of congenital heart defects, surgical treatment of atrial fibrillation, heart transplantation, repair or replacement of heart valves, aortic surgery, aneurysm repair or a combination of these surgical procedures. A computerized Clinical information system tracks previous patients' outcome data completely and correctly and potentiate surgeries with substantial and accurate data for clinical research. Once the research results for decisions are stored sagaciously, queries can be performed to investigate clinical issues, treatments, procedures, and outcomes.

Many prediction models for cardiac surgical outcome apply logistic or multivariable regression to assess preoperative risk (Barnes, Boult, Maddern, & Fitridge, 2008; Bohm et al., 2008; Stijn, Wouters, Freek, & Rene, 2002). Most risk assessment tools used traditional statistical method to derive prediction models for patients who were undergoing open abdominal aortic aneurysm repair, and to evaluate whether or not the surgical procedures would be successful. Similarly, postoperative morbidity is a key factor in recovery and through-put of cardiac hospital patients. Prediction of surgical mortality and postoperative morbidity is important in selecting low-risk patients for operation, and in counseling patients about the risks of undergoing surgical operation. The development of a robust prediction model can therefore both assist vascular surgeons in evaluating the expected outcome for a given patient and facilitate counseling and preoperative decision-making. Reliable and accurate prediction of operative mortality and morbidity is an essential criterion for any such risk evaluation models.

The original paper-based special chart for cardiovascular disease was developed based on foreign experience (STS Adult Cardiac Surgery data collection form, http://www.sts.org/sections/stsnationaldatabase/), fused with local population characteristics (Cheng-Hsin Cardiac Surgery registries Version 2.52.2). The data in the special chart are scattered over three surgical stages: preoperative, intraoperative and postoperative. To facilitate the diagnostic and therapeutic interventions performed by surgeons, nurses and professional staff members, the particular diagnosis or procedure is structured as an interdisciplinary care plan and sequenced on a uniform timeline. All related personnel were collaborators concerning these clinical pathways.

The use of machine learning models has become widely accepted in medical applications. Delen et al. (2007) developed a Web-based Decision Support Systems, which are used to build different types of prediction models , including neural networks, decision trees, ordinal logistic regression and discriminant analysis to classify a movie. Murphy presented a Clinical Decision Support System use of Decision Tree is one of the most popularly applied methods. Artificial Neural Network (ANN) has featured in a wide range of medical applications, often with promising results (Murphy, 2001). Eom et al. developed a classifier ensemble-based, including Artificial Neural Network (ANN), Decision Tree (DT), and Support Vector Machine (SVM), clinical decision support system for predicting cardiovascular disease level. SVM have been successfully used in a wide variety of medical applications (Eom, Kim, & Zhang, 2008). Polat and Güne used a least square support vector machine to assist breast cancer diagnosis (Polat & Güne, 2007). Babaoĝlu et al. first used principle component analysis method to reduce data features, and acquired an optimum support vector machine model for the diagnosis of coronary artery disease(Babaog˘lu, Fındık, & Bayrak, 2010). Choi proposed the detection of valvular heart disorder (VHD) by wavelet packet decomposition and SVM techniques (Choi, 2008).

This study describes the development of an informative ensemble prediction model in web-based system, consisting of DT, ANN and SVM for the prediction of postoperative morbidity between preoperative variables and complication outcomes in Cardiovascular disease patients. For a better understanding of our study, Section 2 of this paper begins with an overview of study background and system overview in general. Section 3 describes the procedures used in this study, ensemble model for the prediction of postoperative morbidity. Section 4 discusses the experimental findings and offers observations about practical applications and directions for future research.

2. Clinical Information Systems

Hospital mortality is an important clinical endpoint in cardiovascular disease. A prognostic model is often used for clinical decision-making prior to and during such surgery. The predicted outcome can be used by surgeons and patients to evaluate whether or not the surgical procedures is likely to be successful. Similarly, postoperative morbidity is a key factor in recovery and throughput of cardiac hospital patients. Prediction of surgical mortality and postoperative morbidity is important in selecting low-risk patients for operation, and in counseling patients about the risks of undergoing surgical operation. The development of a robust prediction model can therefore both assist vascular surgeons in evaluating the expected outcome for a given patient and facilitate counseling and preoperative decision-making. Reliable and accurate prediction of operative mortality and morbidity is an essential criterion for any such risk evaluation models.

The development of a Clinical information system allows surgeons to create large, integrated databases of disease-specific patient information, which helps surgeons begin the real-time management of populations of similar patients. Mining data in these databases may provide insights into new relationships between disease states and how to effectively manage them. In addition, system can help surgeons access data for several purposes: (1) to interpret these patient data for population-based clinical decisions, (2) to analyze the patient data for outcome studies, (3) to develop evidence-based decision-aided models, and (4) to monitor the performance of individual surgeons and the quality of the care delivered by the healthcare organization.

Using data mining techniques to support medical decisions would increase diagnostic accuracy and provide additional knowledge to the medical staff. The increased use of these techniques provides expanded opportunities for determining the utility of medical decision-making models from longitudinal clinical data. Mining these clinical data provides additional insights through advanced analytics, including visualize analysis, clustering analysis, and classification/predictive modeling. The newly discovered insights can serve as evidence for producing accurate judgments. The predictive and interpretable descriptive models contribute greatly to handling medical data gathered through the systematic use of clinical, laboratory, and existing clinical systems (Bellazzi & Zupan, 2008). In this study, the knowledge extracted with the Intelligence Data Analysis (IDA) tool may be in the form of diagrams, charts, or predictive/descriptive models. The goal of predictive data mining is to derive models that can use medical data to predict a patient's mortality and morbidity and thus support clinical surgical decision making.

The prototype IDA tool was developed as a module within Clinical information System. Figure 1 shows how architecture was based on dot.Net framework with SQL server and machine learning functionalities accomplished in a Weka open source library, which was designed to leverage efficient access to clinical data and support for clinical decision making. Web services are used to achieve interoperability between the Medical Database and IDA modules because web-based systems are platform- and programming language-independent, standard protocols are accessible, and they facilitate communication. IDA tool adheres to a flexible design because the open source license allowed us to freely access the library and make it available for other uses without constraints. We combined Medical Database and asynchronous web service calls, making accessing immediate data analysis services from the WEKA library possible (Shetty, S.Vadivel, & Vaghella, 2010). Figure 1 shows the web-based IDA tool that centralizes the support prognosis task. Users can select the desired fields for building classifiers and check the performance of the predictive model results.

The Intelligent Analysis Tool descriptive statistics functions and Data Mining functions were supported by Weka library. The Intelligent Analysis Tool descriptive statistics functions and Data Mining functions were supported by Weka library. User can choose ensemble model including DT, ANN and SVM to construct a set of hypotheses to predict risk. This study designed intelligent Analysis Tool facilitates different types of users to access functionalities in Weka without the need to learn new data mining software.

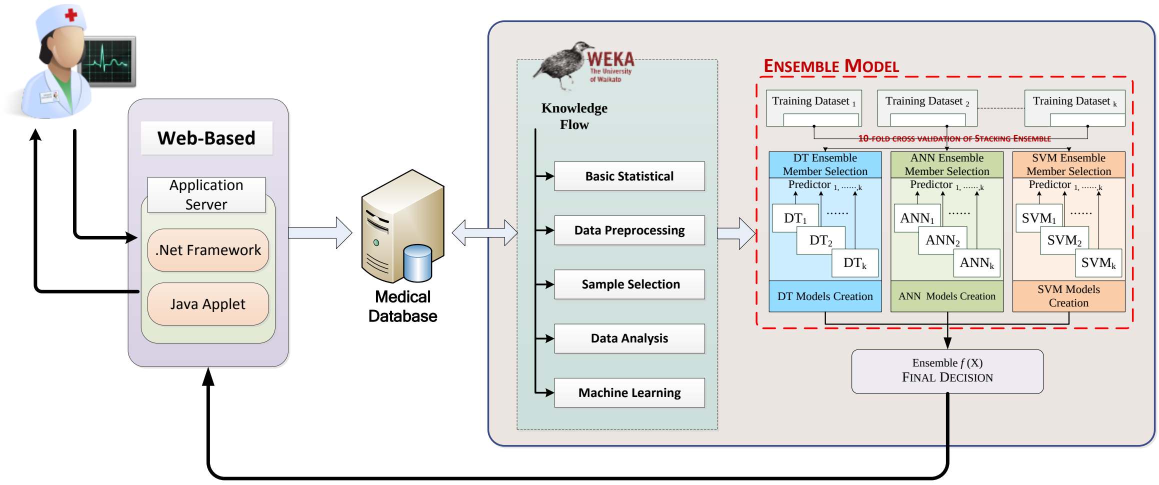


Figure 1. The proposed architecture

3. The Functionality of Intelligent Data Analysis

Data mining tools are usually run from the desktop as stand-alone application software. This approach for the use of data analysis functions is inconvenient and not easily accessible for those users without training and the installed software. Because web services provide easy-to-deliver and accessible information resources on the network, in this study, we developed an IDA tool as a web application, and any client in the network can install and access the IDA tool. The corresponding algorithms in the Weka library enabled data analysis tasks by interaction with the web service and only returned the analyzed results, thus not requiring the downloading of the entire dataset. An IDA tool is therefore seamlessly integrated with medical databases.

3.1 Data Input

We retrospectively examined 122 consecutive patients who underwent Open Repair surgery at Taipei Veteran General Hospital, a teaching hospital in Taiwan, between 2007 and 2009. The dataset contains preoperative patient characteristics, details of the operative information, and pathological and laboratory findings from the emergency room, operating room and intensive care unit. The dataset also included length of ICU stay, variables that describe postoperative complications that frequently occur in Open Repair surgery, death during hospitalization, and time of death for patients who expired. Postoperative complication was used as the binary outcome variable of the ensemble model, and types of complications were used as subsidiary outcome variables. The original dataset contained 154 variables, but included many missing values. Preliminary inspection of the dataset showed that many variables contained missing values for at least 50% of the patients; these variables were not included in further analysis. In order to identify significant variables for use in the ensemble model, a number of criteria were employed. Variables that were subjective, ambiguous or inadequately defined were excluded; variables that were frequently incomplete were also excluded from subsequent analysis. Data collected included preoperative patient characteristics, risk factors, details of the operative information, physical characteristics of the aneurysm, postoperative physiological and laboratory findings, and postoperative complications as the outcome variable.

3.2 Ensemble model

DTs, ANNs, and SVMs were chosen as basis models for the ensemble model, because they represented different approaches, each of which is simple to implement and has been shown to perform well in medical applications. The rationale of employing these models is that Decision Tree have simplicity and capacity for humanly understandable inductive rules so many researchers use Decision Tree to resolve problems and error analysis. Artificial Neural Network is generally superior to conventional statistical models, and Support Vector Machine performs reasonably well in most prediction problems and can be used as a benchmark technique. Then, each individual model makes its own prediction estimating probabilities for each class. The final prediction of stacking is computed using multiple-linear regression as a Meta classifier.

The detailed configurations of each individual model are as follows: DTs with C4.5 search algorithm, ANNs are RBF neural network with radial basis functions as activation functions, SVMs use John Platt's sequential minimal optimization algorithm with logistic to the outputs for proper probability estimates.

3.3 Developing Intelligent Data Analysis Functions

Machine learning techniques have gradually received attention in medical informatics. Several well-known machine learning tools have already been announced, including Weka (Witten, Frank, & Hall, 2011), Yale/RapidMiner (Mierswa, Wurst, Klinkenberg, Scholz, & Euler, 2006), and Java-ML (Abeel, Y. Van de Peer, & Saeys, 2009). These open-source data mining suites provide a graphic user interface for data analysis and model visualization, but this may not be a major concern for a system developer. The IDA tool should integrate with Information System to support uninterrupted medical data analysis tasks, including interactive exploring interesting patterns and deriving useful models. Figure 2 shows the web-based Intelligent Data Analysis tool that centralizes to support prognosis task. Users can select the desired fields for building classifiers. The designed tool facilitates different types of users to access functionalities in Weka without the need to learn new data mining software.

In this study, we intended to develop custom-designed medical data analysis components and schemata, which access the data mining suite’s components through Java programming. The following functions were embedded : (1) Basic statistical functions for primary inspection of the data; (2) Visualize data exploratory functions for the interactive of interesting patterns in datasets; (3) Data sources connective and data preprocessing functions for retrieving medical data from databases, sample selection; (4) Unsupervised and supervised data analysis functions for the intelligent data analysis, includes statistical and machine learning algorithms, ex: regression analysis, clustering techniques, association rules, decision trees and support vector machines; (5) Model evaluation functions for the evaluation of built models' reliability and validity, ex: accuracy, sensitivity, specificity, ROC curves and lift chart.

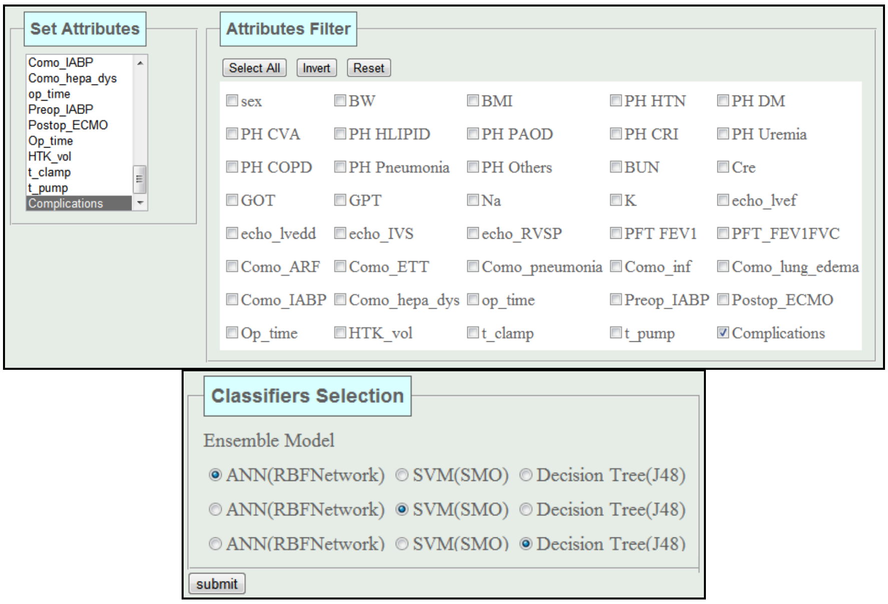
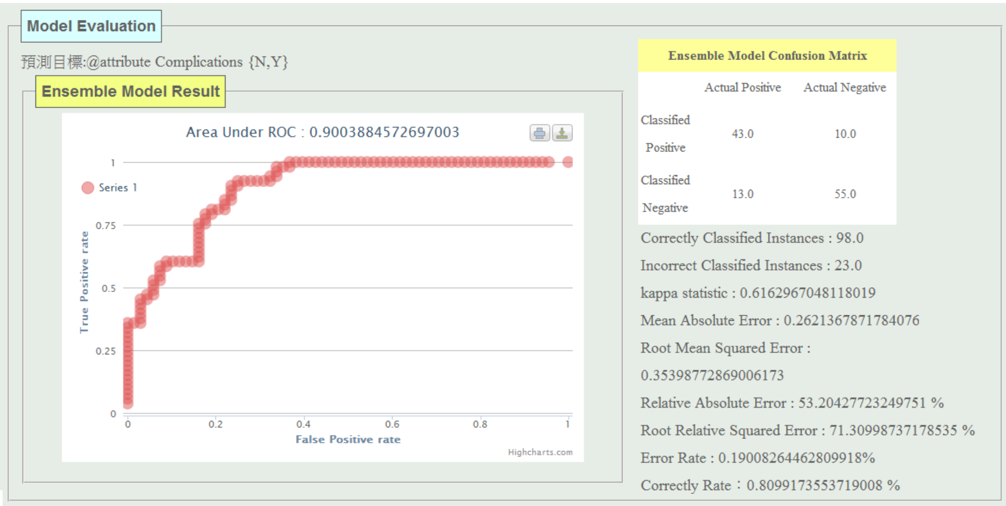


Figure 2. Attributes filter and model selection

The model selection scheme is a mixture of bagging and cross-validation (CV-bagging) that aims to improve the classification by combining models trained on randomly generated subsets of the entire training set. We first applied a cross validation scheme for model selection on each subset; subsequently, for the sake of simplicity and so as not to run into over-fitting problems, we combined the selected models in a uniform weights approach to reach a final decision. The concept of uniform weights voting is both appealing and simple to implement; it can generally be applied to any type of classifier without relying on specific interpretations of the output. In order to validate models, we used random subsets as cross-validation folds for reasons of simplicity. In k-fold cross validation, the dataset is partitioned into k subsets. Of these k subsets, a single subset is retained as the validation dataset, and the remaining k-1 subsets are used for training datasets. The cross validation process is then repeated k times with each of the k subsets used only once as the validation data. The k results from the folds can then be averaged to produce a single estimation. In cases where specific inside medical knowledge is not available, such a cross validation method can be used to select a classification method empirically, because it seems to be obvious that no classification method is uniformly superior. We trained several times for individual classifiers and select the best performance one as the final model. The rationale is if we trained each model with different initial conditions, we can find leverage performance for the final model. The result, a heterogeneous ensemble, allows classification methods to be used more effectively.

To show the performance of ensemble model, we used visualization tool, called Receiver operating characteristic (ROC) and Confusion Matrix, which is typically used for model evaluation in the field of artificial intelligence. The area under the ROC Curve is based on a non-parametric statistical sign test, the purpose of which is to estimate the probability of survival for each of a pair of patients. In this study, the area under the ROC was assessed through stratified 10-fold CV-bagging. Each row of the Confusion matrix represents the instances in a predicted class, while each column represents the instances in an actual class. The detailed accuracy of individual models and of the ensemble model is shown in Figure 3.



Models	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
10 folds CV- bagging weighted average						
DT	0.81	0.169	0.829	0.81	0.81	0.872
ANN	0.736	0.264	0.739	0.736	0.736	0.809
SVM	0.744	0.258	0.746	0.744	0.744	0.743
Ensemble	**0.81**	**0.19**	**0.812**	**0.81**	**0.81**	**0.900**

Figure 3. The results of the prediction

4. Results and Discussion

In this study, we developed a clinical information system to aid the care of cardiovascular patients. The system was integrated as a medical database system to monitor patients' status changes following the preoperative, intraoperative, and postoperative surgical stages and facilitate communication among interdisciplinary team members in the time between preoperative assessment and follow-up tracking. In Cardiovascular disease, these patients are likely to have a relatively high postoperative morbidity rate with complications, and will highly influence longer-term postoperative out-comes. It is essential to create reliable and satisfactory risk prediction models for postoperative morbidity as an aid to clinical decision-making. Although several risk prediction systems have been proposed for patients undergoing Cardiovascular disease, they basically rely on traditional statistical methods. We have proposed an ensemble model to predict postoperative morbidity and support clinical decision-making.

The system is accessible through computers that are widely deployed for clinical use in hospitals. Preliminary feedback provides useful results on tracking patient data and assessing patient risk, which gives potential gains from a wider deployment. The proposed ensemble model is constructed by Decision Tree, Artificial Neural Network and Support Vector Machine were used to augment the ensemble model and design a Web-Based Application, showing moderate performance. The experimental result shows our designed Clinical information system within Intelligent Data Analysis Tool can predict and classify the evolution of patients and facilitates different types of users to access functionalities in Weka, the tool can install and access by any client in the network. The proposed ensemble model predicts postoperative morbidity with relatively satisfactory accuracy, even when data is missing and/or sparse, showing its usefulness in support of clinical decision-making. The supplementary nature of multi-models distinguishes the proposed model from existing risk scoring systems that are based on conventional statistical methods and from various machine learning models. To summarize, these data analysis methods can indicate the reliability of stored data.

References

Abeel, T., Y. Van de Peer, & Saeys, Y. (2009). Java-ML: A Machine Learning Library. *The Journal of Machine Learning Research, 10*(Apr), 931-934.

Babaogˇlu, I., Fındık, O., & Bayrak, M. (2010). Effects of principle component analysis on assessment of coronary artery diseases using support vector machine. *Expert Systems with Applications, 37*(3), 2182-2185.

Barnes, M., Boult, M., Maddern, G., & Fitridge, R. (2008). A model to predict outcomes for Endovascular aneurysm repair using preoperative variables. *European Journal of Vascular and Endovascular Surgery, 35*(5), 571-579.

Bellazzi, R., & Zupan, B. (2008). Predictive data mining in clinical medicine: Current issues and guidelines. *International Journal of Medical Informatics, 77*(2), 81-97.

Bohm, N., Wales, L., Dunckley, M., Morgan, R., Loftus, I., & Thompson, M. (2008). Objective risk-scoring systems for repair of abdominal aortic aneurysms: applicability in Endovascular repair. *European Journal of Vascular and Endovascular Surgery, 36*(2), 172-177.

Choi, S. (2008). Detection of valvular heart disorders using wavelet packet decomposition and support vector machine. *Expert Systems with Applications, 35*(4), 1679-1687.

Delen, D., Sharda, R., & Kumar, P. (2007). Movie forecast Guru: A Web-based DSS for Hollywood managers. *Decision Support Systems, 43*(4), 1151-1570.

Eom, J.-H., Kim, S.-C., & Zhang, B.-T. (2008). AptaCDSS-E: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction. *Expert Systems with Applications, 34*(4), 2465-2479.

Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., & Euler, T. (2006). Yale: Rapid prototyping for complex data mining tasks. *The 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-06)*.

Murphy, C. K. (2001). Identifying diagnostic errors with induced decision trees. *Medical Decision Making, 21*(5), 368-375.

Polat, K., & Güne, S. (2007). Breast cancer diagnosis using least square support vector machine. *Digital Signal Processing, 17*(4), 694-701.

Shetty, S. D., S.Vadivel, & Vaghella, S. (2010). Weka based desktop data mining as web service. *World Academy of Science, Engineering and Technology 2010, 40,* 697-715.

Stijn, C. W., Wouters, L. N., Freek, W. A. V., & Rene, M. H. J. B. (2002). Preoperative prediction of early mortality and morbidity in coronary bypass surgery. *Cardiovascular Surgery, 10*(5), 500-505.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical machine learning tools and techniques* (3rd ed.). San Francisco, CA: Morgan Kaufmann.