Using Intelligence Techniques to Predict Postoperative Morbidity of Endovascular Aneurysm Repair

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

[1] Department of Information Management, National Taipei University of Nursing and Health Sciences, Taiwan, Republic of China
[2] Department of Computer Science and Information Engineering, Tamkang University, Taiwan, Republic of China
[3] Division of Cardiovascular Surgery, Department of Surgery Heart Center, Cheng-Hsin General Hospital, Taiwan, Republic of China

Abstract. Endovascular aneurysm repair (EVAR) is an advanced minimally invasive surgical technology that is helpful for reducing patients' recovery time, postoperative mortality and morbidity. This study proposes an ensemble model to predict postoperative morbidity after EVAR. The ensemble model was developed using a training set of consecutive patients who underwent EVAR between 2000 and 2009. All data required for prediction modeling, including patient demographics, preoperative, co-morbidities, and complication as outcome variables, was collected prospectively and entered into a clinical database. A discretization approach was used to categorize numerical values into informative feature space. The research outcomes consisted of an ensemble model to predict postoperative morbidity, the occurrence of postoperative complications prospectively recorded, and the causal-effect decision rules. The probabilities of complication calculated by the model were compared to the actual occurrence of complications and a receiver operating characteristic (ROC) curve was used to evaluate the accuracy of postoperative morbidity prediction. In this series, the ensemble of Bayesian network (BN), artificial neural network (ANN) and support vector machine (SVM) models offered satisfactory performance in predicting postoperative morbidity after EVAR.

Keywords: Endovascular aneurysm repair (EVAR), postoperative morbidity, ensemble model, machine learning.

1 Introduction

Aortic surgery is a complex surgical operation that is indicated for patients with severe insufficiency in cardiac function. 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. During the operation and the postoperative stay at the ICU and nursing ward, there is considerable morbidity for aortic surgery patients with postoperative complications, which results in increased hospital mortality and postoperative morbidity. Many prediction models for

N.T. Nguyen, C.-G. Kim, and A. Janiak (Eds.): ACIIDS 2011, LNAI 6591, pp. 197–206, 2011.

cardiac surgical outcome apply logistic or multivariable regression to assess preoperative risk [1-3]. Most of the risk prediction models in current use were derived for patients undergoing open abdominal aortic aneurysm (AAA) repair and appear to lack utility when applied to EVAR patients. 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 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. EVAR is an advanced minimally invasive surgical technology that helps reduce patients' recovery time as well as postoperative mortality and morbidity[4]; it is especially helpful in the treatment of patients judged to be high surgical risk for conventional surgery. EVAR benefits patients with medical co-morbidities, and postoperative complications highly significantly influence longer-term postoperative outcomes in EVAR patients.

Data mining techniques are currently used in medical decision support to increase diagnostic accuracy and to provide additional knowledge to medical stuff. Their increased use provides expanded opportunities for determining the utility of medical decision-making models from retrospective data. Compared to data mining for business applications, medical data mining includes prediction data mining and descriptive data mining two distinct concepts. That is, medical data mining not only requires a prediction model with satisfactory accuracy, but also requires a safety context in which decision-making activities require explanatory support. The main distinction is that predictive data mining requires that the training dataset include an outcome variable, while descriptive data mining uses a global strategy to find the characteristics of each affinity granulation of the data. Both these data mining techniques can produce accurate, predictive and interpretable descriptive models that contribute greatly to handling medical data gathered through systematic use of clinical, laboratory, and hospital information systems. The goal of predictive data mining in clinical surgery is to derive models that can use medical data to predict patient's mortality and morbidity and thereby support clinical surgical decision-making. Predictive data mining can also aid in prognosis, diagnosis and treatment planning for surgical procedures. In contrast, descriptive data mining considers the data as affinity granulations, and aims at finding interpretable patterns and associations among data.

The use of machine learning models has become widely accepted in medical applications. Various machine learning models including BNs, ANNs, and SVMs have been tested in a wide variety of clinical and medical applications[5]. Soft-computing, including fuzzy set and rough set techniques that work well in descriptive data mining, is also a promising technique. BN is a probability-based inference model that has a wide range of applications and is increasingly used medically as a prediction and knowledge representation modeling technique. Verduijn, M., et al. [6] presented the prognostic BN as a new type of prognostic model that builds on the BN methodology and implements a dynamic, process-oriented view of cardiac surgical prognosis. Lin and Haug [7] proposed BN suitable for exploiting missing clinical data for medical prediction. ANNs have featured in a wide range of medical applications, often with promising results. Eom et al. [8] developed a classifier ensemble-based, including ANNs, DTs, and SVMs, clinical decision support system for predicting cardiovascular disease level. SVMs have been successfully used in a wide variety of medical applications. Polat and Güne [9] used a least square support vector machine to assist breast cancer diagnosis. Babaoĝlu et al. [10] 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. Choi [11] proposed the detection of valvular heart disorder (VHD) by wavelet packet decomposition and SVM techniques.

This study describes the development of an informative ensemble prediction model consisting of BNs, ANNs and SVMs for the prediction of postoperative morbidity between preoperative variables and complication outcomes in EVAR patients. For a better understanding of our study, Section 2 of this paper begins with an overview of study background and experimental methods in general. Section 3 describes the experimental design and procedures used in this study, including entropy/MDL-based method for the discretization of numerical features, feature selection method, 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 Study Background and Materials

Abdominal aneurysm (AAA) is an enlargement that occurs in a weakened area within the largest artery in the abdomen (http://www.vascularweb.org/patients/NorthPoint/Abdominal_Aortic_Aneurysm.html). If an AAA is not treated in due time, the pressure generated by heartbeats causes the aneurysm to continuously grow larger, and the aortic wall continues to weaken. Finally, rupture occurs and massive internal bleeding occurs. The best way to prevent the high mortality associated with AAA is to find the lesion before rupture occurs. However, patients with aortic diseases are often elderly with severe co-morbidities and sometimes devastating morbidity, making them extremely challenging candidates for surgery. For such patients, EVAR represents a lower risk approach than conventional open surgery and is associated with shorter operating times, shorter hospitalizations, more rapid recovery and improved quality of life during the perioperative period and postoperative follow-up. Although long-term data on the clinical outcomes of patients who received EVAR are not yet available, given its importance, building a prediction of postoperative morbidity after EVAR is critical.

We retrospectively examined 140 consecutive patients who underwent EVAR surgery at Taipei Veteran General Hospital, a teaching center hospital in Taiwan, between 2000 and 2009. The dataset contains preoperative patient characteristics, details of the operative information, and pathological and laboratory findings from the general ward, operating room and intensive care unit (ICU). The dataset also included length of ICU stay, variables that describe postoperative complications that frequently

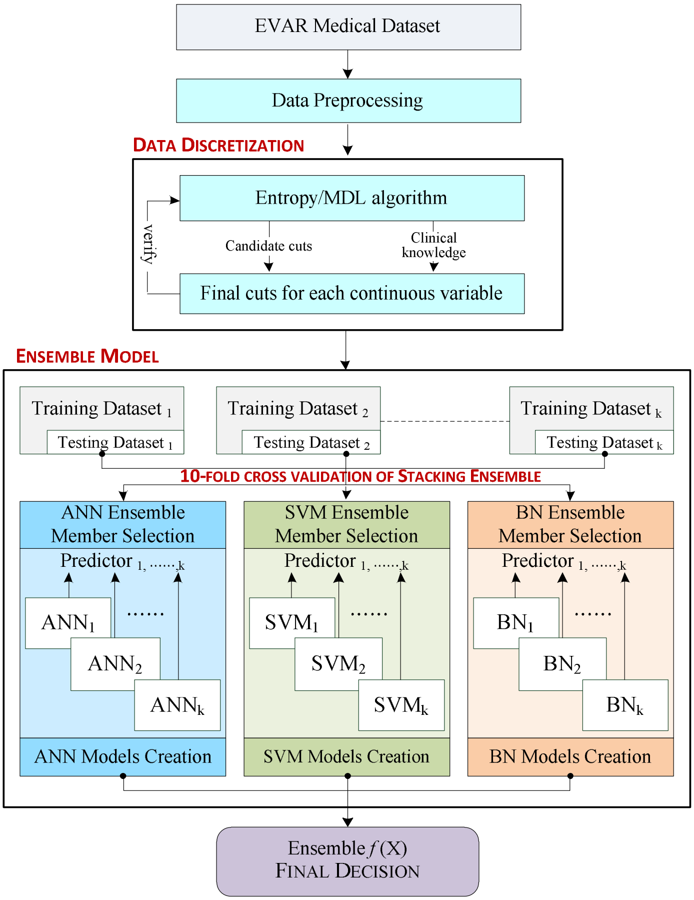


Fig. 1. Proposed architecture for ensemble model development

occur in EVAR 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 137 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. Sixty-seven of the 140 patients experienced postoperative complications during their stay at hospital. Data collected included preoperative patient characteristics, risk factors, details of the operative information, and physical characteristics of the aneurysm, postoperative physiological and laboratory findings, and postoperative complications as the outcome variable.

The development of this prediction model proceeds as follows Fig.1. First, an informative discretization method for numerical values, which employs a data discretization method to guide the categorized numerical features on the basis of entropy/MDL algorithm and laboratory surgeon' s knowledge, is developed. Through data discretization, numerical values are converted into discrete values. Second, the proposed ensemble-based architecture focuses on fusing three types of models. BNs, ANNs, and SVMs). During the training process, the same training data set is used for all individual models in order to reduce the diversity among individual models, keeping in mind that, in an ensemble model, it is important to construct appropriate training data sets that maintain good balance between accuracy and diversity among individual models. The model selection scheme, designated a stacking scheme, is a mixture of stacking and cross-validation that is chosen in order to improve overall classification by combining models trained on randomly generated subsets of the entire training set.

3 Experimental Design

3.1 Entropy/MDL-Based Method for the Discretization of Numerical Values

Most studies dealing with cardiac surgery prediction models have applied logistic or multivariable regression to assess the preoperative risk. Few studies have utilized machine learning algorithms i.e., decision trees, Bayesian networks or artificial neural networks in analyzing clinical data. These state-of-the-art machine learning algorithms are often informative and can represent more knowledge in the clinical data. Generally, these algorithms require discrete categorical values but clinical datasets usually involve numerical variables. To satisfy the requirements of machine learning algorithms, the employment of a discretization approach is necessary. Discretization is defined as a process that divides numerical values into states of discrete categorical values, leading to informative expressed categorical values. For example, the CART model originally was not designed to handle numerical attributes. During the construction of a CART model, numerical attributes were divided into discrete categorical values. Aside from CART, discretization techniques were frequently adopted in other popular learning paradigms, such as C4.5, BNs, ANNs, and genetic algorithms.

This study employed entropy/MDL to divide numerical domains into intervals. Incorporation investigates data discretization; the intervals are characterized by discrete categorical values. The steps for automatically finding discrete categorical values from a given dataset are described herein. Assuming that the domain of a numerical attribute ranges from v1 to v2, and that {c1, c2,…,ck} denote the k cut-points obtained by the entropy/MDL algorithm, using these k cut-points, k discretize values can be determined. For example, AAA_Size is a numerical attribute obtained from the dataset. The domain of AAA_Size ranges from 5.0 to 9.6. Six cut-points, (5.5, 5.7, 5.8, 7.3, 8.1, 9.6), were computed using the entropy/MDL algorithm. As depicted in Fig. 2, six discrete categorical values were obtained. Within each analysis, the laboratory system should contain an interval that delimits life-compatible values. Therefore, if cut

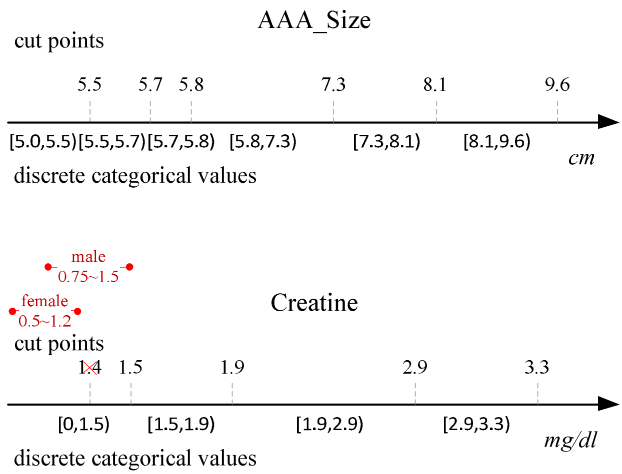


Fig. 2. The discrete categorical values of AAA_Size and Creatine

points fall within the interval, they should be eliminated. For example, one cut point {1.4 mg/dl} of creatine lay within an interval, i.e., [0.75~1.5] for male, [0.5~1.2] for female; therefore the cut point {1.4 mg/dl} was eliminated.

3.2 Significant Attribute Selection

Feature selection methods can be carried out by attribute ranking methods and attribute subset evaluation methods [12]. The attribute ranking methods assess the goodness of individual variables for prediction independently of other attributes and only the high ranked attributes are used to build the prediction model. For practical applications, the former methods can be used as an initial screening to reduce dimensionality in highly dimension datasets. The later methods can be used to find more relevant subset of attributes simultaneously. Hall and Holmes [12] compared six feature selection methods and benchmarking attribute selection techniques for discrete class data mining. The best performing methods according to their study were Information Gain, Recursive Elimination of Features (Relief), Correlation-based Feature Selection (CFS), Consistency-based Subset Evaluation, and Wrapper Subset Evaluation. The former two are attribute ranking methods and the latter three are attribute subset evaluation methods. The results of this benchmark study provide guidelines for the choice of feature selection methods. This search property is in favor of small feature subsets with high class consistency which is more proper for clinical application.

This study used entropy/MDL to discrete numeric attributes. The attributes are then ranked according to their overall contribution to the consistency of the attribute set. As shown in Table 1, we selected attributes by three attribute subset evaluation algorithms. The obtained attribute subsets were employed as major input variables for training ensemble prediction models based on StackingC algorithm.

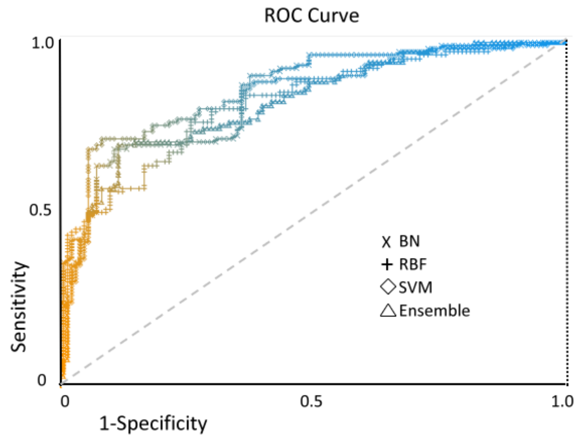
Table 1. Selected attributes by attribute subset evaluation algorithms

Variable	Definition	CFS	Consistency	Wrapper (C4.5)
Gender	Male/female			●
Age	0-100 years			
Smoking	No smoking, smoking, uncertain		●	●
Hypertension	Yes/no		●	
DM	Diabetes mellitus, yes/no	●	●	●
Hyperlipidemia	Yes/no		●	●
COPD	Chronic obstructive pulmonary disease, yes/no		●	●
CVA	Cerebral vascular accident, yes/no	●		
Heart Disease	Yes/no	●	●	
CRI	Chronic renal insufficiency, yes/no	●		
Hgb	Hemoglobin level, g/dl	●	●	●
Hct	Hematocrit percentage, 0-100%	●		
PLT	Platelet count, /CUMM	●	●	●
BUN	Blood urea nitrogen level, mg/dl	●	●	
Creatine	Creatine level, mg/dl	●	●	
AAA size	Size of abdominal aortic aneurysm, mm		●	
AAA site	Site of abdominal aortic aneurysm			●

3.3 Ensemble Model for the Prediction of Postoperative Morbidity

In this study, BNs, ANNs, and SVMs were chosen as based models 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 BNs can easily model complex relationships among variables, ANNs are generally superior to conventional statistical models, and SVMs perform 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. For the implementation of models, we chose BayesNet, MultilayerPerceptron and SMO as base models in WEKA, and StackingC as ensemble method [13]. All the default parameters in WEKA were used. Besides, we used 10-fold cross-validation to lower the variability of training set.

The model selection scheme is a mixture of stacking [14] and cross-validation 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



Models	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
10 folds CV- bagging weighted average						
BN	0.736	0.270	0.737	0.736	0.734	0.834
NN	0.675	0.330	0.657	0.657	0.674	0.755
SVM	0.757	0.247	0.758	0.757	0.756	0.755
Ensemble	0.757	0.249	0.760	0.757	0.755	0.813

Fig. 3. The results of the postoperative morbidity prediction

into over-fitting problems, we combined the selected models by stacking approach to reach a final prediction. 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 [15]. The result, a heterogeneous ensemble, allows classification methods to be used more effectively. The detailed accuracy of individual models and of the ensemble model is shown in Fig 3. The results of the experiment show that individual models' performance and improvements in performance were achieved by applying the ensemble of models. This indicates that model combination techniques indeed yield stable performance improvements for ensemble models.

4 Discussions

In comparison to open repair, endovascular aneurysm repair of abdominal aortic aneurysms provides a lower risk approach associated with a shorter operating time, shorter hospitalization stay, more rapid recovery time and improved quality of life during the perioperative period and postoperative follow-up. Identification of the reasons for postoperative complications and risk factors for re-intervention during follow-up to maintain aneurysm exclusion remain challenges for surgeons. Most surgeons perform EVAR on sicker patients; however, patients with aneurysm anatomy are usually elderly and might be considered marginal for EVAR repair. These patients are likely to have a relatively high postoperative morbidity rate with complications, and will highly influence longer-term postoperative outcomes. 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 open aneurysm repair, they basically rely on traditional statistical methods and provide scant accuracy and utility when applied to EVAR patients.

We have proposed an ensemble model to predict postoperative morbidity after EVAR and support clinical decision-making. The proposed ensemble model is constructed by incorporating discretization of categorical numerical values; BNs, ANNs, and SVMs were used to augment the ensemble model and the dataset was processed by cross validation, showing moderate performance. The experimental result shows that 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 distinguish the proposed model from existing risk scoring systems that are based on conventional statistical methods and from various machine learning models. To summarize, the advantage of using the proposed ensemble model is that it can provide surgeons with practical, relatively accurate aid in their daily diagnostic tasks and enable them to extract meaningful relationships among features of medical datasets through the use of constrained decision rules.

References

1. Barnes, M., Boult, M., Maddern, G., Fitridge, R.: A model to predict outcomes for Endovascular aneurysm repair using preoperative variables. European Journal of Vascular and Endovascular Surgery 35, 571–579 (2008)
2. Bohm, N., Wales, L., Dunckley, M., Morgan, R., Loftus, I., Thompson, M.: Objective risk-scoring systems for repair of abdominal aortic aneurysms: applicability in Endovascular repair. European Journal of Vascular and Endovascular Surgery 36, 172–177 (2008)
3. Stijn, C.W., Wouters, L.N., Freek, W.A.V., Rene, M.H.J.B.: Preoperative prediction of early mortality and morbidity in coronary bypass surgery. Cardiovascular Surgery 10, 500–505 (2002)
4. Bush, R.L., Johnson, M.L., Hedayati, N., Henderson, W.G., Lin, P.H., Lumsden, A.B.: Performance of endovascular aortic aneurysm repair in high-risk patients: Results from the Veterans Affairs National Surgical Quality Improvement Program. Journal of Vascular Surgery 45(2), 227–235 (2007)
5. Roques, F., Michel, P., Goldstone, A.R., Nashef, S.A.: The logistic EuroSCORE. European Heart Journal 24, 1–2 (2003)
6. Bellazzi, R., Zupan, B.: Predictive data mining in clinical medicine: Current issues and guidelines. International Journal of Medical Informatics 77, 81–97 (2008)
7. Verduijn, M., Peek, N., Rosseel, P.M.J., Jonge, E.d., Mol, B.A.J.M.d.: Prognostic Bayesian networks I: Rationale, learning procedure and clinical use. Journal of Biomedical Informatics 40, 609–618 (2007)
8. Lin, J.H., Haug, J.: Exploiting missing clinical data in Bayesian network modeling for predicting medical problems. Journal of Biomedical Informatics 41, 1–14 (2008)

9. Rowan, M., Ryan, T., Hegarty, F., O'Hare, N.: The use of artificial neural networks to stratify the length of stay of cardiac patients based on preoperative and initial postoperative factors. Artificial Intelligence in Medicine 40, 211–221 (2007)
10. Lisboa, P.J., Taktak, A.F.G.: The use of artificial neural networks in decision support in cancer: A systematic review. Neural Networks 19, 408–415 (2006)
11. Dokur, Z.: A unified framework for image compression and segmentation by using an incremental neural network. Expert Systems with Applications 34, 611–619 (2008)
12. Hall, M.A., Holmes, G.: Benchmarking attribute selection techniques for discrete class data mining. IEEE Transactions on Knowledge and Data Engineering 15, 1–16 (2003)
13. Witten, I.H., Frank, E.: Data mining: Practical machine learning tools and techniques. Morgan Kaufmann Publishers, Boston (2005)
14. Seewald, A.: How to make stacking better and faster while also taking care of an unknown weakness. In: Proceedings of the 19th International Conference on Machine Learning, pp. 554–561 (2002)
15. Quinlan, J.R.: Comparing connectionist and symbolic learning methods. In: Computational Learning Theory and Natural Learning Systems, vol. 1. MIT Press, Cambridge (1994)