Türk Göğüs Kalp Damar Cerrahisi Dergisi 2013;21(3):574-580
doi: 10.5606/tgkdc.dergisi.2013.7680

Original Article / Özgün Makale

Risk group classification for bleeding after coronary artery bypass graft surgery: a comparison of the logistic regression with decision tree models

Koroner arter baypas greft cerrahisi sonrasında kanama açısından riskli grupların sınıflandırılması: Lojistik regresyon ve karar ağacı modellerinin karşılaştırılması

Reza Safiarian,[1] Payam Amini,[2] Elham Khodayari Moez,[2] Fatemeh Mohammadzadeh,[2] Mohammad Tavakoli,[3] Farid Zayeri[4]

[1]Baqiyatallah University of Medical Science, Tehran, Iran
[2]School of Medical Sciences, Tarbiat Modares University, Tehran, Iran
[3]Ministry of Health Treatments and Medical Education, Tehran, Iran
[4]Proteomics Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Background: This study aims to identify high-risk patient groups for bleeding after coronary artery bypass graft (CABG) surgery.

Methods: We retrospectively evaluated 205 patients (143 males, 62 females; mean age 59.7±10.1 years; range, 28 to 83 years) undergoing CABG surgery between June 2001 and August 2008 at Jamaran Heart Hospital, Tehran, Iran. Baseline characteristics of the patients and postoperative bleeding status were recorded. For classifying the bleeders and non-bleeders, classic logistic regression and decision tree models were utilized.

Results: Logistic regression analysis showed that sex was significantly related to postoperative bleeding. Decision tree model revealed that age (score= 100), diabetes mellitus (score= 16.38), sex (score= 13.67), capital residency (score= 7.31) and dyslipidemia (score= 5.06) were found to have an impact on bleeding. We also observed that the decision tree model provided a better classification of the patients than logistic regression.

Conclusion: Surgeons should be aware of risk indicators of bleeding such as older age, male sex, absence of diabetes mellitus and presence of dyslipidemia in patients with three bypassed vessels before CABG. We also recommend statisticians to utilize the decision tree model instead of logistic regression analysis in classification of risk groups.

Key words: Bleeding; classification; coronary artery bypass graft; decision tree; logistic regression.

Amaç: Bu çalışmada koroner arter baypas greft (KABG) cerrahisi sonrasında kanama açısından yüksek riskli hasta grupları belirlendi.

Çalışma planı: Haziran 2001 - Ağustos 2008 tarihleri arasında İran Tahran Jamaran Kalp Hastanesi'nde KABG cerrahisi yapılan 205 hasta (143 erkek, 62 kadın; ort. yaş 59.7±10.1 yıl; dağılım 28-83 yıl) retrospektif olarak değerlendirildi. Hastaların başlangıçtaki özellikleri ve ameliyat sonrası kanama durumları kaydedildi. Kanama olan ve kanama olmayan hastaları sınıflandırırken, klasik lojistik regresyon ve karar ağacı modelleri kullanıldı.

Bulgular: Lojistik regresyon analizinde cinsiyetin ameliyat sonrası kanama ile anlamlı düzeyde ilişkili olduğu görüldü. Karar ağacı modelinde ise, yaş (skor= 100), diabetes mellitus (skor= 16.38), cinsiyet (skor= 13.67), başkentte ikamet etme (skor= 7.31) ve dislipideminin (skor= 5.06) kanama üzerinde etkisi olduğu belirlendi. Lojistik regresyona kıyasla, karar ağacı modelinde hastaların daha iyi sınıflandırıldığı da gözlendi.

Sonuç: Cerrahlar KABG öncesinde üç damarına baypas yapılan hastalarda ileri yaş, erkek cinsiyeti, diabetes mellitus yokluğu ve dislipidemi varlığı gibi kanamanın risk göstergelerini göz önünde bulundurmalıdır. Ayrıca, istatistik uzmanlarına risk grubu sınıflandırmasında lojistik regresyon analizinin yerine karar ağacı modelini kullanmalarını önermekteyiz.

Anahtar sözcükler: Kanama; sınıflandırma; koroner arter baypas greft; karar ağacı; lojistik regresyon.

An association between bleeding after coronary artery bypass graft (CABG) surgery and postoperative mortality and morbidity has been well proven in many previous studies.[1-4] In addition, a higher risk of an adverse outcome is predicted for patients who undergo reoperations because of bleeding after a CABG operation.[2] This risk increases in proportion to the time that elapses before the reoperation. Furthermore, other

Available online at
www.tgkdc.dergisi.org
doi: 10.5606/tgkdc.dergisi.2013.7680
QR (Quick Response) Code

Received: September 10, 2012 *Accepted:* December 09, 2012

Correspondence: Farid Zayeri, M.D. Proteomics Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, 1971653313, Tehran, Iran.
Tel: +98-912-527 74 88 e-mail: fzayeri@yahoo.com

complications such as stroke, renal failure, transfusion reactions, infections, and cardiac tamponade are expected among patients for whom the reoperation is delayed.[1,2] These patients usually need prolonged intensive care and hospital stays, which increases the cost of treatment.[4] Much research has been conducted to identify potential changes during cardiac surgery that could decrease the rate of bleeding among patients undergoing CABG[2,4] since on-pump CABG and longer cardiopulmonary bypass (CPB) duration affects platelet function. Because of the decreased incidence of clotting abnormalities, off-pump CABG has become more popular.[1,3]

Classification assigns a new object to a specific class from a given set of classes and can be performed by using various methods like data mining.[5] For example, classifying patients in order to predict periventricular leukomalacia by employing monitoring variables such as heart rate, blood pressure, and central venous filling pressure[6] helps to identify patients who are prone to coronary heart disease because of risk factors such as smoking, blood pressure, and total cholesterol levels.[7] Among the many classification methods, decision trees (DTs), Bayesian networks (BNs), k-nearest neighbor (k-NN) algorithms, and fuzzy logic are the most commonly applied approaches.[5] Aside from evaluating the predictive power of these methods, knowing the ease with which the results can be interpreted is vital when choosing the appropriate analytical method. With this in mind, the DT classifier is a common choice among data analysts.[8-10] This procedure is preferable when the subjects are described through a predetermined set of attributes, the target variable is discrete, disjunctive results are required, or the data contains errors and missing attribute values.[11] In addition to the modern analytical techniques such as DT, regression analysis is one of the traditional statistical methods which can be applied to mark the most influential variables.[9] Predicting the presence or absence of an attribute by means of some predictors is a widely known application of the logistic regression (LR) model.[12]

Our objective was to identify the high-risk groups for bleeding after on-pump CABG according to the preoperative characteristics of the patients in order to give these patients the extra care they require. Therefore, we compared LR and DT, the two most popular classification methods.

PATIENTS AND METHODS

In this study, we analyzed the data from 205 patients (143 males, 62 females; mean age 59.7±10.1 years; range 28 to 83 years) who underwent CABG surgery between June 2001 and August 2008 at Jamaran Heart Hospital in Tehran, Iran. Since reoperation after CABG surgery is rare (2-6%),[3] we randomly selected 40 patients from the reoperated subjects for the cohort, and the remainder were selected from non-reoperated cases in order to have valid inferences.[13,14]

Reoperation took place in these patients as a result of bleeding after initial departure from the operating room. The decision criteria for the reoperation were as follows: *(i)* drainage of more than 500 mL during the first hour, more than 400 mL during each of the first two hours, more than 300 mL during each of the first three hours, or a total of more than 1000 mL in the first four hours; *(ii)* sudden massive bleeding; *(iii)* obvious signs of cardiac tamponade; *(iv)* excess bleeding despite correction of coagulopathies; and *(v)* cardiac arrest in a patient who continued to bleed.[15]

In order to eliminate any bias in the estimation, we applied the following restrictions on the sampling design: *(i)* patients with a history of abnormal coagulation were excluded; *(ii)* the CABG operations were performed by the same surgeon; and *(iii)* patients with three bypassed vessels were eligible for this study.

In this archival study, bleeding after CABG was treated as a binary response variable (0= patients who did not experience bleeding, 1= patients who bled postoperatively). Patients' preoperative characteristics such as gender, age at the time of surgery, dyslipidemia, diabetes mellitus (DM), and hypertension (HT) were considered as predictors (independent variables) along with whether they were a resident of Tehran [capital resident (CR)]. This information was obtained from the patients' medical files at the hospital. To find the variables that affected bleeding after CABG, we applied both the LR model and the DT to the data. The details are shown in the following sections.

Details of statistical analysis

Logistic regression (LR): Logistic regression is a common statistical tool for predicting a binary outcome. In this model, the relationship between the independent variables and the dependent variable is a logit function (the natural logarithm of odds), not a linear one. The model can be written as:

$$\log\left(\frac{\text{pr}(Y = 1|X_1,X_2,\ldots,X_p)}{(1 - \text{pr}(Y = 1|X_1,X_2,\ldots,X_p)}\right) = \alpha + \beta_1X_1 + \beta_2X_2 + \ldots + \beta_pX_p + \varepsilon$$

In this model, X1,X2,…,Xp are the independent variables, and β1,β2,...,βp are the regression parameters which should be estimated through the data.[16,17]

Decision tree (DT): The DT is one of the most popular classification methods as it can be applied to many medical diagnosis problems.[18] In the majority of cases in which the aim of the research is to identify or discriminate high-risk subjects, the DT is an excellent analytical choice.[19] It involves three basic components: decision nodes, branches, and leaves. The path begins at the decision node and extends to the leaf. This corresponds to a conjunction of test features. The tree can be considered as a disjunction of these conjunctions,[18] and these disjunctions function to separate the branch population into groups with a similar likelihood of events. At each branching stage, the set of disjunctions causes the highest possible predictive power. This method provides the graphic feature of choices which allows one to find alternatives for each decision and possible outcome and to compare the different alternatives.[19] Several algorithms have been introduced to construct a decision tree, such as classification and regression trees (CARTs).[20]

Comparison: In addition, the Hosmer-Lemeshow test evaluates the adequacy of LR by using indices such as sensitivity, specificity, diagnostic accuracy (DA), positive predictive value (PPV) and negative predictive value (NPV) to determine the accuracy of the methods. To clarify the results of our study, a receiver operating characteristic (ROC) curve was plotted for the both the LR model and the DT. The area under the ROC curve (AUC) was calculated as a measure of discrimination, and McNemar's test was used to evaluate the differences in proportions between the methods. To find the association between the observed and predicted values in both methods, measures such as the ∅ coefficient, contingency coefficient, and Kendall tau-b correlation coefficient were calculated.[21] The CART® version 6.0 (Salford Systems, San Diego, CA, USA) and IBM SPSS Statistics version 19.0 (IBM Corporation, Armonk, NY, USA) software programs were then used to analyze the data.

RESULTS

About 50% of the patients were CRs, and the majority of these were either nondiabetic (72.7%), nondyslipidemic (63.4%), or nonhypertensive (68.8%). Forty patients (19.5%) experienced bleeding after the CABG surgery. To identify the clinical indicators for this bleeding, we used both the LR model and the DT to analyze the data.

The test sample was composed of 22 randomly selected patients. The remaining 173 subjects made up the learning sample, and these were classified using the DT method. The result derived from the learning sample was then evaluated by utilizing the test sample.

Gender was the only significant variable in the LR model (Table 1), but age (score= 100), DM (score= 16.38), gender (score= 13.67), CR (score= 7.31), and dyslipidemia (score= 5.06) were significant variables in the decision tree analysis. According to the results of LR, the odds of bleeding for men were 2.57 times higher than for women, and the Hosmer-Lemeshow test showed a good fit for the LR model (p=0.524).

In Figure 1, each node shows the probability of bleeding for patients who met the conditions mentioned on the corresponding branches. For example, the probability of bleeding for female patients who are younger than 66.5 was 0.07.

The 14 rules extracted from the DT are shown in Table 2. Our data revealed that regardless of residency and diabetes status, 7.4% of women younger than 66.5 of age experienced postoperative bleeding. Additionally, male diabetics older than 53.5 years old who were not CRs had a 34% probability of bleeding after CABG surgery.

Regarding the higher sensitivity, specificity, DA, NPV, PPV, and AUC, a comparison was made concerning the outperformance of the DT versus the LR model (Table 3). We depicted the behavior of the methods through their ROC curves, and for the DT, it exhibited higher specificity and sensitivity when measured against the LR model (Figure 2).

Table 1. Logistic regression results for assessing the effect of different risk factors on bleeding

Variable	Estimate	SE	*p*	OR	95% CI for OR
Gender	0.946	0.476	0.04	2.574	1.012-6.549
Age	0.005	0.020	0.80	1.005	0.967-1.044
Dyslipidemia	-0.005	0.393	0.98	0.995	0.461-2.147
Diabetes mellitus	-0.648	0.400	0.10	0.523	0.239-1.146
Hypertension	-0.084	0.437	0.84	0.919	0.390-2.165
Capital resident	0.668	0.379	0.07	1.951	0.927-4.103

SE: Standard error; OR: Odds ratio; CI: Confidence interval.

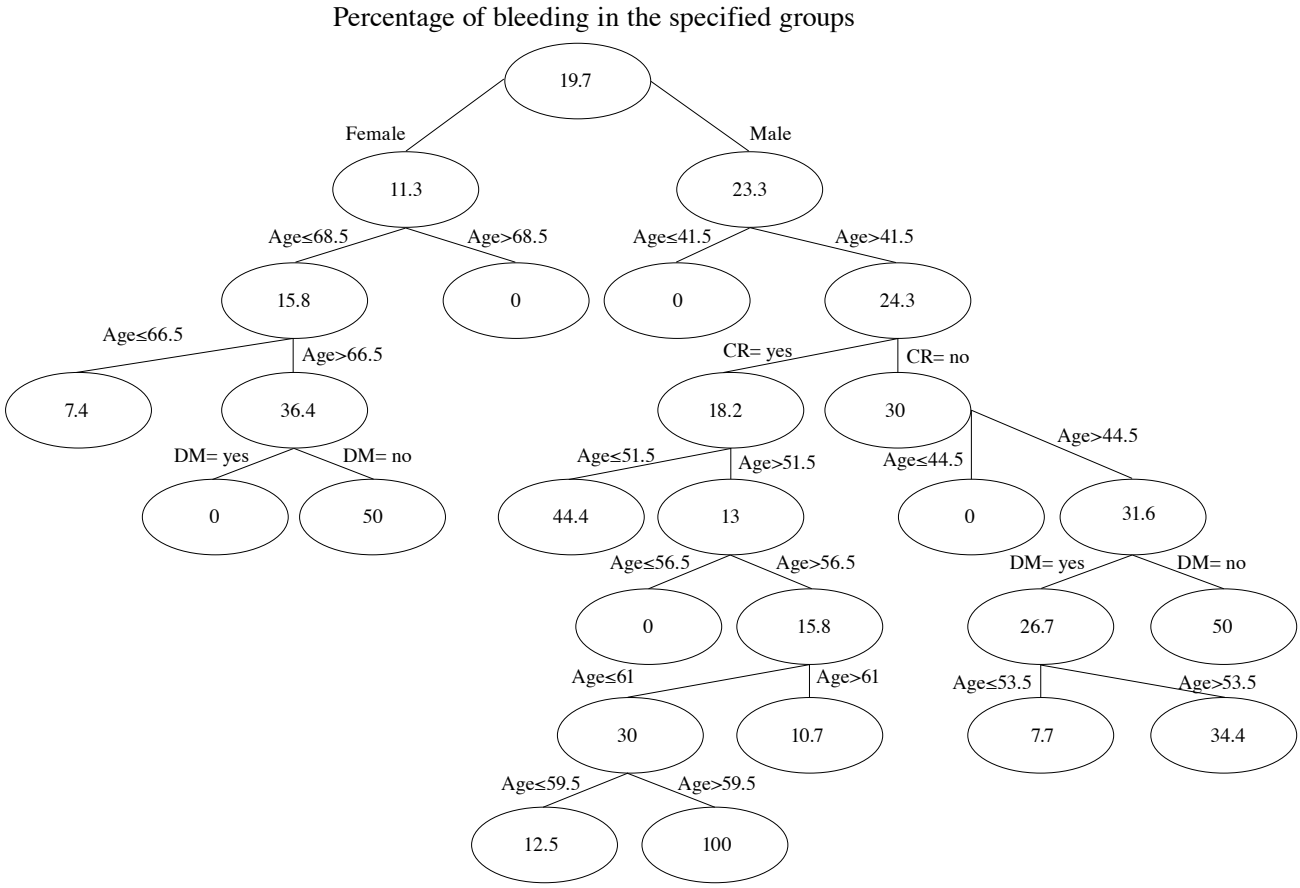

Figure 1. The classification tree model. DM: Diabetes mellitus; CR: Capital resident.

We also used McNemar's test and association measures to compare the diagnostic accuracy between the fitted models (Table 4), and this strongly confirmed the significant difference between the two methods. The ∅ coefficient, contingency coefficient, and Kendall tau-b between the observed and predicted values were 0.174, 0.172, and 0.174 for LR and 0.438, 0.401, and 0.438 for the DT, respectively. Obviously,

Table 2. Risk group classification results for bleeding using the decision tree analysis

	Variables				
Rules	Gender	Age	CR	DM	Bleeding (%)
1	Female	≤66.5	–	–	7.4
2	Female	(66.5-68.5)	–	Positive	0
3	Female	(66.5-68.5)	–	Negative	50
4	Female	>68.5	–	–	0
5	Male	≤41.5	–	–	0
6	Male	(41.5-51.5)	Positive	–	44.4
7	Male	(51.5-56.5)	Positive	–	0
8	Male	(56.5-59.5)	Positive	–	12.5
9	Male	(59.5-61.0)	Positive	–	100
10	Male	>61.0	Positive	–	10.7
11	Male	(41.5-44.5)	Negative	–	0
12	Male	(44.5-53.5)	Negative	Positive	7.7
13	Male	>53.5	Negative	Positive	34.4
14	Male	>44.5	Negative	Negative	50

CR: Capital resident; DM: Diabetes mellitus.

Table 3. Diagnostic values of the decision tree and logistic regression models

Method	Sensitivity	Specificity	PPV	NPV	Accuracy	AUC
Decision tree						
Learning sample	0.79	0.74	0.42	0.93	0.75	0.82
Test sample	0.83	0.69	0.38	0.94	0.71	0.83
Logistic regression	0.62	0.60	0.27	0.86	0.60	0.65

PPV: Positive predictive value; NPV: Negative predictive value; AUC: Area under ROC curve.

the higher correlation signified the method with less misclassification.

DISCUSSION

The shared results of both the DT and LR indicated that men are more prone to bleeding after the CABG surgery. The same result can be found in the study by Mehta et al.,[22] who proved that bleeding in men is 1.39 times more likely than in women. However, when assessing the risk factors of reexploration caused by hemorrhage in CABG patients in 1998, Dacey et al.[3] found that gender was not a significant variable.

We found that the likelihood of post-CABG bleeding was significantly influenced by the age of the patient. In the aforementioned study by Mehta et al.,[22] bleeding in patients over the age of 60 was 1.02 times more probable than for other patients. Choong et al.[2] Dacey et al.[3] and Al-Fayes et al.[4] also determined that increased age was a significant risk factor for bleeding.

According to our results, diabetics were less likely to experience bleeding. This may be because their rate of blood perfusion is less than for non-diabetics.[23] Mehta et al.[22] examined the possibility of DM being a risk factor for bleeding in CABG patients and found that non-diabetics were 1.16 times more likely to have bleeding than diabetics in their study of 528,686 patients.

In this study, HT was included in the models, but bleeding was not affected by this variable. Although Mehta et al.[22] found HT to be a significant risk factor for bleeding after CABG surgery, Choong et al.[2] concluded, just as we did, that the effect of HT was insignificant.

In addition, we found few studies in the literature which considered dyslipidemia to be a risk factor for bleeding after CABG surgery. We also found it to be an insignificant indicator, and Mehta et al.[22] also arrived at the same conclusion.

To identify the relationships between geographic status and clinical outcomes following CABG surgery, Dao et al.[24] concluded that rural patients experience longer hospital stays as well as higher in-hospital mortality rates. When we took into account geographic status as a risk factor for an adverse outcome after

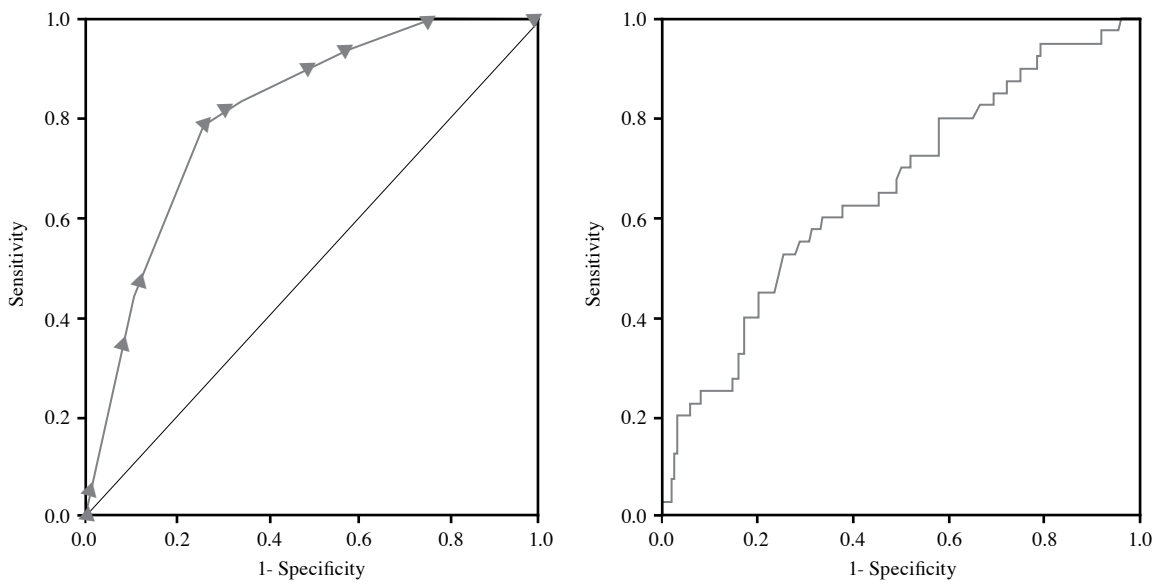

Figure 2. Receiver operating characteristic curves for decision tree (left) and logistic regression (right) predictions.

Table 4. Assessment of the correlation between the decision tree and logistic regression classification results

	Decision tree classification		
	Bleeding	Non-bleeding	*Total*
Logistic regression classification			
Bleeding	25	67	92
Non-bleeding	15	98	113
Total	40	165	205

surgery, our data showed that patients living in Tehran experienced less bleeding than those living in other locations. Patients in the capital have more access to progressive medical equipment, special physicians, and health coverage, thus they are followed up regularly and start their treatment at the beginning stages of the disease. Consequently, fewer complications, for example mortality and morbidity after a surgery, are seen.

In order to classify the patients who underwent CABG into high-risk and low-risk categories, we compared the LR and DT classification models. The DT classification provides a rapid and effective method for determining categories and can be applied to a wide range of issues.[10,25] As previous research has proven, DTs have the ability to cope with noisy data[8] while also satisfying the need for accuracy and precision.[26] In addition, this method is strongly recommended because of its distribution-free nature as well as its ability to classify both categorical and numerical data. Furthermore, its tree-shaped structure can be easily interpreted and understood,[8,9] and its results can be described using a set of if-then rules,[11] which is an advantage for many medical applications.[27] In other words, in contrast to many other commonly used methods, for instance LR, there is no need to assume any distribution for the response when using the DT.[19,28] Clearly, the LR model has the benefit of parametric properties while the DT is non-parametric in nature.[29] However, the DT is able to deal with outliers[30] and missing data,[11,31] whereas the LR model estimates are usually biased when using this type of data.[32]

In this study, we found that the DT performed better than the LR model. Samanta et al.[6] also preferred the results of the DT over LR when they selected the hemodynamic features of periventricular leukomalacia in 2009. Sledjeski et al.[20] reached a considerably higher sensitivity (95% for the DT versus 37% for LR) and lower specificity (39% for the DT versus 80% for LR) for the DT when they sought to determine the high-risk group with regard to recurrent maltreatment by using data collected from investigations carried out in one Connecticutt county by the Connecticut Department of Children and Families. In contrast to these studies, other research, such as that conducted by Dreiseitl and Ohno-Machado[27] in 2003 and Mirta et al.[33] in 2005, concluded that both methods were equally effective.

Based on this study, surgeons should pay special attention to the potential for postoperative bleeding after CABG in men older than 44.5 and women older than 66.5 of age, especially those who are non-diabetics and who live in areas with medical facilities that are not as well equipped. For these high-risk groups, we strongly recommend performing off-pump CABG surgery or in the case of on-pump CABG surgery, the CPB duration should be shortened. Furthermore, patients should discontinue the use of antiplatelet medication prior to their surgery.[1-3] We also recommend that statisticians utilize the CART methods instead of LR for the purpose of classification.

Declaration of conflicting interests

The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The authors received no financial support for the research and/or authorship of this article.

REFERENCES

1. Karthik S, Grayson AD, McCarron EE, Pullan DM, Desmond MJ. Reexploration for bleeding after coronary artery bypass surgery: risk factors, outcomes, and the effect of time delay. Ann Thorac Surg 2004;78:527-34.
2. Choong CK, Gerrard C, Goldsmith KA, Dunningham H, Vuylsteke A. Delayed re-exploration for bleeding after coronary artery bypass surgery results in adverse outcomes. Eur J Cardiothorac Surg 2007;31:834-8.
3. Dacey LJ, Munoz JJ, Baribeau YR, Johnson ER, Lahey SJ, Leavitt BJ, et al. Reexploration for hemorrhage following coronary artery bypass grafting: incidence and risk factors. Northern New England Cardiovascular Disease Study Group. Arch Surg 1998;133:442-7.
4. Al-Fayes M, Allaham A, Shawabkeh Z, Al-Naser Y, Edwan H, Abu Anzeh R. Reopening for bleeding after adult cardiac surgery. Journal of the Royal Medical Services 2011;18:67-71.

5. Han J, Kamber M. Data mining: concepts and techniques. 2nd ed. San Francisco: Morgan Kauffman; 2006.
6. Samanta B, Bird GL, Kuijpers M, Zimmerman RA, Jarvik GP, Wernovsky G, et al. Prediction of periventricular leukomalacia. Part I: Selection of hemodynamic features using logistic regression and decision tree algorithms. Artif Intell Med 2009;46:201-15. doi: 10.1016/j.artmed.2008.12.005.
7. Karaolis MA, Moutiris JA, Hadjipanayi D, Pattichis CS. Assessment of the risk factors of coronary heart events based on data mining with decision trees. IEEE Trans Inf Technol Biomed 2010;14:559-66. doi: 10.1109/TITB.2009.2038906.
8. Xiao-Bai L. A scalable decision tree system and its application in pattern recognition and intrusion detection. Decis Support Syst 2005;41:112-30.
9. Bakır B, Batmaz I, Güntürkün FA, İpekçi I, Köksal G, Özdemirel N. Defect cause modeling with decision tree and regression analysis. World Acad Sci Eng Technoly 2006;24:1-4.
10. Aitkenhead M. A co-evolving decision tree classification method. Expert Syst Appl 2008;34:18-25.
11. Mitchell TM. Machine learning. 2nd ed. New York: McGraw-Hill; 1997.
12. Kurt I, Ture M, Kurum AT. Comparing performances of logistic regression, classificationand regression tree, and neural networks for predicting coronary artery disease. Expert Syst Appl 2008;34:366-74.
13. King G, Zeng L, Explaining rare events in international relations. International Organization 2001;55:693-715.
14. King G, Zeng L. Logistic regression in rare events data. Ipolitical Analysis 2000;9:137-63.
15. Kirklin JW, Barratt-Boyes BG. Cardiac surgery. New York: John Wiley & Sons; 1986. p. 158-9.
16. Hosmer DW, Lemeshow S. Applied logistic regression. 2nd ed. New York: John Wiley & Sons; 2000.
17. Rudolfer SM, Paliouras G, Peers IS. A comparison of logistic regression to decision tree induction in the diagnosis of carpal tunnel syndrome. Comput Biomed Res 1999;32:391-414.
18. Jenhani I, Ben N, Elouedi Z. Decision trees as possibilistic classifiers. Int J Approx Reason 2008;48:784-807.
19. Detsky AS, Naglie G, Krahn MD, Redelmeier DA, Naimark D. Primer on medical decision analysis: Part 2--Building a tree. Med Decis Making 1997;17:126-35.
20. Sledjeski EM, Dierker LC, Brigham R, Breslin E. The use of risk assessment to predict recurrent maltreatment: a Classification and Regression Tree Analysis (CART). Prev Sci 2008;9:28-37. doi: 10.1007/s11121-007-0079-0.
21. Agresti A. An Introduction to categorical data analysis. 2nd ed. New York: John Wiley & Sons; 2007.
22. Mehta RH, Sheng S, O'Brien SM, Grover FL, Gammie JS, Ferguson TB, et al. Reoperation for bleeding in patients undergoing coronary artery bypass surgery: incidence, risk factors, time trends, and outcomes. Circ Cardiovasc Qual Outcomes 2009;2:583-90. doi: 10.1161/CIRCOUTCOMES.109.858811.
23. McPhee S, Papadakis MA, Rabow MW. Current medical diagnosis and treatment. Fifty-first edition. New York: McGraw-Hill Medical; 2012.
24. Dao TK, Chu D, Springer J, Hiatt E, Nguyen Q. Depression and geographic status as predictors for coronary artery bypass surgery outcomes. J Rural Health 2010;26:36-43. doi: 10.1111/j.1748-0361.2009.00263.x.
25. Sangjae L. Using data envelopment analysis and decision trees for efficiency analysis and recommendation of B2C controls. Decis Support Syst 2010;49:486-97.
26. Skinner KR, Montgomery DC, Runger GC, Fowler JW, McCarville DR, Rhoads T, et al. Multivariate statistical methods for modeling and analysis of wafer probe test data. IEEE Transactions on Semiconductor Manufacturing 2002; 15:523-30.
27. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform 2002;35:352-9.
28. Rousseeuwa PJ, Christmannb A. Robustness against separation and outliers in logistic regression. Comput Stat Data Anal 2003;43:315-32.
29. Zhu M, Philpotts D, Sparks R, Stevenson M. A Hybrid approach to combining CART and logistic regression for stock ranking. Journal of Portfolio Management 2011;38:100-9.
30. Hodge VJ, Austin J. A survey of outlier detection methodologies. Artif Intell Rev 2004;22:85-126.
31. Twala BETH, Jones MC, Hand DJ. Good methods for coping with missing data in decision trees. Pattern Recognit Lett 2008;29:950-6.
32. Das U, Maiti T, Pradhan V. Bias correction in logistic regression with missing categorical covariates. J Stat Plan Inference 2010;140:2478-85.
33. Mirta B, Natasa S, Marijana ZS. Modelling small-business credit scoring by using logistic regression, neural networks and decision trees. Intell Sys Acc Fin Mgmt 2005;13:133-50.