International Journal of Advanced Culture Technology Vol.7 No.1 8-13 (2019)
DOI 10.17703/IJACT.2019.7.1.8

IJACT 19-3-2

Tree-based Approach to Predict Hospital Acquired Pressure Injury

Sookyung Hyun[a], Susan Moffatt-Bruce[b], Cheryl Newton[c], Brenda Hixon[d], Pacharmon Kaewprag[e]

[a]College of Nursing, Pusan National University, Yangsan-si, Korea,
[b]Department of Surgery, The Ohio State University, Columbus, Ohio, USA,
[c]Central Quality and Education, [d]Health System Nursing Education, The Ohio State University Wexner Medical Center, Columbus, Ohio, USA, [e]Department of Computer Engineering, Ramkhamhaeng University, Bangkok, Thailand.

sookyung.hyun@pusan.ac.kr, moffatt-bruce.1@osu.edu, Cheryl.Newton@osumc.edu, Brenda.Hixon@osumc.edu, pacharmon@gmail.com

Abstract

Despite technical advances in healthcare, the rates of hospital-acquired pressure injury (HAPI) are still high although many are potentially preventable. The purpose of this study was to determine whether tree-based prediction modeling is suitable for assessing the risk of HAPI in ICU patients. Retrospective cohort study has been carried out. A decision tree model was constructed with Age, Weight, eTube, diabetes, Braden score, Isolation, and Number of comorbid conditions as decision nodes. We used RStudio for model training and testing. Correct prediction rate of the final prediction model was 92.4 and the Area Under the ROC curve (AUC) was 0.699, which means there is about 70% chance that the model is able to distinguish between HAPI and non-HAPI. The results of this study has limited generalizability as the data were from a single academic institution. Our research finding shows that the data-driven tree-based prediction modeling may potentially support ICU sensitive risk assessment for HAPI prevention.

Keywords: *hospital-acquired pressure injury, intensive care units, decision tree*

1. INTRODUCTION

Hospital-acquired pressure injury (HAPI, formally known as a pressure ulcer) is a localized injury to the skin or underlying tissue caused by pressure [1]. HAPI result in additional cost and negatively affects the well-being of patients. HAPI is largely preventable. Braden scale is one of the most often used pressure ulcer risk assessment tools. It is sensitive [2]; however, it is not widely recommended for ICU patients [3-5].

Regression modeling is a standard statistical method that is widely used to identify significant variables and to construct a predictive model. Linear regression is used for regression problems and logistic regression is used for classification problems. However, if there is non-linear relationship between variables, a tree model performs better than a classical regression model. When compared to traditional statistical techniques, classification and prediction algorithms using machine learning methods have shown higher prediction capabilities in non-linear data relationships [6-9]. In addition, these algorithms have also demonstrated better performance on heterogeneous data sets from various sources [10-14].

Decision trees are one of the most commonly used classification algorithms. Trees are constructed by a so-

Manuscript received: January 3, 2019 / revised: January 25, 2019 / Accepted: February 15, 2019
Corresponding Author: sookyung.hyun@pusan.ac.kr
Author's affiliation: College of Nursing, Pusan National University, Yangsan-si, Korea

called *top-down*, recursive splitting approach. The process begins with a root node which represents the whole sample and is split into two or more homogeneous sub-nodes, of the root. The splitting operation is then applied recursively on each of the sub-nodes, and so on, unless a terminating condition prevents a sub-node from being split. Thus, the process ends with a tree with two types of nodes: those divided into sub-nodes are called *decision*, or *internal* nodes; and those which are not further divided are called *terminal* or *leaf* nodes [15]. A node that has sub-nodes is called a *parent* of the sub-nodes while the sub-nodes are *children* of the parent node. Decision trees are easy to understand and they can manage categorical variables as well as numerical variables. The purposes of this study were to apply a tree-based approach to predict HAPI in ICU settings and to evaluate the accuracy of the prediction models.

2. METHODS

Data were retrieved from an information warehouse from an academic medical center. We described the dataset in detail in another journal publication, including data extraction, cleaning, and preparation [16]. For the purpose this study, we created a subset of the data that included relevant ICU patient encounters. The research study was approved by the institutional review board. The dataset included data elements, such as Gender, Age, Race/Ethnicity, Weight, Number of comorbid conditions, diabetes, Vasopressor, Isolation, endotracheal tube (eTube), and Braden score. Bivariate analyses were performed to determine whether they were associated with HAPI development. Descriptive statistics was used to summarize the data. We used RStudio (Version 1.1.463) to construct and validate tree-based predictive models. The total number of patient encounters in the analysis was 12,652. We split the dataset into training (2/3) and testing (1/3) datasets. We created a decision tree that predicts patients at risk for developing a HAPI by using the training dataset and then validated the model by using testing dataset. The accuracy of a predictive model was evaluated by using the area under the ROC curve (AUC) and correct prediction rate.

3. RESULTS

Table 1 illustrates the demographic profiles of the patient encounters. The average age was 57.5. Fifty-six point seven percent were male patients. In terms of race/ethnicity, the majority of the ICU patients were White, followed by Black. The average number of comorbid conditions was 6.4 with a range of 0 to 20. The average ICU stay was 12.1 days with a range of 3 to 216 days. The total number of patients who developed a HAPI during their ICU stay was 965. Figure 1 graphically compares Braden scores between patients with and without a HAPI. In both groups, 75 percentile of patients' Braden scores were below 16 (upper dotted line), which is a Braden's recommended cut-off score for ICU settings.

Table 1. Demographics of patients (N=12,652)

Category	Group	Mean (SD)/Freq.(%)	Range
Age (years)		57.5 (15.5)	[18, 109]
Gender	Male	7178 (56.7)	
	Female	5474 (43.3)	
Race/Ethnicity	White	10225 (80.8)	
	Black	1929 (15.2)	
	Hispanic	94 (0.8)	
	Asian/Native Hawaii	77 (0.6)	
	Other	327 (2.6)	
Braden score		14.0 (3.6)	[6, 23]
Number of comorbid conditions		6.4 (4.5)	[0, 20]
Length of ICU (days)		12.1(12.6)	[3, 216]

The mean Braden score was below 12.5 (lower dotted line) in the patients with a HAPI whereas it was above in the patients without.

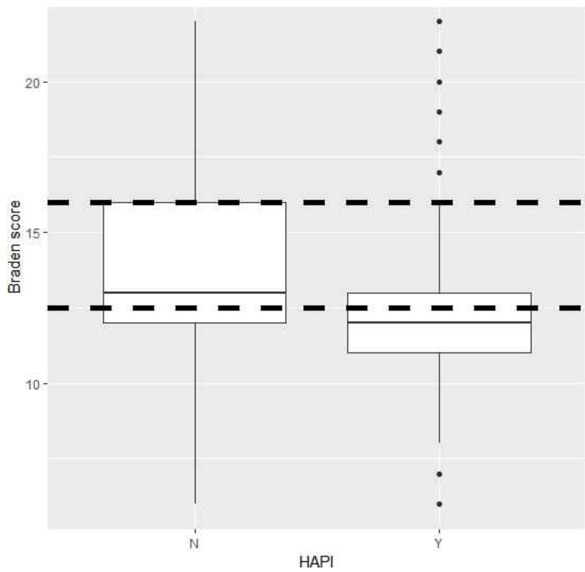

Figure 1. The comparison of the Braden scores

The numbers of patient encounters included in the training and testing datasets were 8,857 (2/3 of the dataset) and 3,795 (1/3), respectively. For model construction, we used the training dataset with HAPI (yes/no) as a dependent variable. In terms of independent variables, all, but Vasopressor were significantly associated with HAPI development in the bivariate analyses. We used the nine data elements as independent variables: Gender, Age, Race/Ethnicity, Weight, Number of comorbid conditions, diabetes, Isolation, eTube, and Braden score. In decision tree construction, tree size is commonly used as a regularization parameter to balance model performance against overfitting risk. We chose optimal tree size by varying complexity parameter (CP), which is a minimum threshold for goodness-of-fit improvement when determining whether to split a node. Figure 2 graphically illustrates relative error as a function of complexity parameter. We used a CP of 0.0027, which showed the minimum cross-validated error rate for our final prediction model.

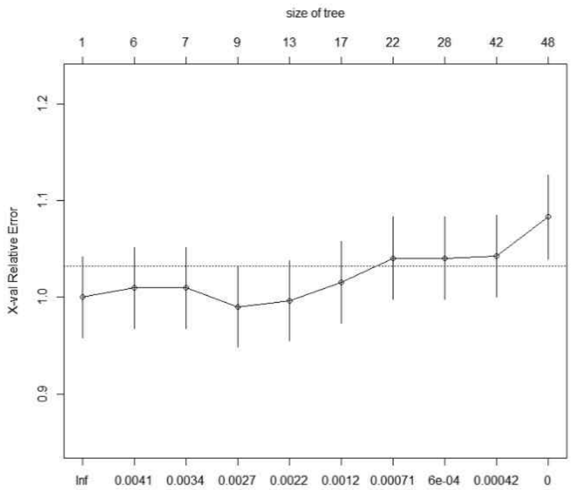

Figure 2. The by complexity parameters

The final model is illustrated in Figure 3. Variables actually selected by the tree construction process were Age, Weight (Weight_lbs; measured in pounds [lbs.]), Number of comorbid conditions (Num_of_dx), diabetes (DM), Isolation, eTube, and Braden score (BradenTotal). The number of terminal nodes was nine in the tree model. Number of comorbid conditions was used as a root node. The median Number of comorbid conditions was 11 in patients with a HAPI while 5 in those without. There were seven decision nodes: Isolation, Braden score (>=13), Age (>=42), eTube, Weight (<249), diabetes, Age (<83). Gender and Race/ethnicity were not selected for use at any decision node in the final tree model.

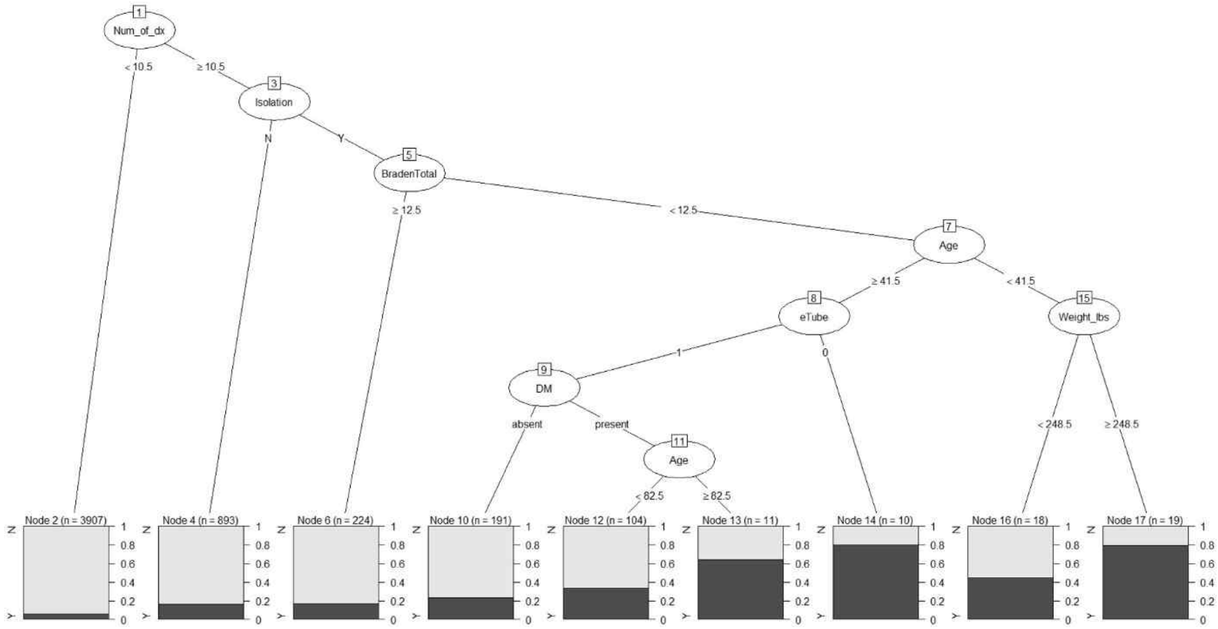

Figure 3. The final model for HAPI prediction

For model validation, we used the testing dataset. Correct prediction rate of the prediction model was 92.4. The ROC curve is displayed in Figure 4. The Area Under the ROC curve (AUC) was 0.699, which means there is about 70% chance that the model is able to distinguish between HAPI and non-HAPI. The accuracy of the decision tree model appeared to be slightly better than the Braden scale (AUC of 0.672) from the findings in the previous research study [17].

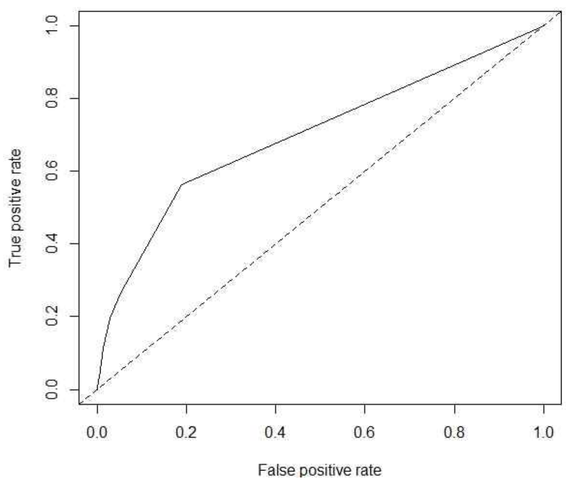

Figure 4. ROC curve of the final model

4. DISCUSSION AND CONCLUSION

We conducted a retrospective cohort study with electronic health data from about 12,000 patient encounters. The dataset included 12 variables such as gender, age, race/ethnicity, weight, number of comorbid conditions, diabetes, Isolation, eTube, vasopressor, Braden score, and HAPI. The final tree model was constructed with age, weight, eTube, diabetes, Braden score, isolation, and number of comorbid conditions. The final model showed slightly better performance than Braden score in our data.

Despite technical advances in healthcare, the rates of HAPI incidence are still unacceptably high. The Braden scale is not considered sufficient for identifying patients at risk for developing HAPI in ICU settings. A paucity of research has been done in the area of ICU HAPI prediction modeling using machine learning [1]. In this research study, we used a large amount of data accumulated in ICU electronic health records in order to determine whether is suitable to apply machine learning to prediction modeling. Decision tree is one of machine learning algorithms and can help identify most significant variables in exploring data. Compared to other machine learning methods, decision trees are not black-box models, but can be easily interpreted [18].

The results of this study have limited generalizability as the data were from a single academic institution; however, our research finding shows that applying a tree-based classification method to HAPI prediction may be potentially suitable to support clinician's clinical judgement with regards to HAPI prediction and prevention in ICU settings. Additionally, the model has demonstrated appropriate performance in distinguishing between HAPI and non-HAPI and has higher predictive capabilities than the Braden scale. We determined that a data-driven, tree-based prediction model may support ICU sensitive risk assessment in critical care.

ACKNOWLEDGMENT

This work was supported by Pusan National University Research Grant, 2015 and grant UL1RR025755 from the National Center For Research Resources. We would like to thank Tara Payne, Marcia Belcher, and Information Warehouse staff for their assistance with data extraction.

REFERENCES

[1] The National Pressure Ulcer Advisory Panel. National Pressure Ulcer Advisory Panel (NPUAP) announces a change in terminology from pressure ulcer to pressure injury and updates the stages of pressure injury. Available from: http://www.npuap.org/national-pressure-ulcer-advisory-panel-npuap-announces-a-change-in-terminology-from-pressure-ulcer-to-pressure-injury-and-updates-the-stages-of-pressure-injury/.

[2] Bergstrom N, Braden BJ, Kemp M, Champagne M, Ruby E., Predicting pressure ulcer risk: a multisite study of the predictive validity of the Braden Scale. Nurs Res, 1998. 47(5): p. 261-9.

[3] Schoonhoven L, Haalboom J, Bousema M, Algra A, Grobbee D, Grypdonck M, et al., Prospective cohort study of routine use of risk assessment scales for prediction of pressure ulcers. BMJ, 2002. 325(7368).

[4] Kottner J, Dassen T, Pressure ulcer risk assessment in critical care: interrater reliability and validity studies of the Braden and Waterlow scales and subjective ratings in two intensive care units. International Journal of Nursing Studies, 2010. 47(6): p. 671-677. 10.1016/j.ijnurstu.2009.11.005.

[5] Cox J, Predictive power of the Braden scale for pressure sore risk in adult critical care patients: a comprehensive review. J Wound Ostomy Continence Nurs, 2012. 39(6): p. 613-21; quiz 622-3. 10.1097/WON.0b013e31826a4d83.

[6] Liew P, Lee Y, Lin Y, Lee T, Lee W, Wang W, et al., Comparison of artificial neural networks with logistic regression in prediction of gallbladder disease among obese patients. Dig Liver Dis, 2007. 39(4): p. 356-62.

[7] Kurt I, Ture M, Kurum A, Comparing performances of logistic regression classification and regression tree, and neural networks for predicting coronary artery disease. Expert Systems with Applications, 2008. 34: p. 366-374.

[8] Worachartcheewan A, Nantasenamat C, Isarankura-Na-Ayudhya C, Pidetcha P, Prachayasittikul V, Identification of metabolic syndrome using decision tree analysis. Diabetes Res Clin Pract, 2010. 90(1): p. e15-8. 10.1016/j.diabres.2010.06.009.
[9] Mullins I, Siadaty M, Lyman J, Scully K, Garrett C, Miller W, et al., Data mining and clinical data repositories: Insights from a 667,000 patient data set. Comput Biol Med, 2006. 36(12): p. 1351-77.
[10] Brickley M, Shepherd J, Armstrong R, Neural networks: a new technique for development of decision support systems in dentistry. J Dent, 1998. 26(4): p. 305-9.
[11] Ridinger M, Rice J, Predictive modeling points way to future risk status. Health Manag Technol, 2000. 21(2): p. 10-2.
[12] Lin S, Lee C, Lu Y, Hsu L, A comparison of MICU survival prediction using the logistic regression model and artificial neural network model. J Nurs Res, 2006. 14(4): p. 306-14.
[13] Ellenius J, Groth T, Dynamic decision support graph--visualization of ANN-generated diagnostic indications of pathological conditions developing over time. Artif Intell Med, 2008. 42(3): p. 189-98.
[14] Tabaton M, Odetti P, Cammarata S, Borghi R, Monacelli F, Artificial neural networks identify the predictive values of risk factors on the conversion of amnestic mild cognitive impairment. J Alzheimers Dis, 2010. 19(3): p. 1035-40. 10.3233/JAD-2010-1300.
[15] Kuhn M, Johnson K, Applied Predictive Modeling. 2016, New York: Springer.
[16] Kaewprag P, Newton C, Vermillion B, Hyun S, Huang K, Machiraju R, Predictive models for pressure ulcers from intensive care unit electronic health records using Bayesian networks. BMC Med Inform Decis Mak, 2017. 17(Suppl 2): p. 65. 10.1186/s12911-017-0471-z.
[17] Hyun S, Vermillion B, Newton C, Fall M, Li X, Kaewprag P, et al., Predictive validity of the Braden scale for patients in intensive care units, American Journal of Critical Care, 2013. 22(6): p. 514-520. 10.4037/ajcc2013991.
[18] Dreiseitl S, Ohno-Machado L, Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform, 2002. 35(5-6): p. 352-9.