J. Phys. Ther. Sci. 31: 69–74, 2019

The Journal of Physical Therapy Science

Original Article

Use of a multilayer perceptron to create a prediction model for dressing independence in a small sample at a single facility

Takaaki Fujita, OTR, PhD[1)*], Atsushi Sato, RPT, MS[2)], Akira Narita, PhD[3)], Toshimasa Sone, OTR, PhD[1)], Kazuaki Iokawa, OTR, PhD[4)], Kenji Tsuchiya, OTR, PhD[5)], Kazuhiro Yamane, OTR[6)], Yuichi Yamamoto, RPT[6)], Yoko Ohira, MD, PhD[6)], Koji Otsuki, MD, PhD[6)]

1) Department of Rehabilitation, Faculty of Health Sciences, Tohoku Fukushi University: 1-8-1 Kunimi, Aoba-ku, Sendai-shi, Miyagi 981-8522, Japan
2) Department of Rehabilitation, Care Center Moriyama, Japan
3) Tohoku Medical Megabank Organization, Tohoku University, Japan
4) Preparing Section for New Faculty of Medical Science, Fukushima Medical University, Japan
5) Department of Rehabilitation Sciences, Gunma University Graduate School of Health Sciences, Japan
6) Department of Rehabilitation, Kita-Fukushima Medical Center, Japan

Abstract. [Purpose] This study aimed to assess the accuracy of a prediction model for dressing independence created with a multilayer perceptron in a small sample at a single facility. [Participants and Methods] This retrospective observational study included 82 first-stroke patients. The prediction models for dressing independence at hospital discharge were created using a multilayer perceptron, logistic regression, and a decision tree, and compared for predictive accuracy. Age, dressing performance, trunk function, visuospatial perception, balance, and cognitive function at admission were used as variables. [Results] The area under the receiver operating characteristic curve, classification accuracy, sensitivity, specificity, positive-predictive value, and negative-predictive value for training data were highest with the multilayer perceptron model. Cochran's Q and multiple comparison tests revealed a significant difference between logistic regression and multilayer perceptron models. Testing of data in 10-fold cross-validation yielded the same results, except for sensitivity. [Conclusion] The present study suggested that higher accuracy could be expected with a multilayer perceptron than with logistic regression and a decision tree when creating a prediction model for independence of activities of daily living in a small sample of stroke patients.
Key words: Activities of daily living, Prediction model, Multilayer perceptron

(This article was submitted Aug. 23, 2018, and was accepted Oct. 7, 2018)

INTRODUCTION

Stroke is considered to be a major cause of serious long-term disability[1)]. Various dysfunctions after stroke, such as hemiplegia, impaired balance, and cognitive impairment, can decrease the level of independence in activities of daily living (ADLs) among patients, making it difficult for many patients to lead an independent lifestyle. To maximize the recovery of functional independence in stroke patients, rehabilitation often includes therapeutic and compensatory programs for improving impairments and the ability to perform ADLs. However, in some patients, ADL ability might not completely recover, and nursing care might be required for such patients, even many years after stroke onset[2)]. This finding indicates that a relatively large number of stroke patients are discharged from hospitals despite having difficulties in daily living.

*Corresponding author. Takaaki Fujita (E-mail: t-fujita@tfu-mail.tfu.ac.jp)

The patient and the patient's family members are usually concerned about whether nursing care will be required after discharge. To facilitate a smooth transition from hospital to home among stroke patients, it is important to predict each patient's level of independence in ADLs at discharge from hospital in order to prepare for necessary personal care plan and appropriate physical environment (e.g., introduction of assistive technologies and welfare services and provision of information to family members on the methods of nursing care). The predicted level of independence in ADLs at discharge also provides important information that is used by rehabilitation staff when preparing a program according to the patient's status at discharge. Although various studies have investigated the methods for predicting ADLs, no decisive method has been identified[3]. In addition, we previously reported that a prediction model should be developed for each facility to improve prediction accuracy[4]. However, in our experience, it is not easy to collect sufficient data to create a prediction model at a single facility because the number of patients is limited at single facilities, except at large hospitals. In brief, the sample size of usable data tends to be small. To overcome this issue, a method for creating a model with high accuracy despite a small sample size should be established.

Recently, previous studies related to stroke in elderly individuals indicated that multilayer perceptron, which is a class of feedforward artificial neural network, can create a prediction model with good accuracy[5, 6]. For example, Cheng et al.[5] reported that the model created by multilayer perceptron achieved a good performance in predicting the occurrence of major adverse cardiovascular events in patients requiring carotid artery stenting treatment. Additionally, Colak et al.[6] compared three prediction models for the outcome of stroke, which were created using multilayer perceptron, knowledge discovery process, and support vector machine, and found that multilayer perceptron had a more predictive performance in predicting stroke compared with support vector machine. However, almost 300 samples were used in these studies[5, 6], and no previous study has investigated small samples, such as less than 100.

Therefore, the present study aimed to investigate whether multilayer perceptron could create a highly accurate prediction model using a small sample at a single facility. We targeted dressing, which is a relatively difficult ADL for stroke patients[2], and we attempted to create a model for predicting dressing independence at discharge from a rehabilitation ward, using the admission status.

PARTICIPANTS AND METHODS

This retrospective observational study included 82 first-stroke patients. These patients underwent rehabilitation following stroke in the rehabilitation ward of a hospital in Japan. The inclusion criteria were as follows: (1) admission for stroke and discharge between April 2011 and February 2014, (2) diagnosis of initial cerebral hemorrhage or cerebral infarction, (3) presence of unilateral supratentorial lesions, (4) inability to dress independently on hospitalization (five points or lower for FIM® instrument dressing upper body, lower body, or both), and 5) no missing information for data analysis. All patients generally underwent occupational therapy, physical therapy, and, if necessary, speech therapy for 2–3 hours per day on weekdays and Saturdays, and for 1–2 hours per day on Sundays and public holidays. The therapies addressed issues such as ADLs, upper limb function, balance, walking, language, and cognitive ability, as necessary in each participant. Informed consent was not required because the design of our study was retrospective without intervention. However, instead of informed consent, our protocol was considered by the Institutional Review Boards of Kita-Fukushima Medical Center and Tohoku Fukushi University, and approved (No. 72, RS180103).

We assessed the medical records and gathered information about clinical variables that have been reported to be associated with dressing performance in stroke patients. These variables were age[7], dressing performance prior to practice[8], trunk function[7], visuospatial perception[7], and balance[9, 10]. Cognitive function, which has been reported to affect the degree of ADL improvement[11], was also considered as a variable in this study. The evaluation methods for each function were decided according to previous studies[7–11]. We used FIM® dressing items[12], Stroke Impairment Assessment Set (SIAS) vertical items, visuospatial deficit items[13], the Berg Balance Scale (BBS)[14], and FIM® cognitive items. These variables at the time of admission were used as independent variables in the prediction model. On the other hand, the index for dressing independence at discharge, which was a dependent variable in the prediction model, used the FIM® dressing item score at discharge. The FIM® dressing items included both upper and lower body items. In this study, we used the lower score of the two items as an index for the degree of dressing independence.

We compared prediction models created by multilayer perceptron with logistic regression and decision tree, which are conventionally used methods. First, we divided the participants into an independent group (FIM® dressing $\geq$6 points) and a non-independent group (FIM® dressing $\leq$5 points) according to the FIM® dressing score at discharge. For the selection of variables to be inserted into the model, we performed inter-group comparison of each variable at admission using Student's t-test, the χ^2 test, and the Mann-Whitney test. Next, we performed multilayer perceptron, logistic regression, and decision tree using the variables at admission that showed significant differences between the groups as independent variables and dressing independence–non-independence upon discharge as the dependent variable. For multilayer perceptron, we used a hierarchical model with one intermediate layer. For threshold adjustment, we added bias items to always output 1 in the input layer and intermediate layer. In multilayer perceptron, taking sample size into account, the variables selected with logistic regression were used as independent variables, and age was added after its division into the following four nominal scales: <65 years, 65–74 years, 75–84 years, and $\geq$85 years. To prevent overtraining, we set the training sample and testing sample

Table 1. Stroke-related characteristics of the study patients

Variables	Overall (N=82)	Dressing at discharge	
		Independent (N=34)	Dependent (N=48)
Age (years)	73.6 ± 12.5	69.3 ± 12.0	76.6 ± 12.0**
Men (%)	56.1	64.7	50.0
Right-side hemiplegia (%)	42.7	44.1	41.7
Post-stroke time at admission (days)	36.6 ± 15.3	34.9 ± 15.6	37.7 ± 15.2
Post-stroke time at discharge (days)	102.6 ± 36.9	101.4 ± 40.2	103.5 ± 34.8
Length of hospital stay (days)	66.1 ± 32.3	66.5 ± 33.8	65.8 ± 31.6
FIM® dressing item at admission (1–7)	2.4 ± 1.5	3.4 ± 1.4	1.7 ± 1.2**
SIAS verticality item at admission (0–3)	2.2 ± 1.1	2.7 ± 0.6	1.8 ± 1.2**
SIAS visuospatial deficit item at admission (0–3)	2.4 ± 1.5	2.8 ± 0.5	2.0 ± 1.2**
Berg Balance Scale at admission (0–56)	16.7 ± 16.0	26.5 ± 16.0	9.7 ± 11.9**
FIM® cognitive item at admission (5–35)	22.9 ± 9.1	27.8 ± 7.4	19.4 ± 8.6**

Data are presented as mean ± SD.
**p<0.01.
SIAS: stroke impairment assessment set.

(used to track errors during training) ratio at 9:1. In models created with multilayer perceptron, the initial values for weighting from the input layer to the intermediate layer were determined randomly, and the prediction accuracy exhibited dependency on these. Thus, we reset the initial values 10 times and used the model with the highest prediction accuracy. For logistic regression, we used the stepwise method (likelihood ratio variable, forward selection). For decision tree, we used classification and regression trees (CART)[15]. The maximum tree depth was set at 3, parent node smallest sample size was set at 10, and child node smallest sample size was set at 3. The Gini index, which represents impurity, was used to determine branches, and the standard error rule (± 1) was used for pruning[15]. To compare the accuracy of the prediction models created with the different methods, we calculated the area under the receiver operating characteristic curve (AUROC), classification accuracy, sensitivity, specificity, positive-predictive value, and negative-predictive value. Classification accuracy was defined as the proportion of true results, either true positive or true negative, in participants. The conformity of the predicted outcomes using each model and the actual outcomes for the patients was assessed using Cochran's Q test and the multiple comparison test (McNemar's test using Bonferroni correction).

In addition, 10-fold cross-validation was performed to evaluate the generalized performance of the models created using multilayer perceptron, logistic regression, and decision tree. For 10-fold cross-validation, all participants were randomly divided into 10 groups. A model was created using the training data of nine groups, while data in the remaining group was used as testing data. We verified accuracy 10 times. To improve evaluation bias and variance, we performed 10-fold cross-validation after adjusting the ratios of patients in the independent and non-independent subgroups to be even among the 10 groups. We calculated the mean values in 10-fold cross-validation and obtained verification data (AUROC, classification accuracy, sensitivity, specificity, positive-predictive value, and negative-predictive value). Additionally, we used Friedman's test and the multiple comparison test (Wilcoxon signed-rank test using Bonferroni correction) for comparisons among the models. Moreover, the conformity of the prediction of independence or non-independence for each model and the actual outcomes was assessed using Cochran's Q test and the multiple comparison test. The level of significance was set at 5% for all tests, and all analyses were performed using SPSS Statistics version 25 (IBM Corp., Armonk, NY, USA).

RESULTS

Table 1 shows the stroke-related characteristics, physical and mental functions, and degree of dressing independence among the patients. Of the 82 patients, 34 showed independent dressing and 48 showed non-independent dressing at discharge. Intergroup comparisons revealed significant differences in age, FIM® dressing at admission, SIAS verticality, SIAS visuospatial perception, BBS, and FIM® cognition. With regard to the creation of prediction models using these items, logistic regression created a model with age, SIAS visuospatial perception, and FIM® dressing, and decision tree created a model with BBS and age (Fig. 1). Multilayer perceptron created a model with age range, SIAS visuospatial perception, and FIM® dressing as independent variables, resulting in a model with three intermediate layer units.

The classification accuracy, sensitivity, specificity, positive-predictive value, and negative-predictive value for the training data were the highest with the multilayer perceptron model, followed by the decision tree and logistic regression models (Table 2). The AUROC was the highest with the multilayer perceptron model, followed by the logistic regression and decision tree models. Cochran's Q test identified differences in the predicted outcomes using each model and the actual outcomes

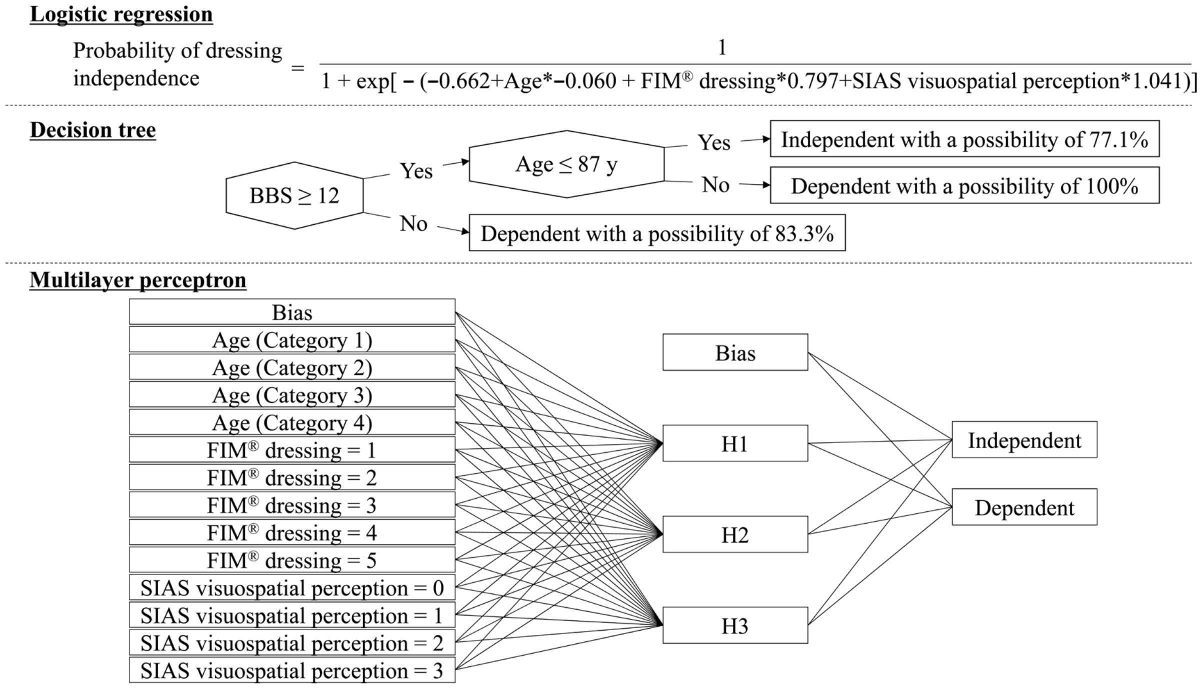

Fig. 1. The models for prediction dressing independence using logistic regression, decision tree, and multilayer perceptron.
If the probability calculated by logistic regression model is more than 0.5, dressing at discharge is predicted as "independence". For example, the probability of a 70 years-patient with a scores of 3 for FIM® dressing and SIAS visuospatial deficit items at admission is 0.66, and dressing performance at discharge is predicted as "independence". The decision tree created a model with BBS and age, and a patient who had over 12 points for BBS and was under 87 years-old was predicted as "independence." Multilayer perceptron created a model with three intermediate layer units, and each connection between neurons were adjusted for optimal weight.
SIAS: stroke impairment assessment set; BBS: Berg balance scale; H: hidden unit.

Table 2. Comparison of the performance of analysis methods

		AUROCC	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Entire dataset	LR	0.865	74.4	79.2	67.6	77.6	69.7
	DT	0.824	81.7	83.3	79.4	85.1	77.1
	MLP	0.937	86.8	87.0	86.7	90.9	81.3
Validation†	LR	0.807	69.5	66.7	70.5	61.5*	75.8
	DT	0.655*	65.8	60.8	68.0	58.9*	75.1
	MLP	0.830*	76.8	62.5	84.5	85.4*	80.3

†Mean value for 10 samples from 10-fold cross-validation.
*p<0.05 on pairwise comparisons.
LR: logistic regression; DT: decision tree; MLP: multilayer perceptron; AUROCC: area under the receiver operating characteristic curve; PPV: positive-predictive value; NPV: negative-predictive value.

among the models (p<0.05), and comparison of each pair of models revealed a significant difference between the logistic regression and multilayer perceptron models (p<0.05).

The AUROC, classification accuracy, specificity, positive-predictive value, and negative-predictive value for the testing data in 10-fold cross-validation were the highest with the multilayer perceptron model, followed by the logistic regression and decision tree models. Only sensitivity was the highest with the logistic regression model. Cochran's Q test identified no significant differences among the groups with regard to conformity between prediction outcomes and results. However, Friedman's test and multiple comparisons indicated that the AUROC and positive-predictive value were higher with the multilayer perceptron model than with the decision tree model (p<0.05) and logistic regression model (p<0.05), respectively.

DISCUSSION

The prediction of ADL independence from admission to discharge in stroke patients is important for achieving a smooth transition to post-discharge living and for planning rehabilitation programs during hospitalization. As factors such as rehabilitation intensity, frequency, participant characteristics, and personal and physical environments differ among facilities, it has been considered important to create an individual prediction model for each facility in order to increase prediction accuracy[4]. However, when creating a prediction model at a single facility, the sample size of usable data tends to be small. The present study compared the accuracy of models created using multilayer perceptron, which has been reported to be useful in recent years, with logistic regression and decision tree, which are conventional techniques, in order to identify the most appropriate method for creating prediction models using small samples at single facilities. The results of the present study suggested that multilayer perceptron was the most appropriate method for creating a prediction model to determine independence of ADL items in a small sample of stroke patients. Multilayer perceptron has been reported to require an extremely large sample for training[16]. However, it appears that this method can be used to create a relatively accurate model even with a small sample if, as in the present study, the number of input layer parameters is limited with approaches, such as the adoption of variables selected with logistic regression, or the intermediate layer set to a single layer.

As in the present study, Omurlu et al.[17] and Colombet et al.[18] created prediction models using multilayer perceptron, logistic regression, and decision tree and then compared accuracy. Omurlu et al.[17] investigated the prediction of the presence of albuminuria in type II diabetes patients and reported that the model created with multilayer perceptron offered the highest prediction accuracy. Colombet et al.[18] investigated the prediction of cardiovascular risk and reported that the predictive ability of decision tree (CART) was slightly lower than the predictive abilities of logistic regression and multilayer perceptron. In addition, HeidarAbadi et al.[19] compared the accuracies of various classification algorithms including decision tree and multilayer perceptron for the prediction of pain in spinal cord injury patients and reported that the multilayer perceptron model had the highest prediction accuracy. As the results of our study are consistent with the results of these previous studies, our results appear to be valid.

The differences noted in predictive ability among multilayer perceptron, logistic regression, and decision tree in the present study were relatively low with a small sample. Therefore, as reported by Colombet et al.[18], these methods can be used in a complementary manner. For example, although multilayer perceptron might be optimal for obtaining the highest prediction accuracy, decision tree offers the advantage of easy visual comprehension of results and ease of use. Our results (that the difference in predictive ability among models was small) indicated that there would be no marked decrease in the predictive ability with the use of multilayer perceptron, logistic regression, or decision tree depending on the clinical situation.

The present study has some limitations. First, the results of this study were prediction models created using SIAS, BBS, and FIM® instrument items. The possibility of obtaining different results when using other parameters cannot be denied. Second, we adopted 10-fold cross-validation that involved the use of internal data to verify validity. In the future, further verification involving a hold-out method using unknown data should be performed. Third, our results were for a single facility. Thus, similar findings need to be confirmed in other facilities for generalizability. Additionally, the purpose of the present study was to only compare the prediction accuracies of different models. Thus, the validities of the constructed models need to be investigated further.

In conclusion, our results suggest that higher accuracy could be expected with multilayer perceptron than with logistic regression and decision tree when creating a prediction model for the independence of ADL items in a small sample of stroke patients.

Presentation at a conference

A part of this research was presented in the 52nd Japanese Occupational Therapy Congress (https://www.mas-sys.com/JOTC52_Abstract/pdf/endai100064.pdf).

Funding

This work was supported by a grant from the Japanese Society for the Promotion of Science Grants-in-Aid for Scientific Research (grant no. 18K17728).

Conflict of interest

There are no conflicts of interest to declare.

ACKNOWLEGMENT

The use of the FIM® instrument to collect data for this research study was authorized and conducted in accordance with the term of a special purpose license granted to the licensee by the Uniform Data System for Medical Rehabilitation (UDSMR). Licensee has not been trained by UDSMR in the use of the FIM® instrument, and the patient data collected

during the course of this research study has not been submitted to or processed by UDSMR. No implication is intended that such data has been or will be subjected to UDSMR's standard data processing procedures or that it is otherwise comparable to data processed by UDSMR. FIM® is a trademark of Uniform Data System for Medical Rehabilitation, a division of UB Foundation Activities, Inc.

REFERENCES

1) Mozaffarian D, Benjamin EJ, Go AS, et al. American Heart Association Statistics Committee and Stroke Statistics Subcommittee: Heart disease and stroke statistics—2015 update: a report from the American Heart Association. Circulation, 2015, 131: e29–e322. [Medline]
2) Edmans JA, Towle D, Lincoln NB: The recovery of perceptual problems after stroke and the impact on daily life. Clin Rehabil, 1991, 5: 301–309. [CrossRef]
3) Veerbeek JM, Kwakkel G, van Wegen EE, et al.: Early prediction of outcome of activities of daily living after stroke: a systematic review. Stroke, 2011, 42: 1482–1488. [Medline] [CrossRef]
4) Sato A, Fujita T, Ohashi Y, et al.: A prediction model for activities of daily living for stroke patients in a convalescent Rehabilitation Ward. J Allied Health Sci, 2016, 7: 1–6. [CrossRef]
5) Cheng CA: Hung-Wen Chiu: An artificial neural network model for the evaluation of carotid artery stenting prognosis using a national-wide database. Conf Proc IEEE Eng Med Biol Soc, 2017, 2566–2569. [Medline] [CrossRef]
6) Colak C, Karaman E, Turtay MG: Application of knowledge discovery process on the prediction of stroke. Comput Methods Programs Biomed, 2015, 119: 181–185. [Medline] [CrossRef]
7) Saito Y, Toshima M, Nori K, et al.: Factors affecting the upper body dressing ability of hemiplegia patients in the convalescent stage −investigation based on functionality. Jpn Occup Ther Res, 2012, 31: 134–140.
8) Suzuki M, Omori M, Hatakeyama M, et al.: Predicting recovery of upper-body dressing ability after stroke. Arch Phys Med Rehabil, 2006, 87: 1496–1502. [Medline] [CrossRef]
9) Fujita T, Nagayama H, Sato A, et al.: Hierarchy of dysfunction related to dressing performance in stroke patients: a path analysis study. PLoS One, 2016, 11: e0151162. [Medline] [CrossRef]
10) Fujita T, Sato A, Yamamoto Y, et al.: Relationship between dressing and motor function in stroke patients: a study with partial correlation analysis. J Phys Ther Sci, 2015, 27: 3771–3774. [Medline] [CrossRef]
11) Imada Y, Tokunaga M, Fukunaga K, et al.: Relationship between cognitive FIM score and motor FIM gain in patients with stroke in a Kaifukuki Rehabilitation Ward. Jpn J Compr Rehabil Sci, 2014, 5: 12–18.
12) Data Management Service of the Uniform Data System for Medical Rehabilitation and the Center for functional Assessment Research: Guide for the Uniform Data Set for Medical Rehabilitation, version 3.0. New York: State University of New York at Buffalo, 1990.
13) Chino N, Sonoda S, Domen K, et al.: Stroke impairment assessment set (SIAS)—a new evaluation instrument for stroke patients. Jpn J Rehabil Med, 1994, 31: 119–125. [CrossRef]
14) Berg K, Wood-Dauphinee S, Williams JI: Measuring balance in the elderly: preliminary development of an instrument. Physiother Can, 1989, 41: 304–311. [CrossRef]
15) Breiman L, Friedman JH, Olshen RA, et al.: Classification and regression trees. Belmont: Wadsworth Publishing, 1984.
16) Pasini A: Artificial neural networks for small dataset analysis. J Thorac Dis, 2015, 7: 953–960. [Medline]
17) Omurlu IK, Ture M, Unubol M, et al.: Comparing performances of logistic regression, classification & regression trees and artificial neural networks for predicting albuminuria in type 2 diabetes mellitus. IJSBAR, 2014, 16: 173–187.
18) Colombet I, Ruelland A, Chatellier G, et al.: Models to predict cardiovascular risk: comparison of CART, multilayer Perceptron and logistic regression. Proceedings. AMIA Symposium, 2000, 156–160.
19) HeidarAbadi NN, Hakemi L, Kolivand P, et al.: Comparing performances of intelligent classifier algorithms for predicting type of pain in patients with spinal cord injury. Electron Physician, 2017, 9: 4847–4852. [Medline] [CrossRef]