International Journal of Medical Informatics 111 (2018) 159–164

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Predicting post-stroke activities of daily living through a machine learning-based approach on initiating rehabilitation

Wan-Yin Lin[a,1], Chun-Hsien Chen[b,c,1], Yi-Ju Tseng[b,d,1], Yu-Ting Tsai[e], Ching-Yu Chang[e], Hsin-Yao Wang[d,*], Chih-Kuang Chen[e,f,**]

[a] Department of Physical Medicine & Rehabilitation, Chang Gung Memorial Hospital at Linkou, Taoyuan City, Taiwan
[b] Department of Information Management, Chang Gung University, Taoyuan City, Taiwan
[c] Department of Neurology, Chang Gung Memorial Hospital at Taoyuan, Taoyuan City, Taiwan
[d] Department of Laboratory Medicine, Chang Gung Memorial Hospital at Linkou, Taoyuan City, Taiwan
[e] School of Medicine, Chang Gung University, Taoyuan City, Taiwan
[f] Department of Physical Medicine & Rehabilitation, Chang Gung Memorial Hospital at Taoyuan, Taoyuan City, Taiwan

A R T I C L E I N F O

Keywords:
Stroke
Activities of daily living
Rehabilitation
Machine learning
Computer-assisted decision making

A B S T R A C T

Objectives: Prediction of activities of daily living (ADL) is crucial for optimized care of post-stroke patients. However, no suitably-validated and practical models are currently available in clinical practice.

Methods: Participants of a Post-acute Care-Cerebrovascular Diseases (PAC-CVD) program from a reference hospital in Taiwan between 2014 and 2016 were enrolled in this study. Based on 15 rehabilitation assessments, machine learning (ML) methods, namely logistic regression (LR), support vector machine (SVM), and random forest (RF), were used to predict the Barthel index (BI) status at discharge. Furthermore, SVM and linear regression were used to predict the actual BI scores at discharge.

Results: A total of 313 individuals (men: 208; women: 105) were enrolled in the study. All the classification models outperformed single assessments in predicting the BI statuses of the patients at discharge. The performance of the LR and RF algorithms was higher (area under ROC curve (AUC): 0.79) than that of SVM algorithm (AUC: 0.77). In addition, the mean absolute errors of both SVM and linear regression models in predicting the actual BI score at discharge were 9.86 and 9.95, respectively.

Conclusions: The proposed ML-based method provides a promising and practical computer-assisted decision making tool for predicting ADL in clinical practice.

1. Introduction

In 2016, a new onset of stroke events occurred every 40 s in the United States, which caused a heavy disease burden on the health care system [1]. In Taiwan, in 2012, approximately 230 patients were hospitalized per day because of acute stroke [2]. Long-term disabilities in patients after stroke may create an enormous physical, mental, and financial burden for the patients, their families, and society. Early rehabilitation of patients improves recovery and reduces disabilities [3]. The ultimate goal of stroke rehabilitation is to ensure that the patients achieve a high quality of life by training them to perform activities of daily living (ADL) as independently as possible. In the initial stage of rehabilitation, accurate prediction of ADL for the subsequent few months is essential. Accurate prediction of ADL is crucial for optimizing stroke management during the first months following a stroke attack. Moreover, it guides realistic goal-setting, facilitates the creation of an early discharge plan, and provides accurate information to patients and their caregivers [3]. Based on accurate prediction of ADL, adequate socioeconomic support and health care resources can be effectively provided by either the government or the patient's family.

In Taiwan, a Post-acute Care-Cerebrovascular Diseases (PAC-CVD) program was implemented since 2014 to improve the quality of care by providing condensed rehabilitation programs to the patients with stroke [4,5]. Various assessments should be evaluated for patients of PAC-CAD program to provide more detailed information from various aspects [6]. The comprehensive assessments include modified Rankin scale (MRS),

* Corresponding author at: Department of Laboratory Medicine, Chang Gung Memorial Hospital at Linkou, No.5, Fuxing St., Guishan Dist., Taoyuan City 333, Taiwan.
** Corresponding author at: Department of Physical Medicine & Rehabilitation, Chang Gung Memorial Hospital at Taoyuan, School of Medicine, Chang Gung University, No.123, Dinghu Rd., Guishan Dist. Taoyuan City, Taiwan.
E-mail addresses: hsinyaowang@cgmh.org.tw (H.-Y. Wang), leonard@cgmh.org.tw (C.-K. Chen).
[1] These authors contributed equally to this manuscript.

https://doi.org/10.1016/j.ijmedinf.2018.01.002
Received 19 August 2017; Received in revised form 22 November 2017; Accepted 2 January 2018
1386-5056/

Barthel index (BI), functional oral intake scale (FOIS), mini nutrition assessment (MNA), European quality of life 5 dimensions questionnaire (QoL), instrumental activities of daily living scale (IADL), Berg balance test (BBT), gait speed, 6-min walk test (6MWT), Fugl–Meyer upper extremity assessment (FuglUE), modified Fugl–Meyer sensory assessment (FuglSEN), mini-mental state examination (MMSE), motor activity log amount of use (MAL, amount), motor activity log quality of use (MAL, quality), and concise Chinese aphasia test (CCAT). Briefly, the assessments could be categorized into the evaluations of four aspects. First, scales of FOIS [7] and MNA [8] could provide status of nutrition. Nutritional intake would be compromised in many stroke patients and decreased nutrition status may limit improvement of rehabilitation [9]. Assessments of BBT [10], gait speed [11], 6MWT [12], FuglUE [13], FuglSEN [14], and MAL [15] give the detailed information about motor/sensory function (both upper and lower extremities) and coordination [6]. Rehabilitation of extremities function is the basis for daily routines such as mobility, bathing or feeding. Third, CCAT [16] and MMSE [17] evaluate the aspect of communication abilities and cognition. Aphasia and cognition impairment are common complications of stroke [18]. CCAT and MMSE could help record and evaluate the improvement of rehabilitation on communication and cognition. Fourth, QoL [19], MRS [20], IADL [21], and BI [22] give the overall evaluation of disability degree, ADL, and life quality. Besides the specific use of individual assessments in clinical practice, the existing multiple assessments results may have other implicit value to the rehabilitation patients.

Multiple assessment features generate more comprehensive information regarding the statuses of patients than single assessment features. However, to date, an accurate and ready-to-use tool for prediction of ADL is not available because most prognostic studies are not of sufficiently high quality [3]. Moreover, a tool for accurate prediction of ADL is not available because the contribution of the features to prognosis of ADL is lacking for this group. To manage this problem, we proposed a machine learning (ML)-based method for predicting ADL. The BI score was used as the indicator of ADL in this study. The BI score is a well-known and widely used tool to assess the independence of functions of ADL, particularly for post-stroke patients [22,23]. Several studies have demonstrated the successful application of ML models in clinical practice [24–26]. In this study, the BI scores of the patients after a period of in-patient rehabilitation training could be accurately predicted using our ML-based approach.

2. Materials and methods

2.1. Patient eligibility

The Chang Gung Medical Foundation Institutional Review Board approved this retrospective cohort research (IRB no. 201700409B0). The cohort included 356 post-stroke patients who were admitted in the PAC-CVD rehabilitation program between March 2014 and December 2016. The patients' medical records were collected and deidentified before analysis. The inclusion criteria for this study were as follows: [1] stroke onset time within 1 month, [2] hemodynamic parameters stable within 72 h, [3] no neurological deterioration within 72 h, and [4] sufficient cognition function and ability to learn rehabilitation exercise, more specifically, modified Rankin Scale (MRS) between 2 and 4. The exclusion criteria included [1] stroke onset time > 1 month and [2] patients with end-stage renal disease or those receiving dialysis therapy. All eligible patients underwent comprehensive post-stroke assessments on admission and every 3 weeks after admission until discharge. The assessments included MRS, BI, FOIS, MNA, QoL, IADL, BBT, gait speed, 6MWT, FuglUE, FuglSEN, MMSE, MAL, and CCAT. All the assessments on admission into the rehabilitation ward were included as the explanatory features for classifying the severity of BI status or predicting the BI score at discharge from the rehabilitation ward. Moreover, the patients' age and length of stay in the acute stroke ward prior to admission to the PAC-CVD ward were also used as explanatory features.

2.2. Importance of each feature for classification models

The BI score at discharge was discretized into 3 ordinal classes, namely low BI status, medium BI status, and high BI status. The low BI status class indicated that BI score at discharge was below 60; Medium BI status class indicated that the BI score at discharge was between 60 and 90; High BI status class indicated that the BI score at discharge was more than 90. Total BI scores below 60, between 61 and 90, and between 91 and 99 indicated extreme dependence, moderate dependence, and slight dependence, respectively [27]. The discretization of "BI score at discharge" into "BI status at discharge" was illustrated in the supplementary material 1. A multivariable logistic regression analysis of the explanatory features was performed for evaluating the importance of each explanatory feature in classifying the BI status at discharge.

2.3. Model development

Patients satisfying the criteria were randomly distributed among 5 folds. We used a 5-fold cross-validation approach to train (4 folds) and test (one fold) the models. We used logistic regression (LR), support vector machine (SVM), and random forest (RF) to construct classification models. Multi-class models (i.e. high, medium, or low BI status) and binary models (e.g. low v.s. non-low BI status) were both developed and validated. The theoretical bases of these algorithms have been described in previous studies. LR measures the relationship between categorical dependent variables and one or more independent variables by using probability scores as the predicted values of the dependent variable [28]. Multinomial LR is used for BI statuses (multi-class) prediction. SVM is a data-mining method that constructs a classification model for a binary-class problem using nonlinear mapping to transform the data into a space of higher dimension. Through appropriate nonlinear mapping to a space of sufficiently high dimension, data from 2 classes are separated using a hyperplane [29]. We use the one-against-one-approach for constructing multi-class SVM models, and the appropriate class is found by a voting scheme [30]. RF, an ensemble classifier proposed by Breiman, comprises many classification trees, the bagging idea, and a random selection of features [31].

In addition to the classification of BI status at discharge, accuracy of the prediction models in predicting actual BI scores was also evaluated. Two prediction algorithms, namely SVM with linear kernel and linear regression were used. The models were constructed and tested using R (version 3.3.2, R Foundation for Statistical Computing, http://www.r-project.org/) with the caret package [32].

2.4. Evaluation methods

Model performance was evaluated through a 5-fold cross validation to ensure minimized bias. An unseen testing set (one fold) was used for performance evaluation after being trained by the training set (4 folds). The process would run 5 times iteratively for each 5-fold cross validation. Moreover, randomized case shuffling was performed prior to 5-fold cross validation. Case shuffling and 5-fold cross validation processes were repeated 10 times to achieve nonbiased evaluation. The area under the receiver operating characteristic (ROC) curve (AUC) was used for evaluating the performance of classification models. By contrast, root mean squared error (RMSE) and mean absolute error (MAE) were used as the indicators for evaluating the accuracy of the BI score prediction models.

2.5. Statistical analysis

Analysis of variance was used to compare continuous variables and determine performance differences among classifiers. The Tukey