

## **Abstract**

**Background:** Patient falls are a serious concern in hospital settings since it is the leading cause of human injuries and accounts for the biggest growth in medical spending. Fall risk assessment tools are preventive measures, but they do not have to signal which items are the most crucial factors for predicting falls. In addition, other fall precautions have not proven effective either. If we apply the health information technology including Machine Learning (ML) to this situation, there is a possibility that we can develop more effective, evidence-based, and personalized fall prevention strategies.

**Objectives:** This study aimed to identify the important risk factors for falls and preventive measures at St. Luke's International Hospital by analyzing large patient data and developing prediction models based on the essential factors identified via statistical and ML approaches, while also validating each via hold-out validation.

**Methods:** A single-institutional observational study was designed using Electronic Health Records (EHR) and an incident report. Data were collected from March 10, 2015, to March 31, 2021, for inpatients aged 50 to 106. The dataset contained a total of 284 predictors classified into intrinsic risk factors (259), extrinsic risk factors (3), and fall preventive measures (22). The primary outcome was the occurrence of falls during hospitalization. Classification And Regression Tree (CART) analysis-based logistic regression model and the Lasso logistic Regression analysis-based logistic regression model were applied to patient data to identify important risk factors and preventive measures. We developed predictive models using these statistical and ML methods. These predictive models were compared with the logistic

regression model for the current fall assessment sheet used in clinical practice via hold-out validation. The sensitivity, specificity, and Area Under the Curve (AUC) were computed.

**Results:** CART analysis by setting the complexity parameter as a value of 0.001 identified 79 variables as important, and the CART analysis-based logistic regression model showed that 34 variables (24 intrinsic risk factors, five extrinsic factors, and five preventive measures) were statistically significant. In contrast, the Lasso logistic regression with all data identified only 11 variables as important, and the Lasso logistic regression analysis-based logistic regression model showed nine variables (seven intrinsic risk factors, one extrinsic risk factor, and one preventive measure) were statistically significant. Regarding the AUCs using test dataset, CART analysis-based logistic regression model with all variables (0.899) and the Lasso logistic regression analysis-based logistic regression model with all variables (0.872) outperformed the CART analysis-based logistic regression model with only risk factors (0.886) and the logistic regression model based on the current fall assessment sheet (0.737). Additionally, in the DeLong's test for comparing the AUCs, there was a significant difference between the lasso logistic regression analysis-based logistic regression model and the logistic regression model based on the current fall assessment sheet ( $p < 0.001$ ).

**Conclusions:** This study identified important risk factors and fall prevention measures related to falls in several hospital wards. The results can help nurses develop fall prevention strategies and decision-making in improving medical safety practices. We also found that the lasso logistic regression analysis-based logistic regression model could predict the outcome with the same accuracy as the other models by fewer variables. This study clarified that the risk factors

obtained from the lasso logistic regression are important to predict inpatient falls of patients in clinical practice.

***Keywords: Accidental Fall, Assessment Sheet, AUC, CART Analysis, LASSO, Logistic Regression Analysis, Nurses, Preventive measures***