



The 65th ASH Annual Meeting Abstracts

ORAL ABSTRACTS

902.HEALTH SERVICES AND QUALITY IMPROVEMENT - LYMPHOID MALIGNANCIES

A Machine Learning Approach for Predicting the Occurrence of Asparaginase-Associated Pancreatitis in Pediatric Patients with Acute Lymphoblastic Leukemia

Suhyun Yoon¹, Hyery Kim, MD², Dongbin Youk³, Jong keon Song³, Sung Han Kang, MD³, Kyung-Nam Koh, MD PhD⁴, Ho Joon Im, MD PhD⁴

¹Asan Medical Center/1Department of Pediatrics, University of Ulsan College of Medicine, Seoul, Korea, Republic of (South)

²Asan Medical Center/Department of Pediatrics, Seoul National Univ. College of Med., SMG-SNU Borame Med. Ctr., Seoul, South Korea

³Asan Medical Center/Department of Pediatrics, University of Ulsan College of Medicine, Seoul, Korea, Republic of (South)

⁴Department of Pediatrics, Asan Medical Center Children's Hospital, University of Ulsan College of Medicine, Seoul, Korea, Republic of (South)

Introduction

Children being treated for acute lymphoblastic leukemia (ALL) are frequently affected by asparaginase-associated pancreatitis. Additionally, pancreatitis is among the most troublesome and frequent side effects of asparaginase therapy and is a significant contributor to early drug discontinuation and poor outcomes. There are inadequate odds ratios for known risk factors, such as asparaginase dosage, advanced age, and single nucleotide polymorphisms, to predict pancreatitis occurrence. The goal of this study was to use machine learning to develop a predictive model for asparaginase-induced pancreatitis in pediatric ALL patients.

Methods

Data were collected from 711 patients who had childhood ALL and received asparaginase. Pancreatitis was defined as serum amylase and/or lipase levels greater than three times the upper limit of normal or acute pancreatitis on abdominal images. One month from the time of asparaginase administration for each patient was defined as one "timestep", and when asparaginase was administered thereafter, it was defined as a new individual timestep. Each timestep was defined as one training case, and a case in which pancreatitis occurred at that timestep was defined as an event. Finally, 3193 training cases were defined in a total of 711 patients. The physical measurement results, prescription codes, blood test results, and blood transfusion history data were collected from electronic health records (HER) during the entire treatment period of the patients. Among these are age, body mass index, body surface area, gender, type of asparaginase (Native, Erwinia, or Pegylated), previous history of pancreatitis, cumulative number of asparaginase administrations, and asparaginase change history before the current time point. The results of 47 blood tests on the start date of asparaginase in each timestep were also used as predictive variables (Figure 1). Using logistic regression, Random forest, and XGBoost as machine learning methods, we assessed a model predicting asparaginase-associated pancreatitis through 5-fold cross-validation. Performance indicators such as area under the receiver operating characteristic curve (AUC) score, Precision Recall (PR) score, F0.5 score, and F2 score were employed to evaluate the binary classification of imbalanced data. The selection of the model was determined based on these two criteria.

Results

When considering the $F(0.5+F2.0)/2$ score as the basis for model selection, the logistic regression model demonstrated an AUC of 81% (PR 32.86%, $F(0.5+F2.0)/2$ score 23.23%). On the other hand, the XGboost model exhibited an AUC of 79% (PR 33.7%, $F(0.5+F2.0)/2$ score 32.07%), while the Random Forest model achieved an AUC of 84% (PR 33.34%, $F(0.5+F2.0)/2$ score 39.48%). Among these models, the Random Forest model demonstrated the highest predictive power.

When the model was chosen using the PR score, the logistic regression model achieved an AUC of 80% (PR 34.97%, $F(0.5+F2.0)/2$ score 22.08%), whereas the XGboost model achieved an AUC of 79% (PR 31.58%, $F(0.5+F2.0)/2$ score 31.6%). Also, it was seen that the Random Forest model had the best performance across all metrics, with an AUC of 85%, a precision-recall (PR) score of 32.26%, and a $F(0.5+F2.0)/2$ score of 36.4% (Figure 2, left).

According to Shapley values, it is evident that some parameters, namely greater lipase levels, higher cumulative asparaginase dosages, higher amylase levels, higher glucose levels, and older age, have significantly contributed to the occurrence of asparaginase-associated pancreatitis (Figure 2, right).

Conclusions

A machine learning model was employed to successfully forecast the occurrence of acute pancreatitis following the administration of asparaginase in pediatric patients with AAL. This study specifically focused on making predictions regarding pancreatitis within a month based on the test results obtained at the commencement of asparaginase treatment. This approach offers the potential for promptly predicting the development of pancreatitis. In further stages, following external validation and prospective observational clinical trials, the prediction model has the potential to be included in the EHR and serve as a Clinical Decision Support System (CDSS).

Disclosures No relevant conflicts of interest to declare.

Figure 1. Data scheme for prediction model

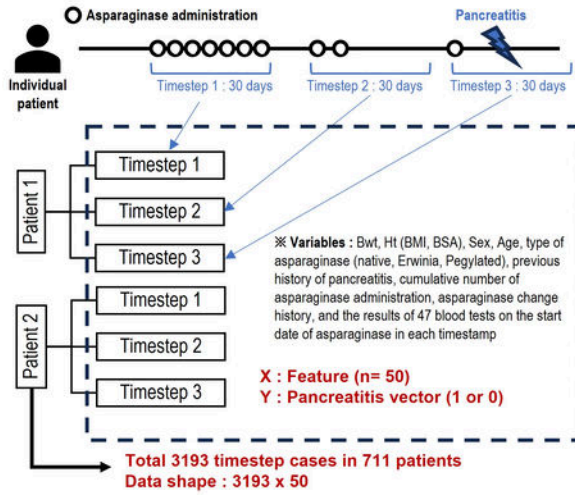


Figure 2. Best model from Random Forest.

AUC & PR curves (left), Feature importance map by SHAP value (right)

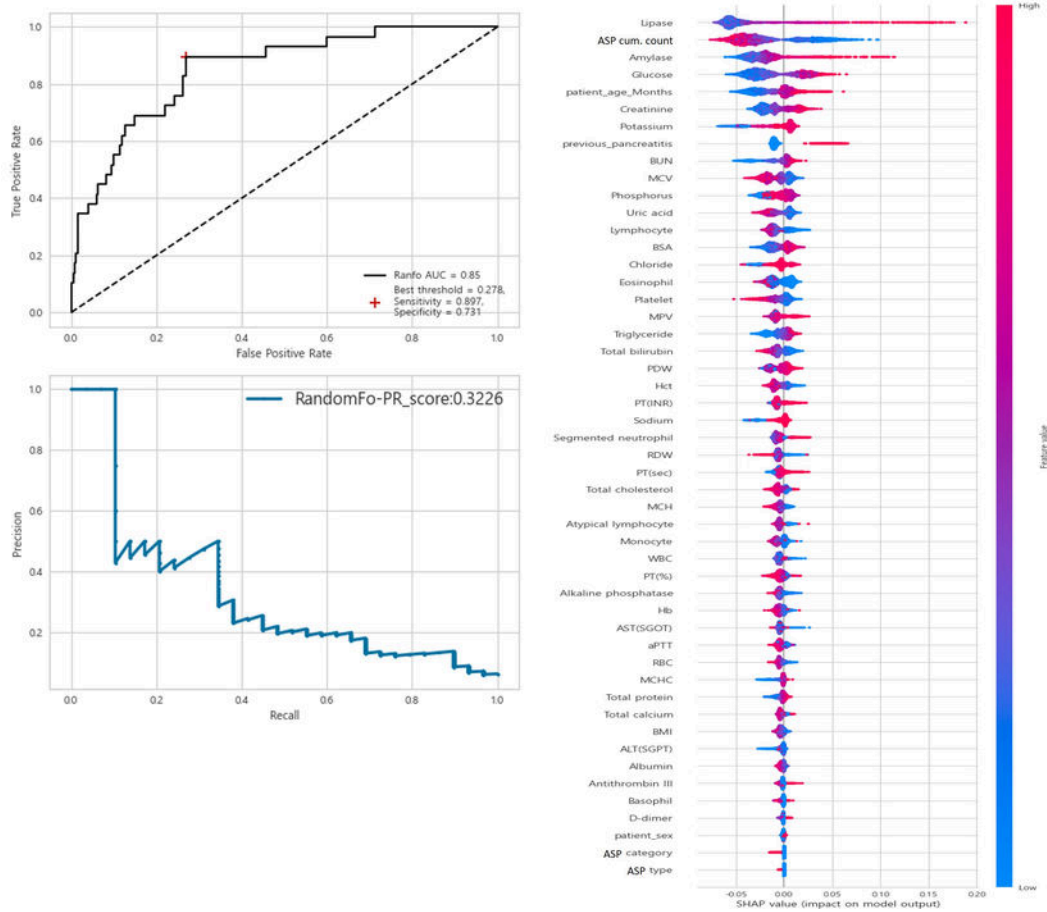


Figure 1

<https://doi.org/10.1182/blood-2023-191154>