Machine Learning for the Prediction of Progression in Patients with Acute Kidney Injury in Critical Care

Lifan Zhang
Sichuan University West China Hospital https://orcid.org/0000-0001-5182-3814

Canzheng Wei
West China Hospital of Medicine: Sichuan University West China Hospital

Yunxia Feng
Mianyang Central Hospital

Aijia Ma
West China Hospital of Medicine: Sichuan University West China Hospital

Yan Kang (Kangyan@scu.edu.cn)
https://orcid.org/0000-0003-4968-1217

Research

Keywords: Acute kidney injury, Critical care, Logistic Models, Extreme gradient boosting

Posted Date: April 26th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-412422/v1

Abstract

Background: Acute kidney injury (AKI) is a serve and harmful syndrome in the intensive care unit. The purpose of this study is to develop a prediction model that predict whether patients with AKI stage 1/2 will progress to AKI stage 3.

Methods: Patients with AKI stage 1/2, when they were first diagnosed with AKI in the Medical Information Mart for Intensive Care (MIMIC-III), were included. We excluded patients who had underwent RRT or progressed to AKI stage 3 within 72 hours of the first AKI diagnosis. We also excluded patients with chronic kidney disease (CKD). We used the Logistic regression and machine learning extreme gradient boosting (XGBoost) to build two models which can predict patients who will progress to AKI stage 3. Established models were evaluated by cross-validation, receiver operating characteristic curve (ROC), and precision-recall curves (PRC).

Results: We included 25711 patients, of whom 2130 (8.3%) progressed to AKI stage 3. Creatinine, multiple organ failure syndromes (MODS), blood urea nitrogen (BUN), sepsis, and respiratory failure were the most important in AKI progression prediction. The XGBoost model has a better performance than the Logistic regression model on predicting AKI stage 3 progression (AU-ROC, 0.926; 95%CI, 0.917 to 0.931 vs. 0.784; 95%CI, 0.771 to 0.796, respectively).

Conclusions: The XGboost model can better identify patients with AKI progression than Logistic regression model. Machine learning techniques may improve predictive modeling in medical research.

Keywords: Acute kidney injury; Critical care; Logistic Models; Extreme gradient boosting

Full Text

This preprint is available for download as a PDF.

Figures

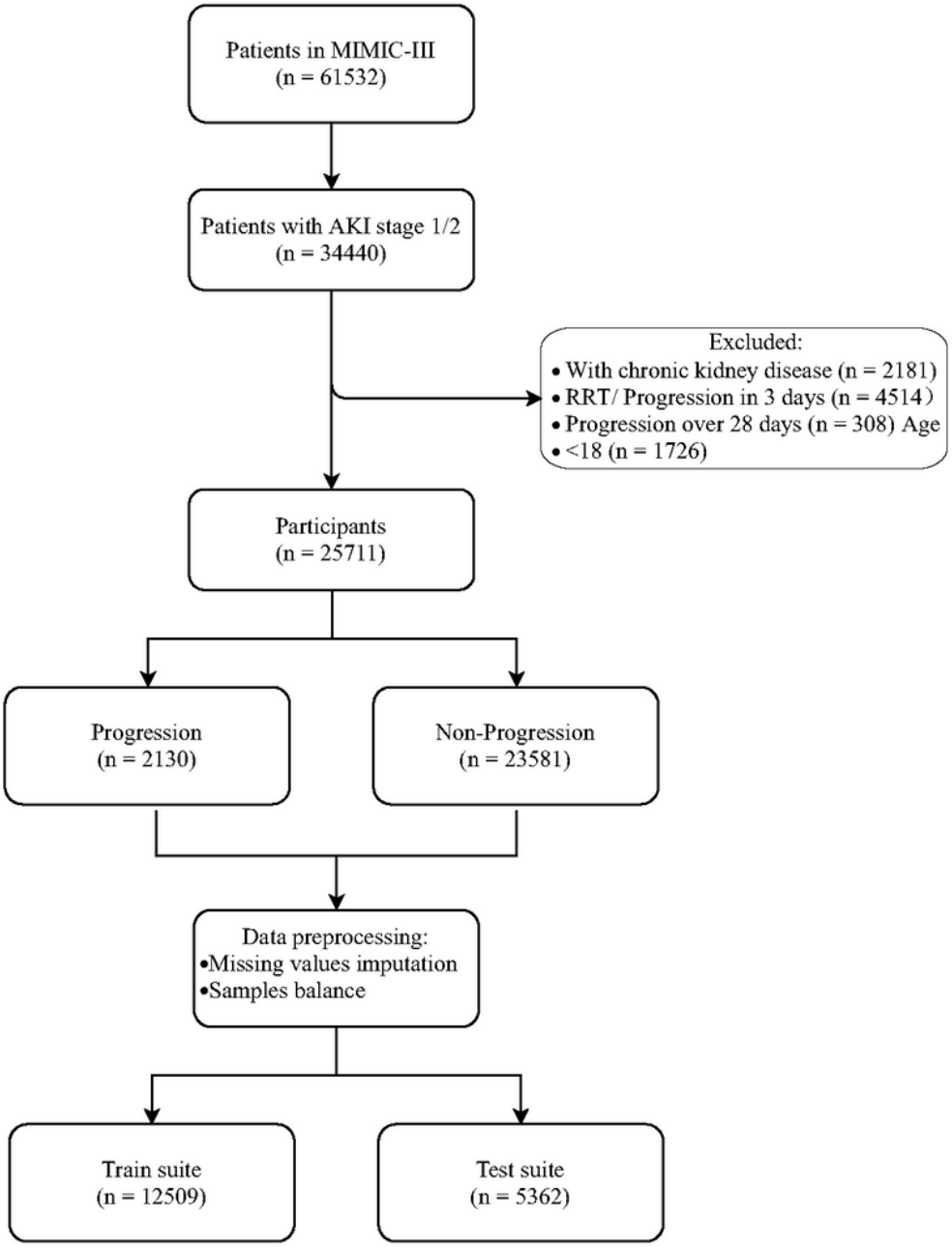

Figure 1

Flow chart of patient selection and data processing

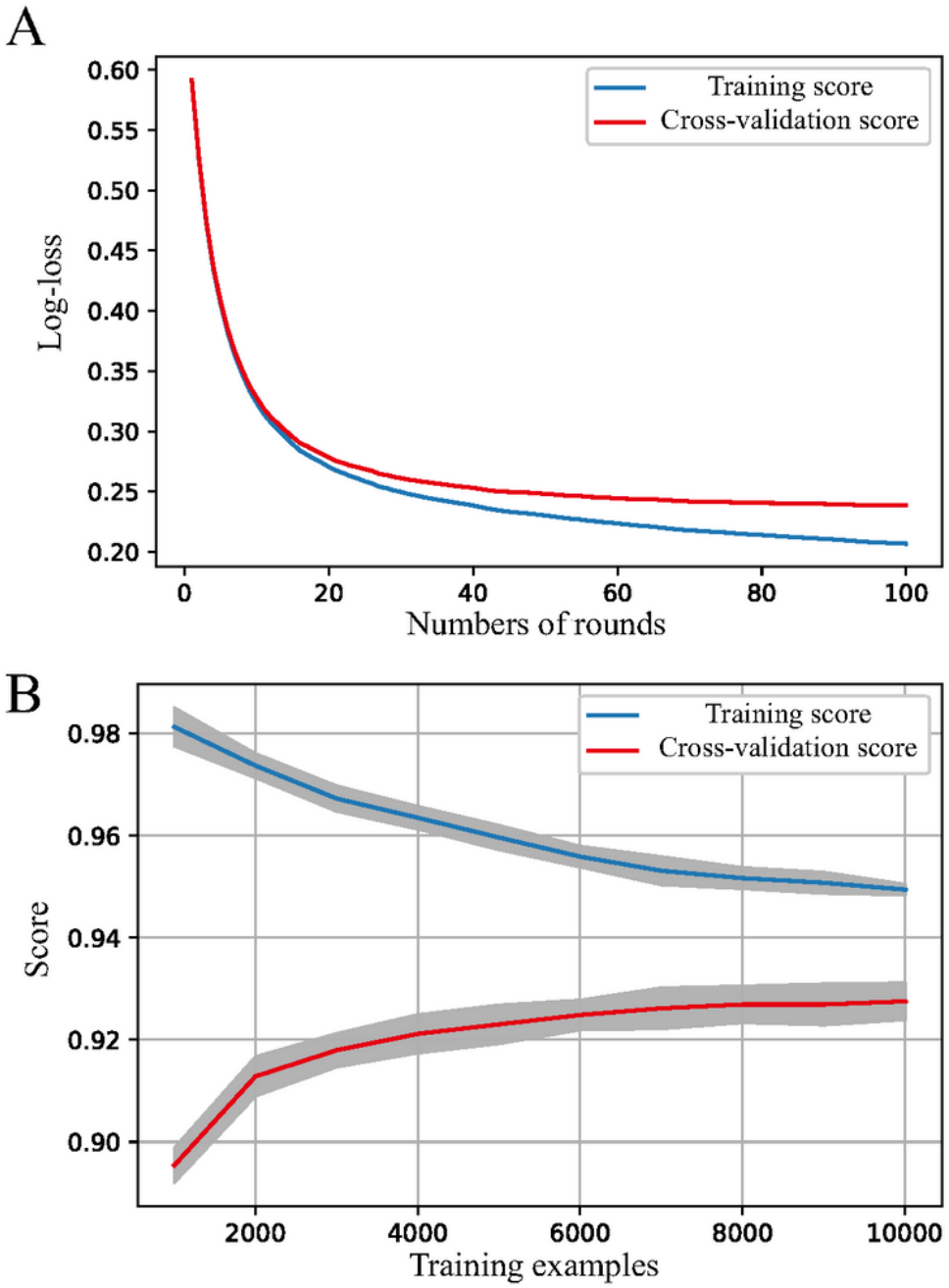

Figure 2

The training process of the extreme gradient boosting machine A: Cross-validation during XGBoost hyperparameter tuning. The log-loss value for the training and testing datasets is shown in the vertical axis. The dashed vertical line indicates the number of rounds with the minimum log-loss in the test sample. B: Learning curve of the XGBoost model after hyperparameter tuning. AU-ROC value for the testing and training datasets is shown in the vertical axis. With the subsample ratio increasing, AU-ROC of

training datasets decreases, and AU-ROC of testing datasets increases. The training score is always higher than the test score

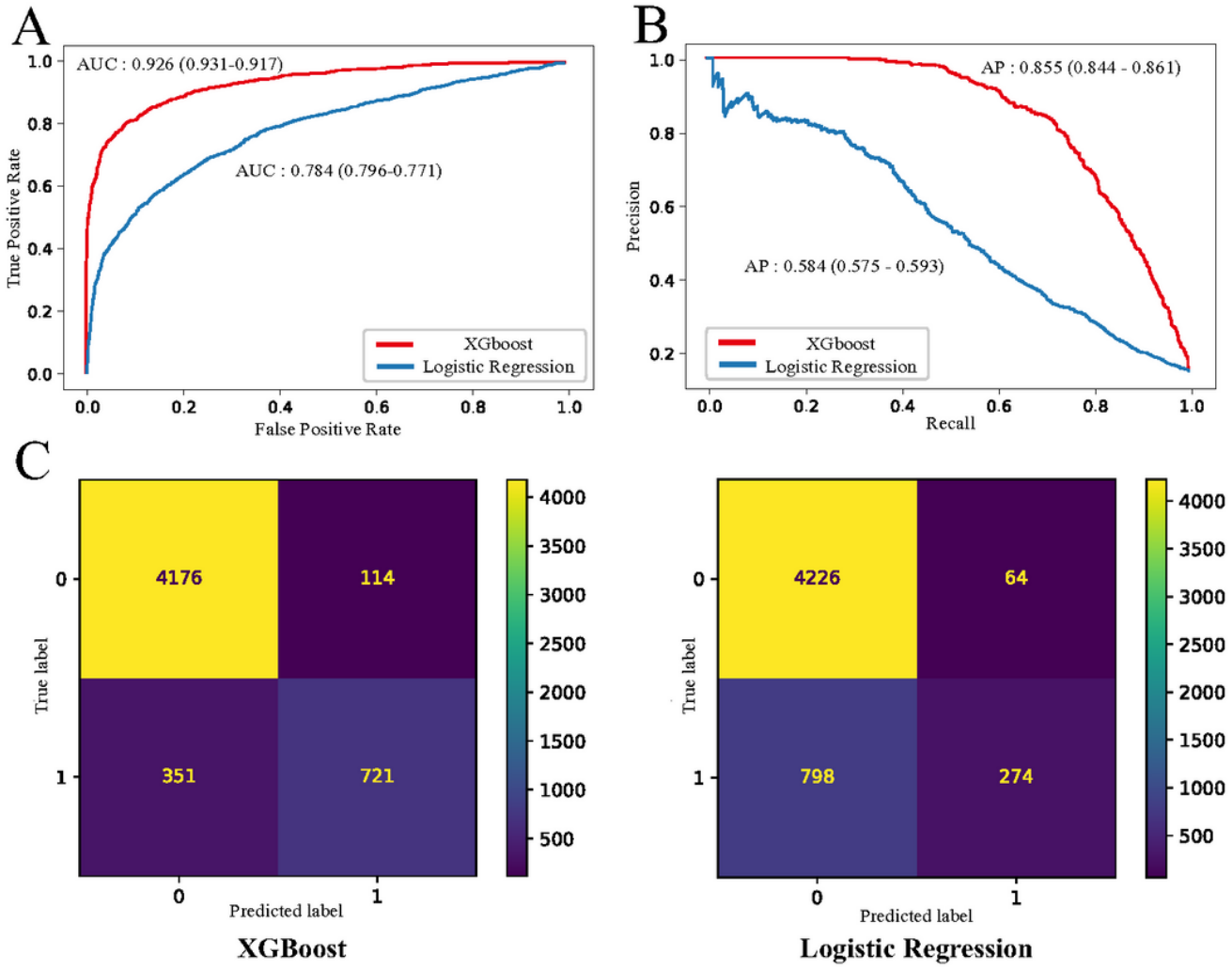

Figure 3

Performance of the XGBoost and Logistic regression model A: Receiver operating characteristic curve for estimating the discrimination between the Logistic regression model and the XGBoost model. B: Precision-Recall curve for estimating the discrimination between the Logistic regression model and the XGBoost model. C: Confusion matrix of the Logistic regression model and XGBoost model. The color represents the number of patients. Whether progress is represented by numbers.

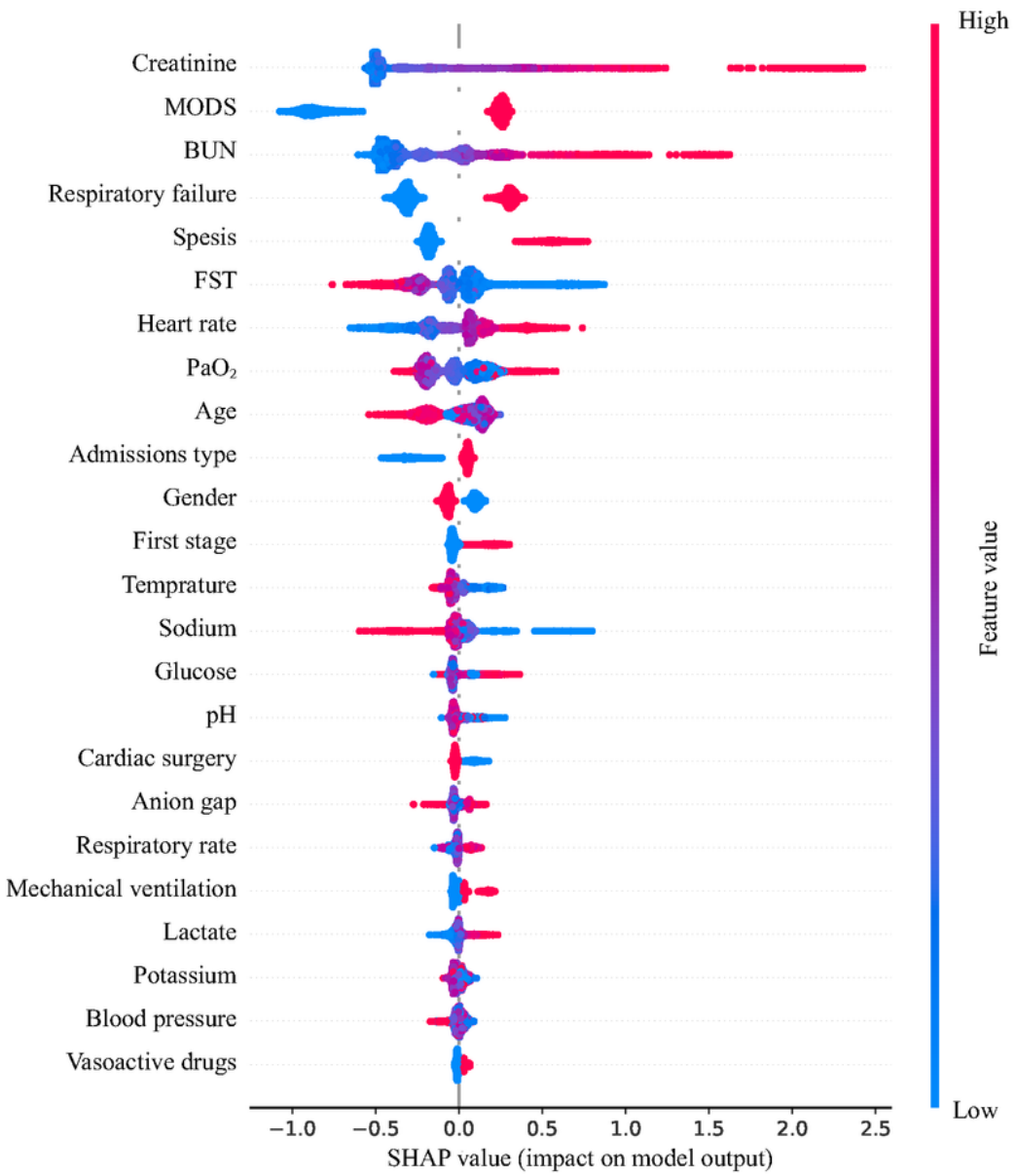

Figure 4

SHAP value of XGBoost model output SHAP value of all patient output. Each point represents a variable for an observation. The color of the point is determined by its relative height in the variable. The blue represents lower and the red represents higher.

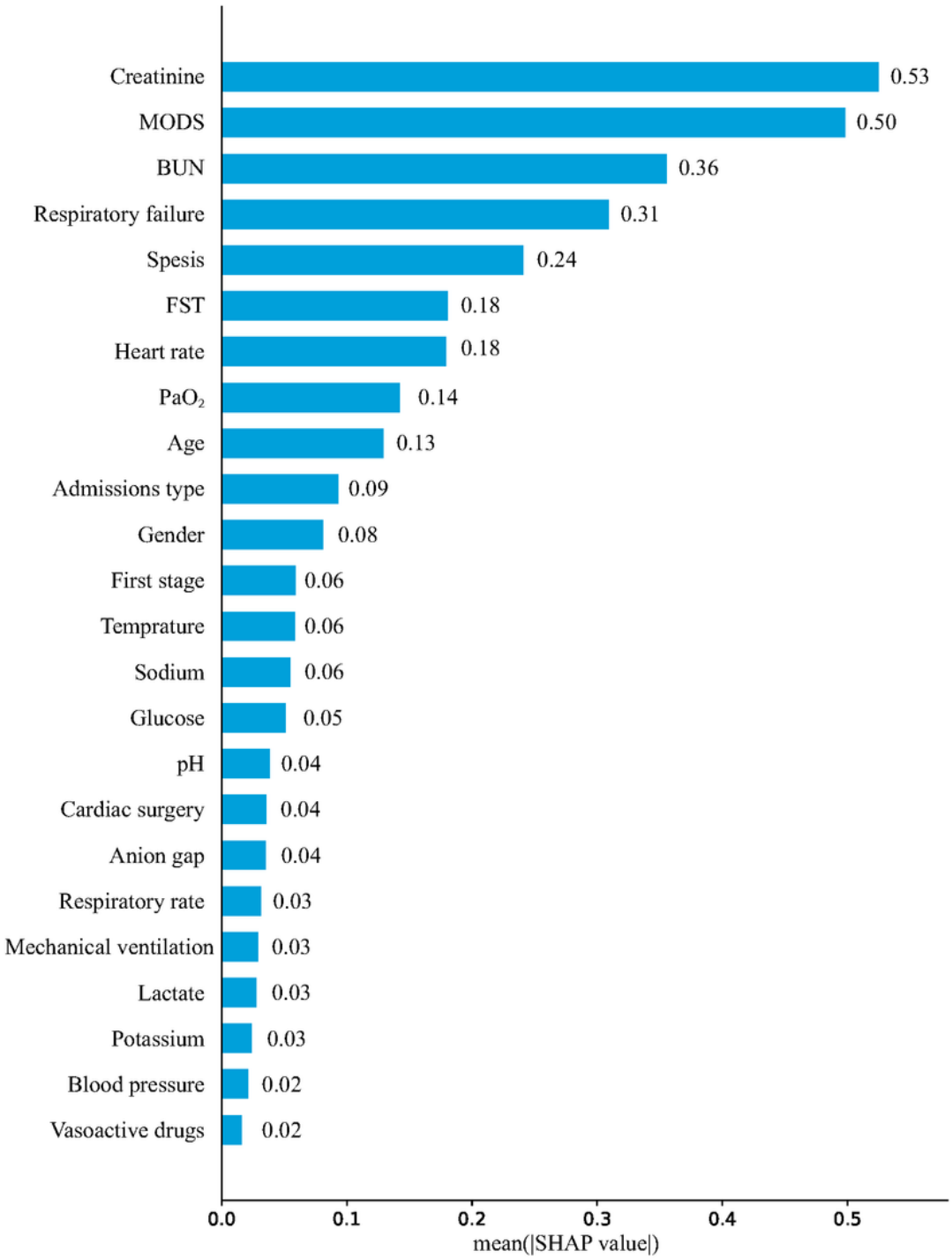

Figure 5

Feature importance derived from the XGBoost model Feature importance was calculated by the mean contribution of every observation, which is equal to the traditional method. Abbreviations and annotations: BUN, blood urea nitrogen; FST, furosemide stress test; MODS, multiple organ failure syndromes

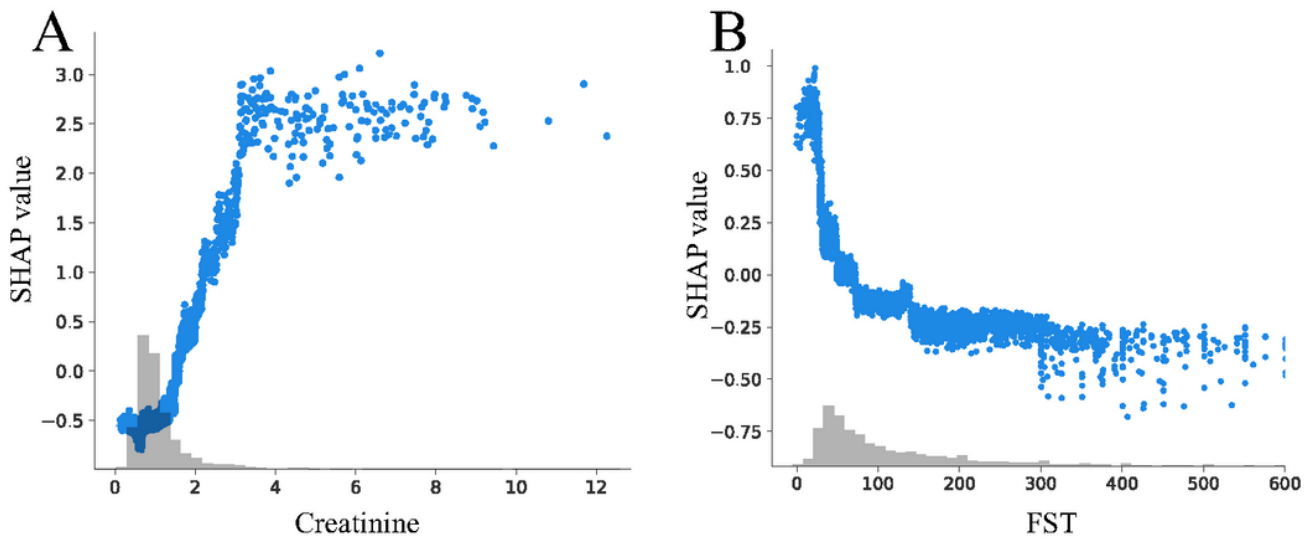

Figure 6

SHAP value for single variable A: SHAP value for creatinine. SHAP value increases with the increase of creatinine until creatinine probably reaches 5 mg/dL. B: SHAP value for FST. SHAP value decreases with the increase of FST until FST reach 100ml/h.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.docx