**Methods(Abstract)**

We retrospectively studied ,,, in MIMIC-III database. Diuretic resist was defined as,,,. Patients were randomly divided into an 75%/25% split for training and validation cohorts, trained modeled were validated using the validation cohort. Random forest models were constructed to predict whether patients would have RRT using 20 variables chosen from 100 clinical variables after lasso regression. Areas under the receiver operating characteristic curves (AUC) were used to quantify the predictive performance.

**Measurements and results**

**Data**

**Feature selection**

**Statistical analysis**

**...**

We applied random forest (RF) to predict whether patients need RRT in the end. The model was trained and tested on a randomly partitioned 75% 25% split of the data. Basic characteristics of the two cohorts were showed in Table 1.

In order to evaluate our models and select variables, we used the 10-fold cross-validation[1] evaluation methods where the data were randomly partitioned into 10 mutually exclusive subsets with approximately equal size. The testing operation was repeated 10 times where 9 folds of the subsets were used as training set and the other as test set at the evaluation iteration. In general, this method gives a lower variance than a single hold-out set evaluator[2].

Two models were developed using RF, the first model contained variables measured on admission and the first measurements after diuretic resist. Variable selection was conducted using lasso regression and RF, finally 20 most important variables were selected and taken into the final model. The second model contained variables selected by step regression.

We chose area under receiver operating curve (AUC) as our primary performance measure for RRT and AUC were reported only for validation dataset.

Data extraction was performed using pgAdmin4 and all analyses were conducted in R 3.4.3

**Results**

**...**

AUC of validation cohorts for model 1 was 88.5%.

...

AUC of validation cohorts for model 2 was 87.5%.

1. Prasad AM, Iverson LR, Liaw A. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems. 2006;9(2):181–99.

2. Sakr, Sherif, Radwa Elshawi, Amjad M. Ahmed, Waqas T. Qureshi, Clinton A. Brawner, Steven J. Keteyian, Michael J. Blaha, and Mouaz H. Al-Mallah. “Comparison of Machine Learning Techniques to Predict All-Cause Mortality Using Fitness Data: The Henry Ford exercIse Testing (FIT) Project.” *BMC Medical Informatics and Decision Making* 17, no. 1 (2017): 1–15. <https://doi.org/10.1186/s12911-017-0566-6.>