**Analysis of Diuretic Resist Project**

**Date : 6/3/2018**

**Author : Yuan Zhang**

**Cohort:**

1.  Patient in MIMIC with heart failure, >18                  -----13568

2. age >18                    -----13426

3. Left Ventricular Systolic Function:  Moderate or Severe            -----2826

4. diuretic resistance drug >2units:                         -----1260

3. exclude ESRD                                                        -----1217

**Re-admission:**

313 in 1217 (25.72%)

While re-admission in 30 days: 60 in 1217 patients(4.93%)

re-admission in 60 days: 117 in 1217 patients(9.61%)

**Variables:**

a .Demographic

b . comobidity and drug use

c . fluid variables

d . survival info

e . lab test and vital signs on four timestamp:

    --    1 ) First Time after Admit

    --    2 ) First Time after DR

    --    3 ) Last Time Before Dialysis

    --    4 ) Last Time Before Discharge

**Note that:**

**The third timestamp for no-rrt patients was defined as:**

patient urine output increase after diuretic resist

  "非rrt患者增加到最大利尿剂后尿量开始增多的时间点. 尿量开始增多:尿量比前一天多且液体出量大于入量,或体重较前下降"

**Prediction**

Consider 2 models:

model 1 : contain variables on admission and DR

model 2 : contain variables of step regression

(result are shown in result folder)

...

Comparing some models, We finally decide to use these two. Results are shown below.

***model 1:***

*Variables: 'age', 'gender', 'add\_drug\_therapy', 'change\_drug\_therapy', 'height', 'weight\_adm', 'weight\_after\_dr',*

*'bbr', 'inotropes\_vasopressor', 'acei', 'mra', 'vasodilator',*

*'ischemic\_heart\_disease', 'cardiomyopathies', 'valvular\_disease', 'atrial\_fibrilation', 'hyperlipidemia', 'hypertension', 'diabetes', 'sleep\_disordered\_breathing', 'renal\_failure', 'anemia', 'infection', 'alcohol\_abuse',*

*'hr\_adm', 'hr\_dr', 'sbp\_adm', 'sbp\_dr', 'dbp\_adm', 'dbp\_dr', 'spo2\_adm', 'spo2\_dr', 'tem\_adm', 'tem\_dr',*

*'base\_excess\_adm', 'bun\_adm', 'bun\_dr', 'creatinine\_adm', 'creatinine\_dr', 'sodium\_adm', 'sodium\_dr', 'potassium\_adm', 'potassium\_dr', 'hemoglobin\_adm', 'hemoglobin\_dr', 'hco3\_adm', 'hco3\_dr', 'gfr\_adm', 'gfr\_dr', 'ph\_adm', 'ph\_dr', 'albumin\_adm', 'albumin\_dr'*

list 20 most important variables of the model:



(ordered by mean accurance)



(ordered by Gini index)

AUC of validation cohorts for model 1 was 87.1%(80.3%-93.8%)

***Model 2:***

*Variables: creatinine\_dr, gfr\_dr, spo2\_adm, age, albumin\_dr, acei, hyperlipidemia*



(ordered by mean accurance)



(ordered by Gini index)

AUC of validation cohorts for model 2 was 88.3%(82.1%-94.4%)

\*Note：AUC（area under the curve），指的是ROC曲线下的面积。它是一个模型评价指标，只能用于二分类模型的评价。AUC越大越好。

\* Note: 在随机森林模型里面，变量的重要性有两种度量指标，一种是：decrease of accuracy， 一种是decrease of node impurity（节点不纯度）。你可以选一个放在论文里（第二个比较合理，因为我们的数据不平衡，看accuracy会有偏差）。如果选第二个，可以写：The variable importance measure is measured by the Gini index.

**Apply linear regression for secondary outcomes:**

LOS hospital, LOS ICU, GFR, total fluid input to discharge, total fluid output to discharge, fluid overload to discharge, urine output to discharge

The results are :

|  |  |  |
| --- | --- | --- |
| RRT’ effect on secondary outcomes | | |
|  | **coef(95%CI)** | **p value** |
| GFR | 1.27 ( -3.68 , 6.21 ) | 0.6155 |
| Fluid output to discharge | 11873.03 ( 7059.12 , 16686.93 ) | < 0.0001 |
| Fluid input to discharge | 2872.87 ( -5059.38 , 10805.12 ) | 0.4774 |
| Fluid overload to discharge | -7153.55 ( -13488.24 , -818.86 ) | 0.0269 |
| Urine output to discharge | -6368.12 ( -10233.3 , -2502.95 ) | 0.0013 |
| LOS hospital | 7.24 ( 4.42 , 10.07 ) | < 0.0001 |
| LOS ICU | 6.33 ( 4.33 , 8.32 ) | < 0.0001 |
|  | **OR(95%CI)** | **p value** |
| Hospital expire flag | 1.34 ( 0.66 , 2.72 ) | 0.4216 |

\*Use linear regression for continuous variables

\*Use logistic regression for binary variables(hospital expire)

\*Apply step regression to select variables for all these models