Online Supplement

Development and Validation of Machine Learning-Based Models for Prediction of Intensive Care Unit Admission and In-Hospital Mortality in Patients with Acute Exacerbations of Chronic Obstructive Pulmonary Disease

Qinyao Jia[1,*] Yao Chen[2*] Qiang Zen[3*] Shaoping Chen[4] Shengming Liu[5] Tao Wang[6#] XinQi Yuan[7#]

*These authors contributed equally to this work and should be considered co-first authors.

[1]School of Pharmacy, North Sichuan Medical College, Nanchong, China

[2]Department of Tuberculosis, Chengdu Public Health Clinical Medical Center, Chengdu, China

[3]Department of Pulmonary and Critical Care Medicine, The Third Hospital of Mianyang, Sichuan Mental Health Center, Mianyang, China

[4]Department of Pulmonary and Critical Care Medicine, Affiliated Hospital of North Sichuan Medical College, Nanchong, China

[5]Department of Pulmonary and Critical Care Medicine, The First Affiliated Hospital of Jinan University, Guangzhou, China

[6]Department of Pulmonary and Critical Care Medicine, University of Chinese Academy of Sciences Shenzhen Hospital, Shenzhen, China

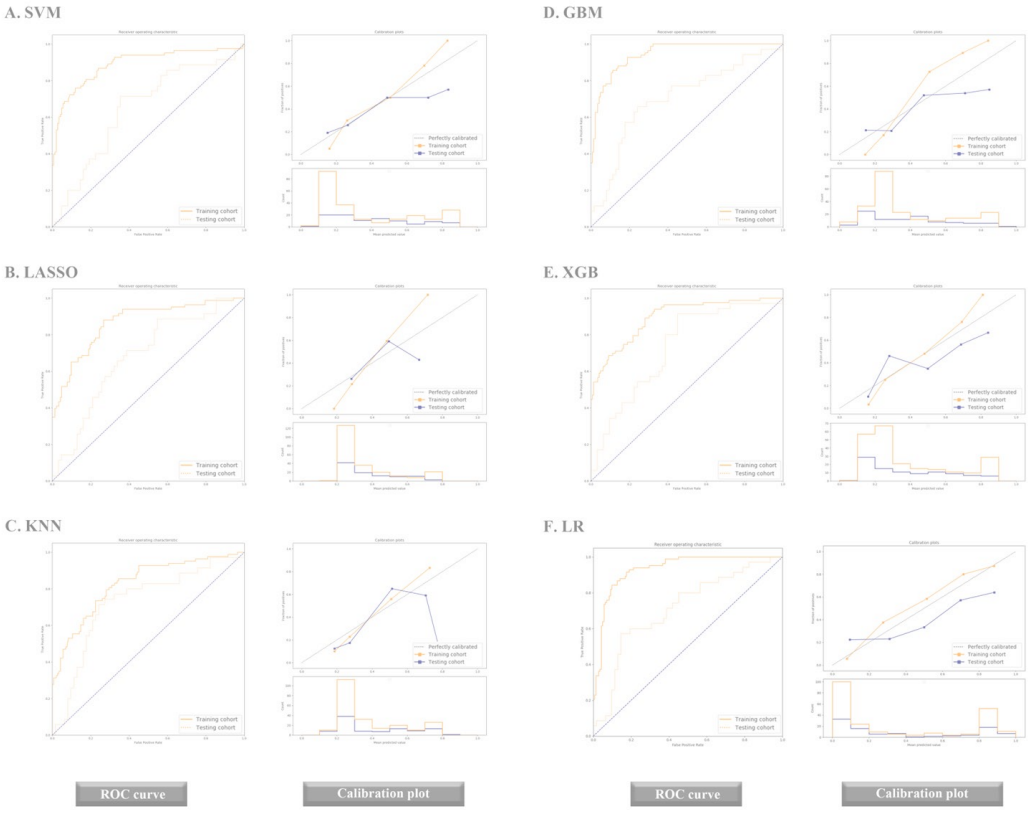
[7]Department of Pulmonary and Critical Care Medicine, The Fifth People's Hospital of Sichuan Province, Chengdu, China

Supplementary table 1 The list of 90 features for present study.

Gender	Length of stay	PE situation	Age	Smoking
Temperature (℃)	Pulse(times/min)	Breathing (times/min)	Systolic pressure(mmHg)	Diastolic pressure(mmHg)
Height (cm)	Weight (kg)	Hb (g/L)	HCT	PLT (10^{-9}/L)
LYMPH (*10^{-9}/L)	WBC (*10^{-9}/L)	N%	pH	PCO_2 (mmHg)
PO_2 (mmHg)	SpO_2 (%)	Oxygen saturation	INR fibrinogen	D-dimer (mg/dlt)
TBIL (umol/l)	ALT(IU/L)	AST (IU/L)	Albumin (g/L)	Globulin (g/L)
Antithrombin III(%)	Blood glucose (mmol/l)	Urea (mmol/l)	Creatinine (umol/l)	Uric acid (umol/l)
Triglyceride (mmol/L)	Cholesterol (mmol/L)	Creatinekinase (IU/L)	LDH (IU/L)	Serum Na (mmol/l)
Myoglobin (ng/ml)	Creatine kinase isoenzyme (ng/ml)	Troponin (ng/ml)	＞BNP (ng/ml)	Serum K (mmol/l)
History of inflammatory bowel disease	Anti-platelet therapy	varicose veins	severe lung disease	COPD
malignant tumor (previous history)	malignant tumor(current disease)	cerebral apoplexy (within one month)	Rheumatic disease	Craniocerebral trauma
Atrial fibrillation	Asthma	Interstitial lung disease	Respiratory failure	Bronchiectasis
Pulmonary heart disease	Pulmonary arterial hypertension	Obstructive Sleep Apnea-Hypopnea Syndrome,	Hypertension	Coronary heart disease
Cardiomyopathy	Peripheral vascular disease	Congenital heart disease	Rheumatic heart disease	Chronic hepatitis
Liver cirrhosis	Alimentary tract hemorrhage	Acute coronary syndrome	Chronic gastritis	Chronic nephritis
Nephrotic syndrome	Chronic renal insufficiency	Acute renal insufficiency	Myelodysplastic syndrome	Paroxysmal nocturnal hemoglobinuria
HIV	Anti-coagulant therapy	Drinking history	Dementia	Parkinson’s disease
Dyslipidemia	Hyperthyroidism	Hypothyroidism	Diabetes	BMI

Figure.S1. Evaluation of prediction performance of seven ML models in predicting ICU admission

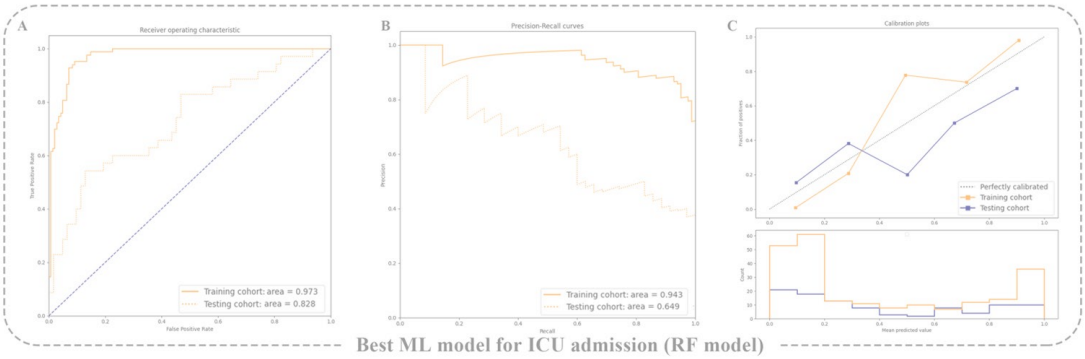
on patients of training and test sets.



Abbreviations: AUC, Area under receiver operating characteristic curve; ROC, Receiver operating characteristic; LR, logistic regression; LASSO, least absolute shrinkage and selection operator; SVM, support vector machine; KNN, K-Nearest Neighbor; RF, random forest; GBM, gradient boosting machine; XGB, extreme gradient boosting.

Figure.S2: Assessment of prediction performances for best ML model (random forest) in

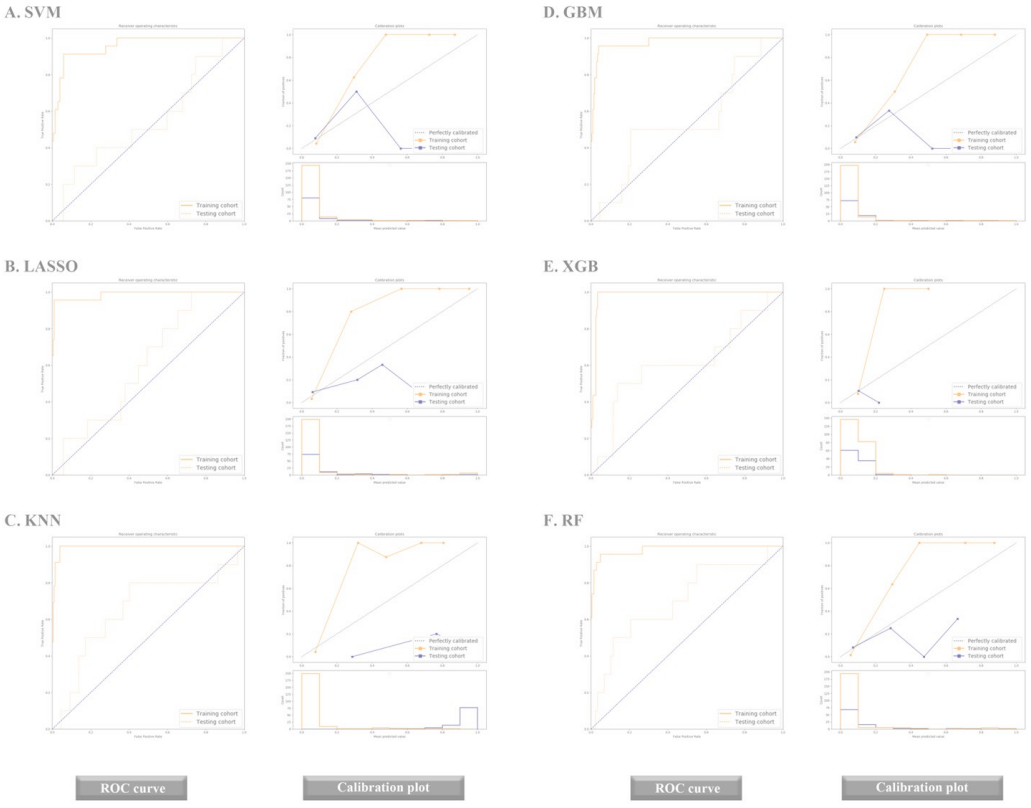
classifying ICU admission, including ROC curves (Left), Precision-recall plots (Middle), and Curves of calibration analysis (Right) for training and test sets, separately.



Abbreviations: AUC, Area under receiver operating characteristic curve; ROC, Receiver operating characteristic; ML, machine learning.

Figure.S3. Evaluation of prediction performances of seven ML models in predicting in-hospital

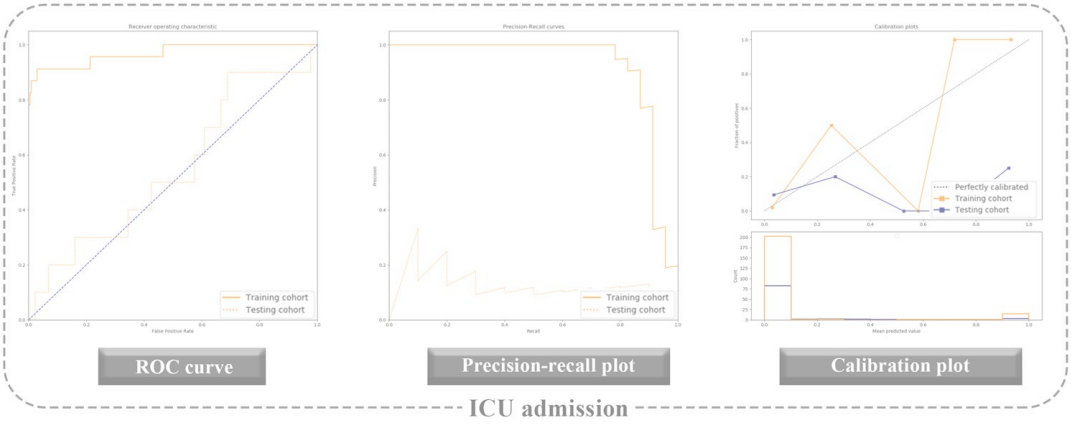
mortality on patients of training and test sets.



Abbreviations: AUC, Area under receiver operating characteristic curve; ROC, Receiver operating characteristic; LR, logistic regression; LASSO, least absolute shrinkage and selection operator; SVM, support vector machine; KNN, K-Nearest Neighbor; RF, random forest; GBM, gradient boosting machine; XGB, extreme gradient boosting.

Figure.S4. Evaluation of the predictive performance of the occurrence of ICU admission in the

prediction of in-hospital mortality on patients of training and test sets.



Abbreviations: AUC, Area under receiver operating characteristic curve; ROC, Receiver operating characteristic.