北京大学第一医院 北京大学泌尿外科研究所 Institute of Urology, Peking University

传统算法和机器学习算法的尿培养预测模型对比：一项基于泌尿系结石患者的多中心回顾性研究

何宇辉[1]，彭畔新[2]，应文伟[2]，王庆伟[3]，王焱[3]、刘贤奎[4]、宋文辉[5]、高跃[5]、李沛哲[2]、王杰[1]、朱伟杰[1]、周晓峰[2*]、李学松[1*] 、周利群[1*]

[1] 北京大学第一医院
[2] 中日友好医院
[3] 郑州大学第一附属医院
[4] 中国医科大学附属第一医院
[5] 天津市第一中心医院
* 通讯作者

Introduction

- The ability to quickly and accurately identify urinary stone patients with positive urinary cultures, who sometimes need aggressive antibiotic management in a timely manner, is a major challenge for urologists.
- As a gold standard, urine culture testing has high requirements for operation and is time-consuming.
- New technologies for urine culture outcome still have a long way to go before being able to be applied in clinical practice.
- Our study aim to determine the predictive value of machine learning algorithms using a urine culture predictive model based on patients with urinary stones.

Materials and Methods

- A total of 2,054 urinary stone patients’ data from four clinical centers were analyzed.
- Predictive models of urine culture outcomes were constructed and validated by logistic regression, random forest, adaboost, and gradient boosting decision tree (GBDT) models.
- ROC with AUC was used to evaluate the performance of each prediction model. Additive NRI and absolute NRI were used to assess the predictive capabilities of the models.

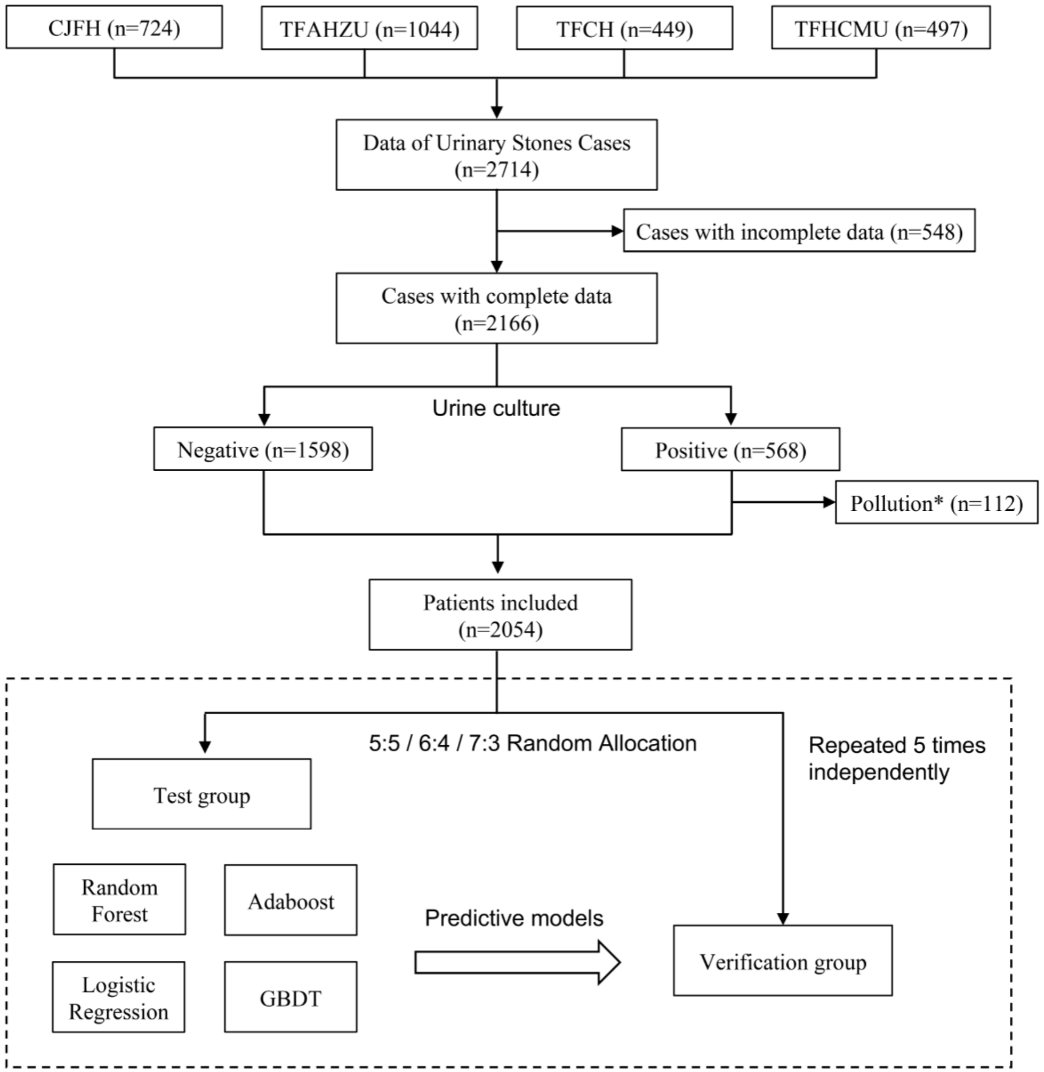

Figure 1 Flowchart of the study. CJFH, China-Japan Friendship Hospital. TFAHZU, The First Affiliated Hospital of Zhengzhou University. TFHCMU, The First Affiliated Hospital of China Medical University. TFCH, Tianjin First Central Hospital. GBDT, gradient boosting decision tree. *, a urine culture with three or more bacteria is considered contaminated.

Results

Patients' Characteristics

- A total of 2,054 cases from four clinical centers were ultimately included in the study.
- Patients had a mean age of 51.3±13.8 and a positive urine culture rate of 22.3%.
- Females, age ⩾60, with hypertension, diabetes, coronary heart disease and smoking habits had a significant association to positive outcomes of urine culture.

Table 1 Correlation between urine culture results and main clinical characteristics in 2,054 urinary calculi patients from four clinical centers

Main Clinical Characteristic	Number of patients	Urine culture: Positive (%)	Urine culture: Negative (%)	p
Total	2054	456 (77.8%)	1598 (22.2%)	
General Information				
Sex				
Male *vs* Female	1355 *vs* 699	194 (14.3%) *vs* 437 (62.5%)	1161 (85.7%) *vs* 262 (37.5%)	<0.001*
Age				
<60 *vs* ≥60	1454 *vs* 600	275 (18.9%) *vs* 81(30.1%)	1179(81.087%) *vs* 419 (69.8%)	<0.001*
BMI#				
<23 *vs* ≥23	454 *vs* 1599	87(19.2%) *vs* 369 (23.1%)	367(80.8%) *vs* 1230(76.9%)	0.077
Past history				
Hypertension				
Yes *vs* No	599 *vs* 1455	159 (26.5%) *vs* 297 (20.4%)	440 (73.46%) *vs* 1158 (79.6%)	0.002*
Diabetes				
Yes *vs* No	300 *vs* 1754	81 (27.0%) *vs* 375 (21.4%)	219 (73.00%) *vs* 1379 (78.6%)	0.030*
Coronary heart disease				
Yes *vs* No	98 *vs* 1956	34 (34.7%) *vs* 422 (21.6%)	64 (65.31%) *vs* 1534 (78.4%)	0.002*
History of abdominal/pelvic surgery				
Yes *vs* No	260 *vs* 1794	69 (26.5%) *vs* 387 (21.6%)	191 (73.46%) *vs* 1407 (78.4%)	0.072
History of cerebrovascular disease				
Yes *vs* No	43 *vs* 2011	9 (21.0%) *vs* 447 (22.2%)	34 (79.07%) *vs* 1564 (77.8%)	0.839
Malformation of urinary system				
Yes *vs* No	120 *vs* 1934	30 (25.0%) *vs* 426 (22.0%)	90 (75.00%) *vs* 1508 (78.0%)	0.447
Personal history				
Smoking				
Yes *vs* No	552 *vs* 1502	146 (26.5%) *vs* 310 (20.6%)	406 (73.6%) *vs* 1192 (79.4%)	0.005*
Drinking				
Yes *vs* No	130 *vs* 1924	27 (20.8%) *vs* 429 (22.3%)	103 (79.2%) *vs*1495 (77.7%)	0.685

BMI, body mass index; *vs*, versus; #, BMI was divided according to national surveys to fit Chinese actual situation.[20] Chi-square tests. *, $p<0.05$

20 Tian Y, Jiang C, Wang M, Cai R, Zhang Y, He Z, Wang H, *Wu D*, Wang F, Liu X, He Z (2016) BMI, leisure-time physical activity, and physical fitness in adults in China: results from a series of national surveys, 2000-14. Lancet Diabetes Endocrinol 4:487-497.

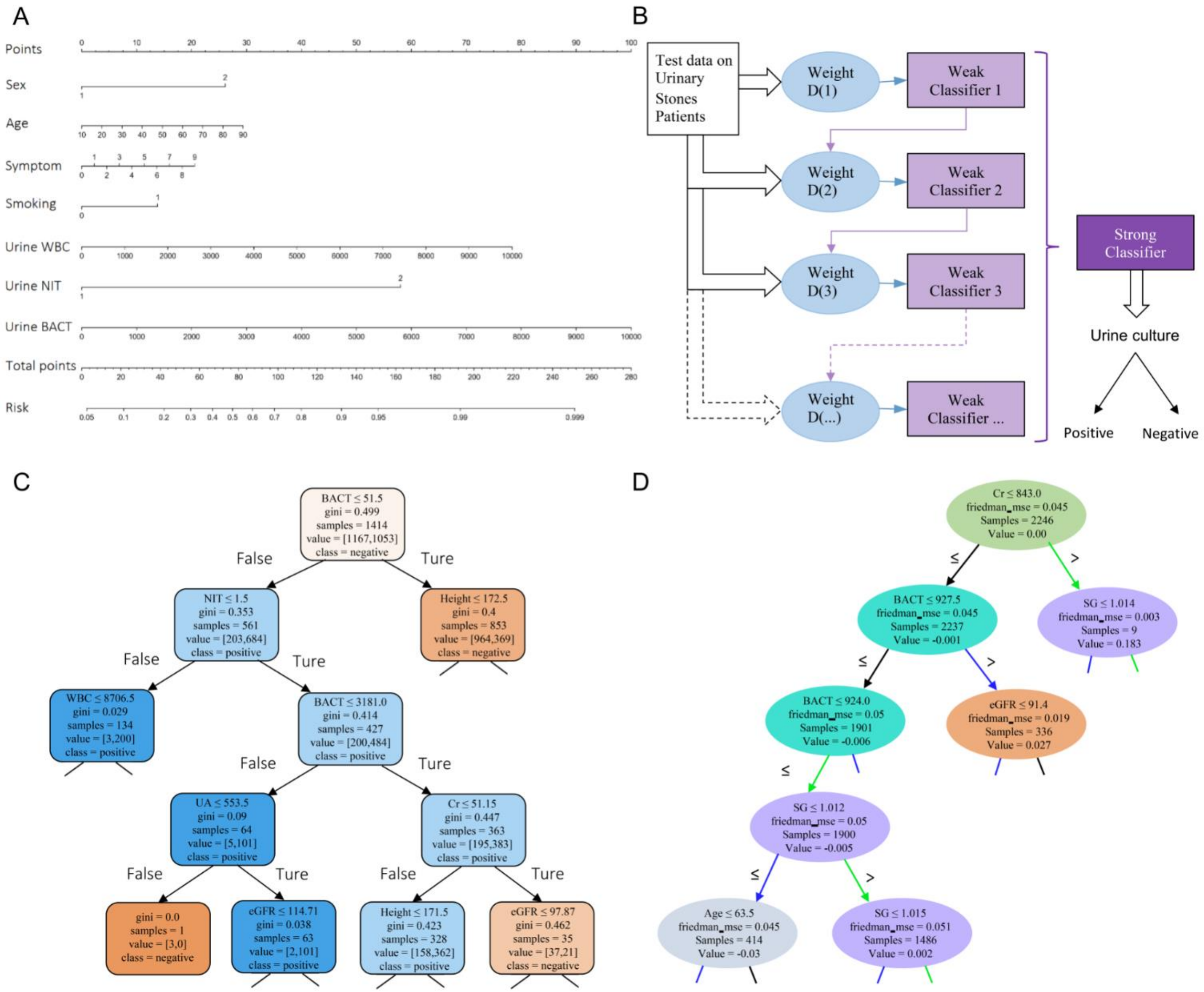

Figure 2 Visualization of models included in this study. (A) Nomogram model to predict the risk of positive urine culture based on a logistic regression algorithm. (B) Visible principle of the algorithm based on adaboost. (C) and (D) are visualizations in parts of leaves based on the random forest and GBDT models. It should be noted that the visual model only shows a part of the leaves or principles of the decision tree and does not represent the entire model.

Supplement Table 1 Effect-size estimation of predictors in association with the risk of positive urine culture

Variables	Univariate analysis			Multivariate analysis		
	OR	95% CI	p	OR	95% CI	p
Sex	3.458	2.621-4.563	<0.001	2.835	2.017-3.983	<0.001
Age	1.026	1.015-1.037	<0.001	1.013	1.0001-1.0251#	0.047
Symptom	1.069	1.001-1.140	0.045	1.094	1.014-1.178	0.020
Diabetes	1.458	1.027-2.071	0.035			
Smoking	1.404	1.047-1.883	0.023	1.766	1.225-2.546	0.002
Urine WBC	1.001	1.0006-1.0011#	<0.001	1.001	1.0001-1.0005#	0.002
NIT	22.452	11.589-43.496	<0.001	10.106	4.878-20.941	<0.001
PH	1.416	1.177-1.703	<0.001			
Urine BACT	1.001	1.0004-1.0008#	<0.001	1.001	1.0002-1.0005#	<0.001

OR>1 suggests risk factors to a positive urine culture. OR, odds ratio; 95% CI, 95% confidence interval; WBC, white blood cell; NIT, nitrite; PH, potential of hydrogen potential of hydrogen, BACT, bacteriuria. #, reserved to four digits after decimal point. *, p<0.05.

Supplement Table 2 Features selection to construct models

Features
Sex, Age, Height, Weight, Symptom, Duration of symptom, Hypertension, Years of hypertension, DM, Years of DM, CHD, Years of CHD, Pelvic surgery, Years of pelvic surgery, Cerebral infarction, Years of cerebral infarction, Urinary system anatomy, Gallstone, Smoking, Number of cigarettes/day, Years of smoking, Drinking, Vol. of drinking/30mL/day, Allergy, Blood type, UA, Cr, Glu, Ca, P, eGFR, SG, PH, NIT, Urine WBC, Urine RBC, Urine BACT, Numbers of stones, (Single)Stone location, (Single)length of stone, (Single)Hight of stone, CT value of stone(s)

DM, diabetes mellitus; CHD, coronary heart disease; UA, uric acid; Cr, creatinine; Glu, glucose; Ca, calcium; P, phosphorus; eGFR, estimated glomerular filtration rate; SG, specific gravity; PH, potential of hydrogen potential of hydrogen; NIT, nitrite; WBC, white blood cell; RBC, red blood cell; BACT, bacteria; CT, computed tomography.

Supplement Table 3 Hyperparameter values of the final models

Classifier	Hyperparameter	Value
GBDT	the maximum depth of variable interactions	3
	number of trees	100
	shrinkage	none
	number of minobsinnode	none
Random forest	number of trees	11
Adaboost	number of estimators	100
	learning rate	1.0

GBDT, gradient boosting decision tree.

Results

Model construction

- A nomogram model, was constructed based on risk factors at a significance level of 5%. (Logistic regression model)
- 42 features were selected to construct the machine learning models including sex, age, smoking habits, etc.
- Urine BACT was the most important feature among the three machine learning models.
- Test No. 3, with a ratio of 6:4, is a typical example of models' performance. (Fig.3 B)

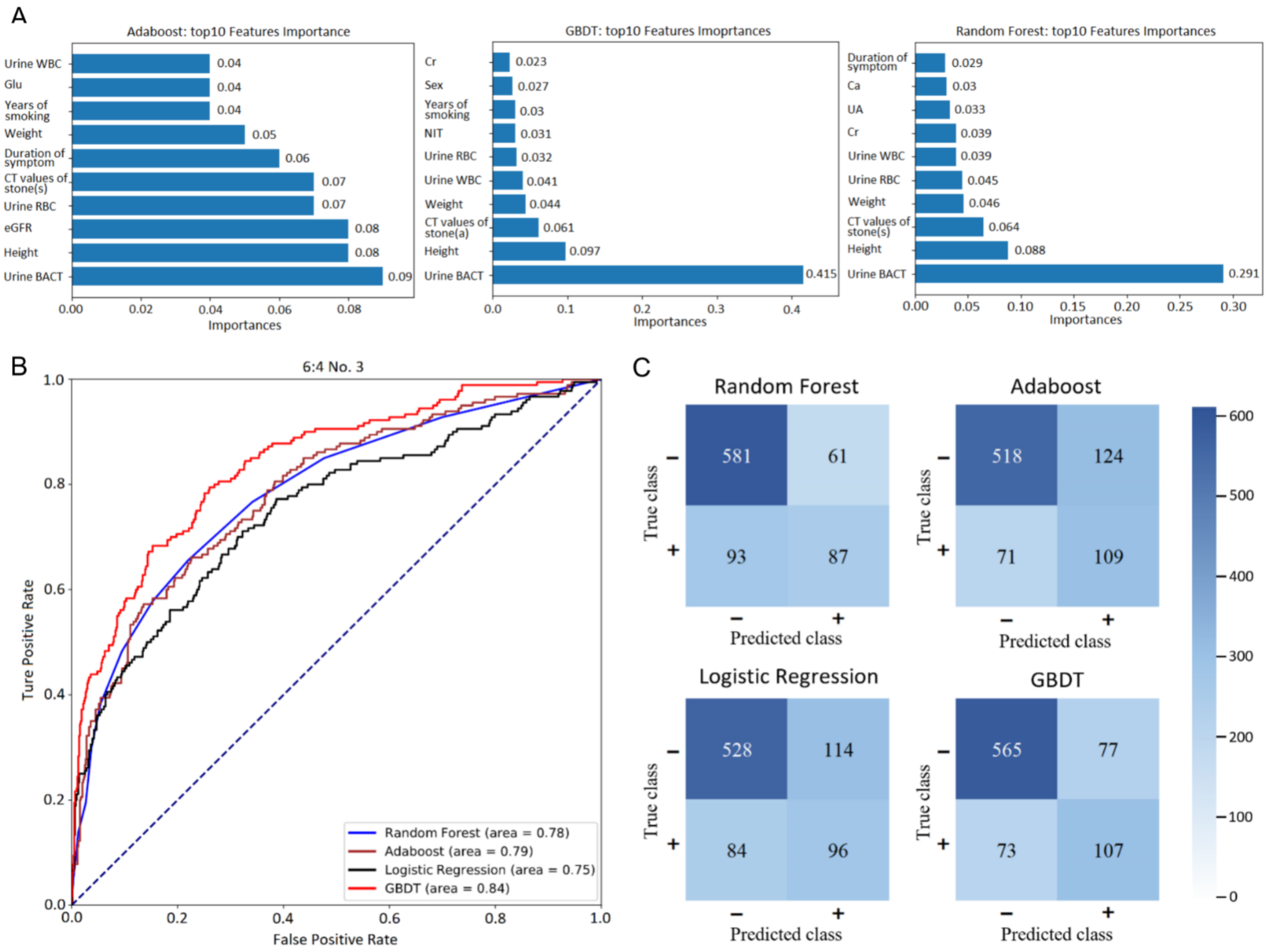

Figure 3 Important features in machine learning models and typical examples of performance predictions. (A) Top 10 important features of each machine learning model. (B) Typical receiver operating characteristic curve in four models in test No. 3 with a ratio of 6:4. (C) Confusion matrix of four models in test No. 3 with a ratio of 6:4. No., number.

Results

Model performance and comparison

- The GBDT model had the highest average AUC among all models, 0.07 higher than the logistic regression model.
- The additive NRI and absolute NRI of the GBDT and logistic regression models were 0.124 (95% CI: 0.106-0.142) and 0.065 (95% CI: 0.060-0.069), respectively.

Table 2 Performance of each model on validation data

Models	AUC		Sensitivity (%)		Specificity (%)		Additive NRI*		Absolute NRI*	
	Average	95% CI	Average	95% CI	Average	95% CI	Average	95% CI	Average	95% CI
Logistic Regression	0.761	0.753-0.770	59.2	0.575-0.766	77.7	0.766-0.788	-	-	-	-
Random Forest	0.790	0.782-0.798	50.4	0.489-0.519	88.6	0.879-0.894	0.020	0.004-0.035	0.065	0.057-0.065
Adaboost	0.779	0.766-0.791	62.3	0.606-0.640	79.8	0.791-0.805	0.051	0.036-0.159	0.023	0.016-0.030
GBDT	0.831	0.823-0.840	64.0	0.619-0.662	84.6	0.836-0.855	0.124	0.106-0.142	0.065	0.060-0.069

AUC, area under curve; 95% CI, 95% confidence interval; NRI, Net reclassification index; GBDT, gradient boosting decision tree; *, Additive NRI and absolute NRI were calculated by comparing learning machine models to the logistic regression model

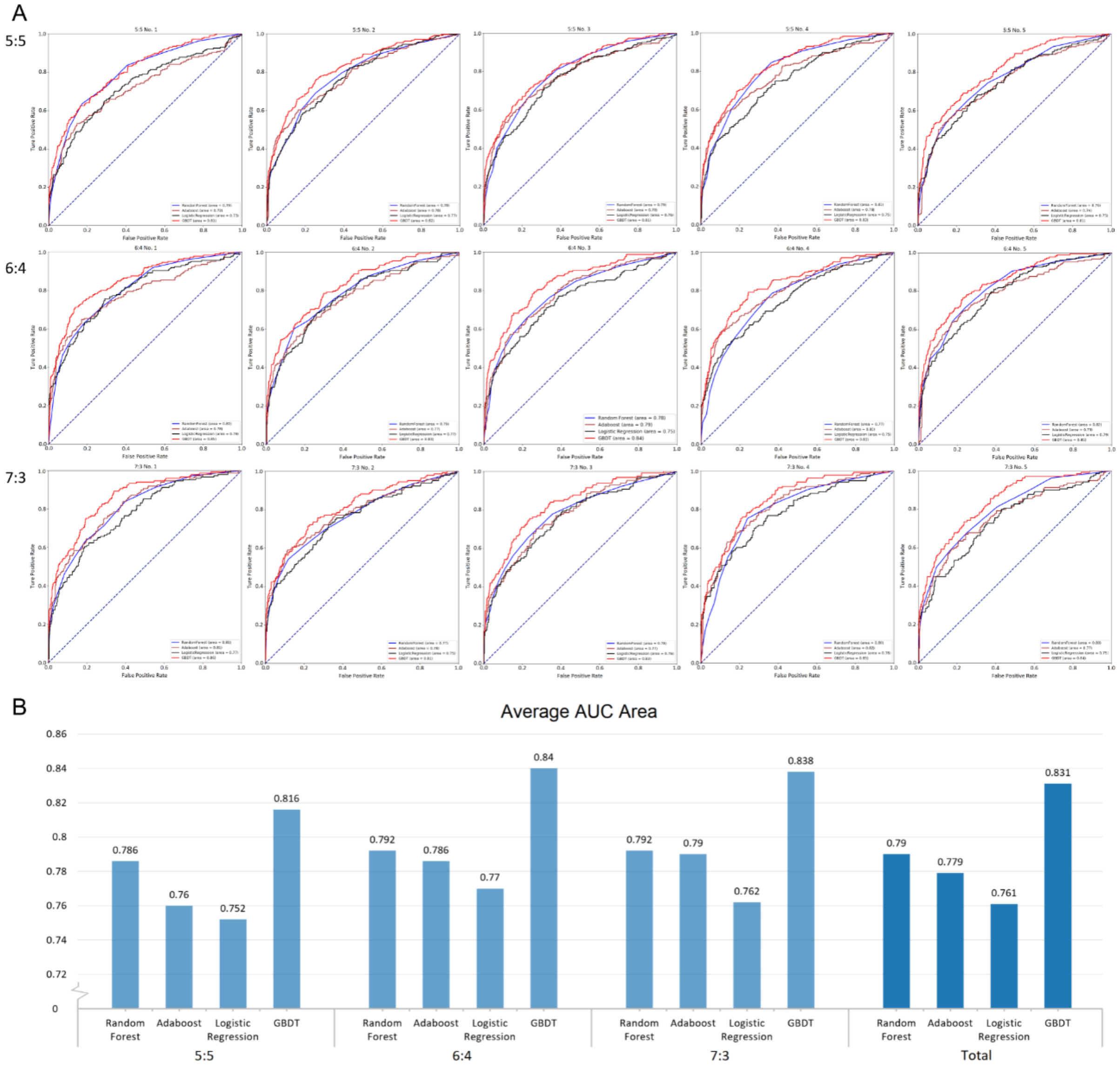

Figure 4 General performance of each model. (A) ROC of different models for cross-validation. (B) Statistical histogram of AUC. ROC, receiver operating characteristic. AUC, area under curve.

Discussion

- Current reports concerning urine culture predictive models are based on logistic regression algorithms, and their results are insufficiently accurate as a consequence of the algorithm's inherent limitations.
- The predictive accuracy of the GBDT model was, on average, 7% higher than that of the logistic regression model.
- The improvement in predictive accuracy is of great significance. For example, compared with laparoscopic, the accuracy of the Da Vinci robot surgery is higher, so the benefit to patients is huge.
- It is believed that in the future, models of deep machine learning will cover all areas of medicine - not just urology.

Strengths and limitations of this study

- Urine culture prediction model can help urologists acquire results in a short time.
- The use of machine learning algorithms improved the accuracy of the urine culture predictive model compare to traditional algorithms.
- In the future, machine learning models may be widely used in other areas of medicine.
- The samples for model training in this study are insufficient compared with those for machine learning models in other fields.
- This is a retrospective study, and further verification with a large sample of prospective studies are required.

敬请各位老师批评指正