Journal of Infection and Public Health 17 (2024) 10–17

Contents lists available at ScienceDirect

Journal of Infection and Public Health

journal homepage: www.elsevier.com/locate/jiph

Original article

Development and validation of artificial intelligence models to predict urinary tract infections and secondary bloodstream infections in adult patients

Min Hyuk Choi [a,b], Dokyun Kim [a,b,1,*], Yongjung Park [a,1,**], Seok Hoon Jeong [a,b]

[a] Department of Laboratory Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine, 211 Eonju-ro, Gangnam-gu, Seoul 06273, South Korea
[b] Research Institute of Bacterial Resistance, Yonsei University College of Medicine, Seoul, South Korea

ARTICLE INFO

Article history:
Received 17 March 2023
Received in revised form 28 September 2023
Accepted 22 October 2023

Keywords:
Urinary tract infection
Secondary bloodstream infection
Machine learning
Artificial intelligence
Urinalysis

ABSTRACT

Background: Traditional culture methods are time-consuming, making it difficult to utilize the results in the early stage of urinary tract infection (UTI) management, and automated urinalyses alone show insufficient performance for diagnosing UTIs. Several models have been proposed to predict urine culture positivity based on urinalysis. However, most of them have not been externally validated or consisted solely of urinalysis data obtained using one specific commercial analyzer.

Methods: A total of 259,187 patients were enrolled to develop artificial intelligence (AI) models. AI models were developed and validated for the diagnosis of UTI and urinary tract related-bloodstream infection (UT-BSI). The predictive performance of conventional urinalysis and AI algorithms were assessed by the areas under the receiver operating characteristic curve (AUROC). We also visualized feature importance rankings as Shapley additive explanation bar plots.

Results: In the two cohorts, the positive rates of urine culture tests were 25.2% and 30.4%, and the proportions of cases classified as UT-BSI were 1.8% and 1.6%. As a result of predicting UTI from the automated urinalysis, the AUROC were 0.745 (0.743–0.746) and 0.740 (0.737–0.743), and most AI algorithms presented excellent discriminant performance (AUROC > 0.9). In the external validation dataset, the XGBoost model achieved the best values in predicting both UTI (AUROC 0.967 [0.966–0.968]) and UT-BSI (AUROC 0.955 [0.951–0.959]). A reduced model using ten parameters was also derived.

Conclusions: We found that AI models can improve the early prediction of urine culture positivity and UT-BSI by combining automated urinalysis with other clinical information. Clinical utilization of the model can reduce the risk of delayed antimicrobial therapy in patients with nonspecific symptoms of UTI and classify patients with UT-BSI who require further treatment and close monitoring.

Introduction

Urinary tract infections (UTIs) are one of the most common infections encountered in communities and in healthcare settings [1–3]. Although UTIs are mostly asymptomatic and self-limiting, they can cause serious complications such as secondary bloodstream infection (BSI), urosepsis, and death, which require early clinical decision [4–7].

Urine culture is the standard test for the definitive diagnosis of UTI [8,9], and blood culture could provide additional information in selected UTI patients who have been treated with antimicrobial agents prior to urine sample collection or who are at high risk of developing secondary BSI [10]. However, traditional culture methods are time-consuming, making it difficult to utilize the results in the early stages of UTI management. Therefore, a presumptive diagnosis of UTI through urinalysis is recommended to determine the initiation of empirical therapy [11]. Automated urinalysis, including test strip analysis and urine sediment analysis, is available in a timely manner and can reduce medical costs and laboratory workload, but

* Corresponding author at: Department of Laboratory Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine, 211 Eonju-ro, Gangnam-gu, Seoul 06273, South Korea.
** Corresponding author.
E-mail addresses: kyunsky@yuhs.ac (D. Kim), YPARK119@yuhs.ac (Y. Park).
[1] These authors contributed equally as co-corresponding authors to this work.

https://doi.org/10.1016/j.jiph.2023.10.021

these tests alone show insufficient performance for UTI diagnostics [12].

Several models have been proposed to predict urine culture positivity based on urinalysis. However, these models predominantly relied on urinalysis data obtained from specific analyzers and lacked standardized and externally validated predictive models [13–18]. This is a critical issue given that commercial urine analyzers operate on different principles and data from one center cannot be directly generalized to other settings without external validation. Therefore, new prediction system that is available early and generally applicable to UTI diagnosis are needed, and more comprehensive models with individualized information can improve on the limitations of existing systems. Artificial intelligence (AI) technology could provide new insights into various fields of clinical medicine by detecting complex and non-linear relationships between variables that cannot be easily captured by conventional statistical methods [19].

The aim of this study was to develop a generalizable AI prediction model for the diagnosis of UTI using large cohort data from two university hospitals that differ in patient characteristics and automated urinalysis systems used. To simultaneously provide information on the indications for blood culture tests in patients with UTIs, additional AI models to predict the progression to urinary tract-related BSI (UT-BSI) were suggested. Furthermore, the UTI and UT-BSI prediction models constructed in this study were externally validated with the independent dataset and identified the variables having the largest impact on the models.

Patients and methods

Patients

We retrospectively collected all cases with urine culture results from two university hospitals (hospital 'S′ and hospital 'G′, tertiary hospitals with more than 2000 and 800 beds, respectively) in South Korea from 2011 to 2021. During the study period, a total of 584,055 patients were subjected to urine cultures for suspected UTIs. Patients who did not have a urinalysis result within six hours before or after urine culture were excluded (n = 77,153) [15]. Other exclusion criteria for cases included patients under 19 years of age, absence of demographic data, or more than 20% missing values [20]. Finally, 259,187 patients were enrolled for the analysis, and cases consisted of a development dataset (196,932 patients from hospital 'S′) and an external validation dataset (62,255 patients from hospital 'G′). According to Centers for Disease Control/National Healthcare Safety Network definitions [21], the following urine culture results were classified as positive: no more than two species of microorganisms growth of ≥ 10^5 colony-forming unit (CFU) per mL; single pathogen growth of ≥ 10^4 CFU/mL; or pathogen growth of ≥ 10^3 CFU/mL in urine samples collected via a straight catheter. Positive findings in urinalysis were defined as dipstick positive for leukocyte esterase (LE) and/or nitrite or when ≥ 10 white blood cells (WBCs)/mL or ≥ 3 WBC/high-power field were observed in a urine sample. In addition, a case in which the same microorganism was isolated from urine and blood culture within three days from the same patient was defined as a UT-BSI.

Patient-level data were collected, including demographics, underlying comorbidities with age-adjusted Charlson comorbidity index, date of urine sample collection, and commercial analyzers on which urinalysis was performed. To extract the worst values within 24 h of urine culture sampling, both maximum and minimum values of laboratory tests and vital signs were obtained. In addition, the use of vasopressors, antimicrobial agents, and mechanical ventilation was investigated, and the worst Glasgow Coma Scale and a Sequential Organ Failure Assessment (SOFA) score were also calculated [22].

Automated urinalysis

We retrieved the results of test strip analysis and urine sediment analysis using the electronic medical record collection programs of the at each institution. Automated urinalysis was conducted in hospital 'S′ using the URiSCAN Pro/Super automated urine chemistry analyzer (YD Diagnostics, Yongin-si, Republic of Korea) with the UF-1000i automated urine particle analyzer (Sysmex Co., Kobe, Japan) from 2011 to 2013, the CLINITEK Advantus Urine Chemistry Analyzer and CLINITEK Novus automated urine chemistry analyzer (SIEMENS Healthineers AG, Erlangen, Germany) with the UF-1000i/UF-5000 automated urine particle analyzer (Sysmex) from 2014 to 2017, and the Atellica 1500 Automated Urinalysis System (CLINITEK Novus & Atellica UAS 800) (SIEMENS) from 2018 to 2021. In hospital 'G′, the same tests were performed with the URiSCAN Pro automated urine chemistry analyzer (YD Diagnostics) and the UF-1000i automated urine particle analyzer (Sysmex) from 2011 to 2017 and the iRICELL system (iChem VELOCITY urine chemistry analyzer & iQ200 SPRINT urine microscopy analyzer) (Beckman Coulter Inc., Brea, CA) from 2018 to 2021.

Model development

We conducted this study in accordance with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guidelines for prediction model development and validation [23]. Before the modeling, all continuous variables were standardized, and missing values were replaced with median values [20]. The development dataset from hospital 'S′ was randomly split, with 80% of the cases serving as the training set and the remaining 20% as the test set for internal evaluation of the models. In addition, the performance of models selected through internal testing with the development dataset was verified with the external validation dataset from the independent hospital 'G′.

Models were developed to predict both urine culture positivity and the occurrence of UT-BSI. Candidate models were trained using the Attentive Interpretable Tabular Learning neural network (TabNet), K-nearest neighbor (KNN), extreme gradient boosting (XGBoost), and light gradient boosting (LightGBM) algorithms. KNN was selected as a representative of non-parametric algorithms known for their adaptability to simple data sets, while XGBoost and LightGBM were chosen as machine learning classifiers to handle complex data sets through ensemble learning techniques [24]. Additionally, TabNet, a deep learning classifier designed for tubular data, was included to evaluate its performance against machine learning classifiers [25]. Hyperparameter tuning was performed via a grid search and five-fold cross-validation for each AI model (Supplementary Table 1). We first developed AI models using all parameters (81 variables, Supplementary Table 2) and then selected the best-performing algorithm as the final model. In addition, we further developed a reduced model using the top 10 predictors to enable faster simulations, providing real-time predictive results [26–28]. The number of parameters constituting the reduced model was chosen as the usable upper bound for the web application implementation of the AI model.

To compare the predictive performances of the models, we generated the highest area under the receiver operator characteristic curve (AUROC) with 95% confidence intervals (CIs) using bootstrapping. The F1 score (the harmonic mean of precision and recall), specificity, the area under the precision-recall curve (AUPRC), and accuracy were used as additional performance metrics for model comparison. Feature importance rankings were visualized as importance plots and Shapley additive explanation (SHAP) summary bar plots to interpret the models. AI analysis was conducted using

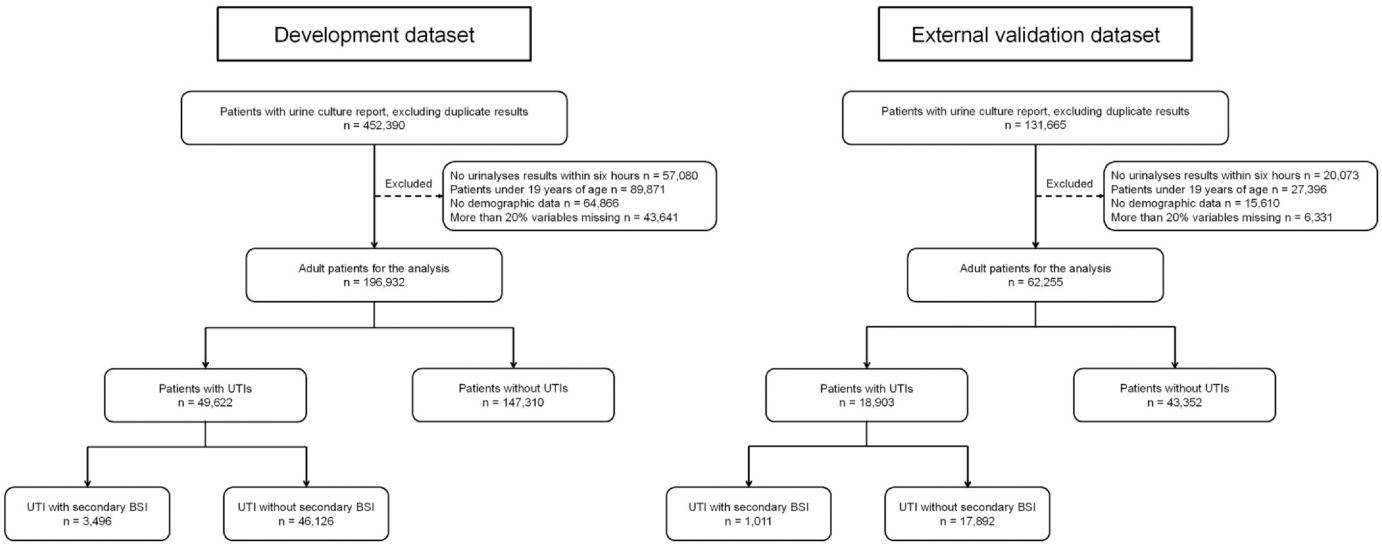

Fig. 1. Flowchart depicting steps in obtaining the dataset. Abbreviations: UTI, urinary tract infection; BSI, bloodstream infection.

Python programming software version 3.7.12 (Python software foundation, Wilmington, DE).

Statistical analysis

Baseline characteristics are presented either as means and standard deviations (SDs) for continuous variables or as numbers and percentages for categorical variables. The statistical significance between groups was tested with either the chi-square test for qualitative data or Student's t-test for quantitative data. All reported p values were two-sided, and statistical significance was assumed if $p < 0.05$. All statistical analyses were performed using R statistical software version 4.1 (R Studio, Inc., Boston, MA).

Results

Baseline characteristics of the study populations

Adult patients for whom urine culture and urinalysis were performed at intervals of less than 6 h were investigated. A total of 259,187 patients including 196,932 patients in the development cohorts and 62,255 patients in the external validation cohorts were enrolled (Fig. 1). The baseline characteristics of the enrolled patients were presented in Table 1. The mean age of patients in total dataset was 59.8 years (SD 17.0), and the most common underlying comorbidity was solid organ cancer (27.1%, n = 70,144), followed by diabetes mellitus (9.3%, n = 24,177), and congestive heart diseases (5.9%, n = 15,337). The positive rate of urine culture was 26.4% (n = 68,525), with progression to secondary bloodstream infection in 1.7% (n = 4507) of patients.

The results of automated urinalysis showed significant differences between patients with and without UTI and patients with UT-BSI (Supplementary Table 3). Urinary WBC counts were highest in the patients with UT-BSI group and were statistically higher in patients with UTI than those without UTI. Similarly, the positive (1 + to 4 +) rates of LE, urinary bacterial counts, and urinary bilirubin were highest in the group diagnosed with UT-BSI and lowest in the non-UTI group ($p < 0.001$). Compared to patients without UTI, patients with UTI and UT-BSI had a significantly higher proportion of a prior history of urologic device use such as an indwelling catheter (26.4%, 38.4%, and 57.5% in the group without UTI, the group with UTI, and group with UT-BSI, respectively).

The baseline characteristics of the patients classified according to the dataset are presented in Supplementary Table 4. There were statistically significant differences in most independent variables between the patient groups in the two cohorts. Patients in the development dataset were older, had a lower proportion of females, and had more underlying comorbidities ($p < 0.001$). The positive rates of urine culture tests were 25.2% and 30.4%, and the cases classified as UT-BSI were 1.8% and 1.6% in the development and external validation datasets, respectively, showing statistically significant differences. There were also significant differences in the composition of pathogenic microorganisms between the two cohorts (Table 2). *Escherichia coli* was the most common causative pathogen of UTI but was more prominent in the external validation cohort and in the secondary BSI group. In contrast, most Gram-positive bacteria were isolated more frequently in the development cohort and had a lower rate of progression to secondary BSI.

Development of AI models

The performance metrics for UTI prediction using automated urinalysis and AI algorithms are presented in Table 3. As a result of predicting UTI from the automated urinalysis, the AUROC was 0.745 (95% confidence interval [CI], 0.743–0.746) and 0.740 (95% CI, 0.737–0.743) in the two cohorts, respectively, showing lower predictive performance compared to the AI models. All AI algorithms presented excellent discriminant performance (AUROC > 0.9) [29] with the exception of the KNN classifier. For the internal test set, the XGBoost classifier showed the highest value in AUROC (0.988; 95% CI, 0.987–0.988), accuracy (0.955), and the F1 score (0.908). Although there was a deterioration in performance in the external validation dataset, the XGBoost model still achieved the best values in AUROC (0.967; 95% CI, 0.966–0.968), accuracy (0.909), and the F1 score (0.851). Thus, the XGBoost algorithm was selected as a classifier for the development of the final model.

Development of the model to predict UT-BSI was conducted using a similar process that is summarized in Table 4. The XGBoost algorithm performed best with AUROC values of 0.968 (95% CI, 0.964–0.971) and 0.955 (95% CI, 0.951–0.959) in the internal test and external validation sets, respectively. In both datasets, the algorithm's F1 scores were also the highest at 0.330 and 0.233, respectively.

Fig. 2 describes the AUROC and AUPRC of the external validation dataset for urine culture positivity (Figs. 2A and 2B) and UT-BSI

Table 1
Baseline characteristics of patients with and without UTI and patients with UT-BSI.

Variables	Total (n = 259187)	No UTI (n = 190662)	UTI (n = 68525)	p	UTI without BSI (n = 64018)	UT-BSI (n = 4507)	p
Age	59.8 ± 17.0	58.3 ± 16.8	63.9 ± 16.8	< 0.001	63.7 ± 17.0	67.2 ± 14.5	< 0.001
Female	123661 (47.7%)	76929 (40.3%)	46732 (68.2%)	< 0.001	43774 (68.4%)	2958 (65.6%)	< 0.001
Charlson comorbidity index	3.7 ± 2.4	3.5 ± 2.3	4.2 ± 2.4	< 0.001	4.2 ± 2.4	4.6 ± 2.2	< 0.001
Underlying comorbidities							
Solid organ cancer	70144 (27.1%)	50636 (26.6%)	19508 (28.5%)	< 0.001	18064 (28.2%)	1444 (32.0%)	< 0.001
Diabetes mellitus	24177 (9.3%)	15736 (8.3%)	8441 (12.3%)	< 0.001	7780 (12.2%)	661 (14.7%)	< 0.001
Kidney diseases	7543 (2.9%)	5838 (3.1%)	1705 (2.5%)	< 0.001	1490 (2.3%)	215 (4.8%)	< 0.001
Congestive heart diseases	15337 (5.9%)	10192 (5.3%)	5145 (7.5%)	< 0.001	4826 (7.5%)	319 (7.1%)	< 0.001
Cerebrovascular diseases	12209 (4.7%)	7207 (3.8%)	5002 (7.3%)	< 0.001	4742 (7.4%)	260 (5.8%)	< 0.001
Indwelling catheter	76727 (29.6%)	50426 (26.4%)	26301 (38.4%)	< 0.001	23711 (37.0%)	2590 (57.5%)	< 0.001
Urinalysis system							
CLINITEK Advantus urine chemistry analyzer	1892 (0.7%)	966 (0.5%)	926 (1.4%)	< 0.001	837 (1.3%)	89 (2.0%)	< 0.001
Atellica 1500 Automated Urinalysis System	46037 (17.8%)	33069 (17.3%)	12968 (18.9%)		12025 (18.8%)	943 (20.9%)	
iRICELL system	38475 (14.8%)	26234 (13.8%)	12241 (17.9%)		11589 (18.1%)	652 (14.5%)	
CLINITEK Novus automated urine chemistry analyzer	94800 (36.58%)	70294 (36.87%)	24506 (35.76%)		22787 (35.59%)	1719 (38.14%)	
URiSCAN Pro/Super automated urine chemistry analyzer with UF-1000i automated urine particle analyzer	54203 (20.9%)	42981 (22.5%)	11222 (16.4%)		10477 (16.4%)	745 (16.5%)	
URiSCAN Pro automated urine chemistry analyzer with UF-1000i automated urine particle analyzer	23780 (9.2%)	17118 (9.0%)	6662 (9.7%)		6303 (9.8%)	359 (8.0%)	
Urine culture positivity	68525 (26.4%)	-	68525 (100.0%)		64018 (100.0%)	4507 (100.0%)	
Urinary tract related bloodstream infection	4507 (1.7%)	-	4507 (6.6%)		0 (0.0%)	4507 (100.0%)	
Ventilator use	5839 (2.3%)	4571 (2.4%)	1268 (1.9%)	< 0.001	1134 (1.8%)	134 (3.0%)	< 0.001
Maximum C-reactive protein (mg/L)	57.4 [14.9–122.9]	55.1 [13.5–120.6]	62.5 [18.1–128.5]	< 0.001	55.9 [15.8–117.8]	150.8 [86.7–231.9]	< 0.001
SOFA score	0 [0–3]	0 [0–2]	1 [0–4]	< 0.001	1 [0–3]	4[2–8]	< 0.001

Abbreviations: UTI, urinary tract infection; BSI, bloodstream infection; UT-BSI, urinary tract-related BSI; SOFA, sequential organ failure assessment; WBC, white blood cell.
Data are presented as number (%), mean ± standard deviation, or median [1st-3rd quartile].

(Figs. 2C and 2D), with XGBoost models showing the best performances.

Top predictors in the final models

The predictors for UTI and UT-BSI obtained in the development dataset, along with their rank, are presented in Fig. 3 and Supplementary Figs. 2–3. Based on the SHAP analysis of the XGBoost algorithm, urinary bacterial count, monocyte count, WBC count, lymphocyte count, urinary WBC count, specific gravity (SG), diastolic blood pressure (BP), systolic BP, patient's age, and c-reactive protein (CRP) level were selected as the top 10 predictors associated with UTI. Reduced models were also developed using these top 10 variables. Meanwhile, the urinary bacterial count in automated urine sediment analysis was the most important variable in predicting UT-BSI using the XGBoost model, followed by urinary WBC count, lymphocyte count, and CRP level. The code for a reduced model derived with the top 10 parameters was disclosed (https://github.com/tcmhwd/uti).

Table 2
Number of distributions of urinary tract infection causative microorganisms.

Organisms	Patients with UTI in the development dataset (n = 49622)	Patients with UTI in the external validation dataset (n = 18903)	Total patients with UTI (n = 68525)	UTI without BSI (n = 64018)	UT-BSI (n = 4507)
Gram-negative bacteria	33844 (68.2%)	13517 (71.5%)	47361 (69.1%)	43529 (68.0%)	3832 (85.0%)
Escherichia coli	23438 (47.2%)	9781 (51.7%)	33219 (48.5%)	30294 (47.3%)	2925 (64.9%)
Klebsiella pneumoniae	4271 (8.6%)	1346 (7.1%)	5617 (8.2%)	5029 (7.9%)	588 (13.0%)
Pseudomonas aeruginosa	1455 (2.9%)	598 (3.2%)	2053 (3.0%)	1976 (3.1%)	77 (1.7%)
Proteus spp.	1089 (2.2%)	354 (1.9%)	1443 (2.1%)	1348 (2.1%)	95 (2.1%)
Citrobacter spp.	707 (1.4%)	244 (1.3%)	951 (1.4%)	930 (1.5%)	21 (0.5%)
Enterobacter spp.	386 (0.8%)	439 (2.3%)	825 (1.2%)	798 (1.2%)	27 (0.6%)
Acinetobacter baumannii	353 (0.7%)	179 (0.9%)	532 (0.8%)	509 (0.8%)	23 (0.5%)
Gram-positive bacteria	11719 (23.6%)	3867 (20.5%)	15586 (22.7%)	15066 (23.5%)	520 (11.5%)
Enterococcus faecalis	4804 (9.7%)	1642 (8.7%)	6446 (9.4%)	6325 (9.9%)	121 (2.7%)
Enterococcus faecium	2914 (5.9%)	1048 (5.5%)	3962 (5.8%)	3837 (6.0%)	125 (2.8%)
Streptococcus agalactiae	840 (1.7%)	236 (1.2%)	1076 (1.6%)	1042 (1.6%)	34 (0.8%)
Staphylococcus aureus	623 (1.3%)	236 (1.2%)	859 (1.3%)	647 (1.0%)	212 (4.7%)
Fungus	4059 (8.2%)	1519 (8%)	5578 (8.1%)	5423 (8.5%)	155 (3.4%)
Candida spp.	3806 (7.7%)	1490 (7.9%)	5296 (7.7%)	5148 (8.0%)	148 (3.3%)
Other fungus	253 (0.5%)	29 (0.2%)	282 (0.4%)	275 (0.4%)	7 (0.2%)
Other microorganisms, including cases of polymicrobial infection	4683 (9.4%)	1281 (6.8%)	5964 (8.7%)	5860 (9.2%)	104 (2.3%)

Abbreviations: UTI, urinary tract infection; BSI, bloodstream infection; UT-BSI, urinary tract-related BSI
Data are presented as number (%).

Table 3
Performance metrics for the prediction of UTI by urinalysis alone and the Artificial intelligence algorithms.

Algorithm	AUROC (95% CI)	Specificity	Sensitivity/Recall	Accuracy	PPV/Precision	F1 score
Internal test set						
Urinalysis alone	0.745 (0.743–0.746)	0.601	0.888	0.673	0.428	0.578
TabNet	0.978 (0.973–0.976)	0.885	0.934	0.897	0.729	0.819
KNN	0.883 (0.880–0.887)	0.951	0.619	0.867	0.81	0.702
LightGBM	0.978 (0.977–0.979)	0.974	0.83	0.938	0.916	0.871
XGBoost	0.988 (0.987–0.988)	0.979	0.883	0.955	0.934	0.908
XGBoost with top 10 variables*	0.963 (0.962–0.965)	0.967	0.79	0.922	0.891	0.837
External validation set						
Urinalysis alone	0.740 (0.737–0.743)	0.591	0.889	0.681	0.487	0.629
TabNet	0.956 (0.955–0.957)	0.779	0.958	0.834	0.655	0.778
KNN	0.843 (0.840–0.846)	0.839	0.691	0.794	0.651	0.671
LightGBM	0.951 (0.949–0.952)	0.934	0.801	0.894	0.841	0.821
XGBoost	0.967 (0.966–0.968)	0.933	0.855	0.909	0.847	0.851
XGBoost with top 10 variables*	0.916 (0.914–0.918)	0.939	0.67	0.857	0.827	0.740

Abbreviations: UTI, urinary tract infection; AUROC, area under the receiver operating characteristic curve; CI, confidence interval; PPV, positive predictive value; TabNet, Attentive Interpretable Tabular Learning neural network; KNN, K-nearest neighbor; LightGBM, light gradient boosting; XGBoost, eXtreme gradient boosting.
*Reduced model using only the top 10 predictors (urinary bacteria, monocyte count, WBC count, lymphocyte count, urinary WBC count, specific gravity, diastolic blood pressure, systolic blood pressure, patient's age, and serum c-reactive protein level) selected by Shapley additive explanation analysis

Discussion

We found that applying AI technology can improve the early prediction of urine culture positivity and UT-BSI by combining automated urinalysis with other clinical information.

Some previous studies have proposed urine culture prediction models through traditional statistical methods [13–15]. These models tried to predict UTIs by suggesting multiple cutoffs for automated urinalysis results but showed relatively low F1 scores, which is an informative value that evaluates classifiers on imbalanced data [30]. Urine cultures obtained in real clinical settings have much higher rates of negative findings, and perhaps conventional logistic regression models could not overcome this fundamental data imbalance. Furthermore, the evaluation of most previous models only included results from one specific automated urinalysis system, and there were no standardized, externally validated predictive models for identifying patients with UTIs [14,31,32].

AI-based algorithms have the inherent advantage over traditional statistical approaches of correcting multicollinearity and nonlinear relationships between independent variables [33], resulting in better predictive performance in various fields of clinical medicine for diagnosis and prognostic prediction [19]. Several prior studies have attempted to use AI techniques to differentiate patients with UTIs [16–18], but none of these results have been externally validated. Accordingly, we developed AI models for predicting positive urine culture and validated it externally in a cohort with different patient characteristics. The final model in our study had an excellent performance of ≥ 0.9 in both the F1 score and AUROC in the internal test cohort. Even in the external validation cohort that used different urinalysis analyzers, the final model maintained high predictive performance with an F1 score of 0.851 and AUROC of 0.967, showing only a slight deterioration compared to that of the internal test set.

Most cases with asymptomatic UTIs are self-limiting and do not require antimicrobial therapy. However, it is difficult to differentiate between asymptomatic and nonspecific systemic symptoms, especially in elderly, unconscious, and immunocompromised patients [34]. Studies assessed on UT-BSI are of particular importance in university hospitals with a high proportion of critically ill patients. Progression from simple UTI to secondary BSI has a high mortality rate of 20–40% and is a clinical indicator of empirical antimicrobial treatment [35]. In this regard, blood cultures could provide additional clues for treatment decisions in certain patients with UTI, but it is time-consuming and has issues with increased laboratory workload. Therefore, we developed the AI model that simultaneously predicts UT-BSI based on urinalysis results and presented the final XGBoost model with an AUROC of 0.9 or higher in both the internal test cohort and external validation cohort. Since the proportions of patients with UT-BSI in the two cohorts were 1.8% and 1.6%, which are much lower than the 25.2% and 30.4% of UTI cases, the performance (especially the sensitivity and F1 score) of the model was deteriorated compared to the UTI prediction. However, our model shows a specificity of ≥ 0.99, which may contribute to

Table 4
Performance metrics for the prediction of UT-BSI by urinalysis alone and the Artificial intelligence algorithms.

Algorithm	AUROC (95% CI)	Specificity	Sensitivity/Recall	Accuracy	PPV/Precision	F1 score
Internal test set						
Urinalysis alone	0.726 (0.723–0.729)	0.486	0.966	0.494	0.033	0.063
TabNet	0.937 (0.926–0.947)	0.907	0.841	0.906	0.138	0.237
KNN	0.706 (0.691–0.721)	0.999	0.064	0.981	0.505	0.114
LightGBM	0.968 (0.964–0.971)	0.998	0.206	0.983	0.634	0.311
XGBoost	0.968 (0.964–0.971)	0.997	0.223	0.983	0.628	0.330
XGBoost with top 10 variables*	0.928 (0.919–0.936)	0.999	0.091	0.982	0.615	0.159
External validation set						
Urinalysis alone	0.706 (0.700–0.712)	0.452	0.960	0.460	0.028	0.055
TabNet	0.883 (0.873–0.892)	0.915	0.655	0.911	0.113	0.192
KNN	0.703 (0.690–0.717)	0.997	0.076	0.982	0.298	0.121
LightGBM	0.954 (0.949–0.958)	0.999	0.132	0.984	0.596	0.216
XGBoost	0.955 (0.951–0.959)	0.998	0.146	0.984	0.569	0.233
XGBoost with top 10 variables*	0.893 (0.885–0.902)	0.999	0.058	0.984	0.504	0.105

Abbreviations: UT-BSI, urinary tract-related bloodstream infection; AUROC, area under the receiver operating characteristic curve; CI, confidence interval; PPV, positive predictive value; TabNet, Attentive Interpretable Tabular Learning neural network; KNN, K-nearest neighbor; LightGBM, light gradient boosting; XGBoost, eXtreme gradient boosting.
*Reduced model using only the top 10 predictors (urinary bacteria, monocyte count, WBC count, lymphocyte count, urinary WBC count, specific gravity, diastolic blood pressure, systolic blood pressure, patient's age, and serum c-reactive protein level) selected by Shapley additive explanation analysis

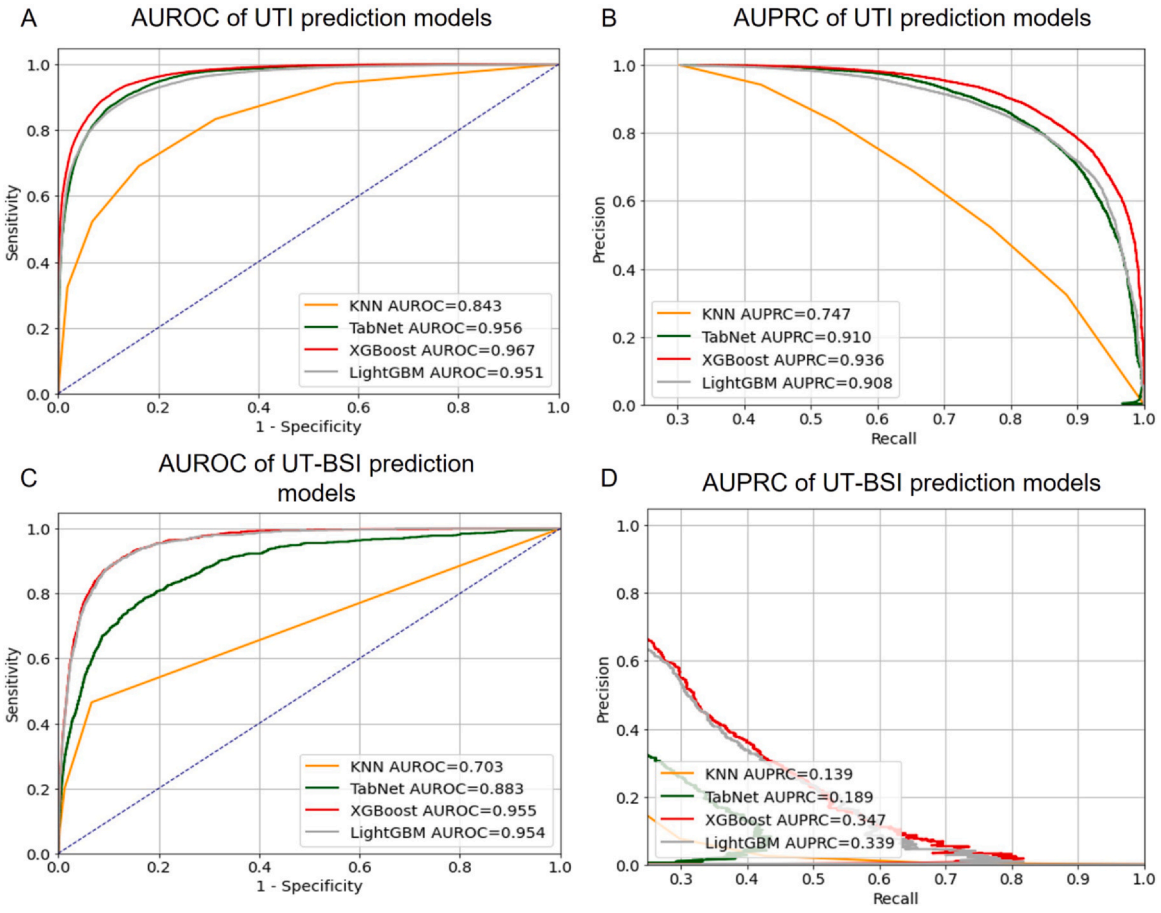

Fig. 2. Comparison of Artificial intelligence-based UTI and UT-BSI prediction models. Abbreviations: UTI, urinary tract infection; UT-BSI, urinary tract-related bloodstream infection; AUROC, area under the receiver operating characteristic curve; TabNet, Attentive Interpretable Tabular Learning neural network; KNN, K-nearest neighbor; LightGBM, light gradient boosting; XGBoost, eXtreme gradient boosting; AUPRC, area under the precision-recall curve.

reducing unnecessary blood culture tests and associated costs in patients with UTIs.

The XGBoost algorithm achieved the best discriminative power in predicting UTI and UT-BSI in this study. In general, ensemble tree-based machine learning models such as XGBoost and LightGBM are known to perform best on tubular data sets, which is consistent with our study [36]. In these final models, we also presented the top variables with the largest impact on the predictions of UTI and UT-BSI. These parameters could be grouped into three categories: 1) automated urinalysis results (urinary bacterial count, urinary WBC count, and SG), 2) other laboratory test results obtained on the same day that the urinalysis was performed (total and differential WBC

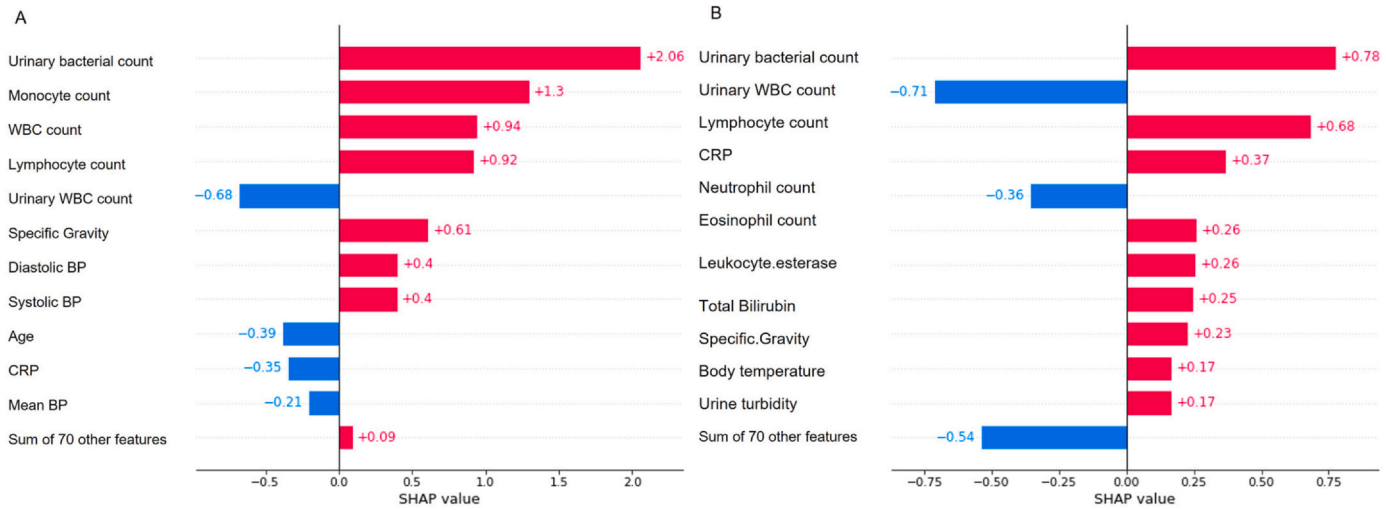

Fig. 3. SHAP value summary bar plot of the final model including critical variables for predicting UTI and UT-BSI. Red bars indicate higher values or affirmative responses for binary features, and blue bars indicate the opposite. A positive SHAP value indicates that the variables increase the likelihood of UTI (A) or UT-BSI (B). Abbreviations: SHAP, Shapley additive explanation; UTI, urinary tract infection; UT-BSI, urinary tract-related bloodstream infection; WBC, white blood cell; BP, blood pressure; CRP, c-reactive protein.

count and CRP level), and 3) patient demographics and vital signs (age and BP). Urinary bacterial count, the best predictor of UTI in this study, is a quantification of the number of bacteria per high-power field, and any amount of urinary bacteria could be a sign of a UTI [37]. Urinary WBC is a metric commonly associated with bacteriuria but has low specificity and positive predictive value to be used alone. SG is known to correlate with urine osmolality and concentration, which could affect the diagnostic performance of dipstick positivity [38]. In addition to conventional urinalysis results, blood WBC count and serum CRP level, which are used as inflammatory markers, were also important predictors for our final model. As in previous studies, we found a high risk of UTI in elderly and female patients [39]. Therefore, the use of the AI approach in conjunction with patient baseline characteristics, automated urinalysis, and routine laboratory tests can provide more useful information for predicting UTI and UT-BSI. For convenience of data input, we derived the reduced model with the top 10 parameters, and this predictive model could support early decision-making in the clinical settings.

Our study is limited by the single country and its retrospective nature. Although it is a large-scale study, residual confounders or hidden biases, such as patient race, the prevalence of different uropathogens by hospital, and commercial urinalysis analyzers not included in our data, may have influenced the generalization of the results. Different in medical practices, patient demographics, and regional prevalence of UTIs are potential biases in these models. In addition, participants with more than 20% missing data or without demographics were excluded from the analysis. The missing values may have affected the predictive performance of our models. The study ensured that data privacy and patient confidentiality were rigorously maintained, and ethical standards required transparent reporting of limitations, potential biases, and any steps taken to mitigate these issues. This ensures the ethics and credibility of the findings. Furthermore, we also tried to minimize the impact of hidden bias and missing data through validation with the external cohort that was significantly different from the internal training cohort.

In summary, we developed prediction models for whether a patient has UTI and/or UT-BSI using AI approaches and validated its performance in the independent external cohort. These early applicable predictive systems consist only of routinely available clinical information and support clinical decision-making. Clinical utilization of the model can reduce the risk of delayed antimicrobial therapy in patients with nonspecific symptoms of UTI and classify patients with UT-BSI who require further treatment and close monitoring. Further adjustment and reinforcement of our model could be achieved through additional external validation and prospective application of the predictive system.

Ethical approval

The study was approved by the Institutional Review Board (approval no.: 3–2021–0178) of Yonsei University Gangnam Severance Hospital (Seoul, Republic of Korea).

Patents

Reduced models have submitted patent applications on the technology disclosure in the publication (10–2022–0153706).

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

M.H.C., D.K., and Y.P. designed the study and wrote the manuscript. M.H.C., and Y.P. analyzed the data. Y.P., and S.H.J supervised the study and provided critical comments. All authors have reviewed the manuscript and approved its submission.

Data Availability

The datasets used in this study are not publicly available because of ethical restrictions for the current study. These are available upon reasonable request from the corresponding author.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jiph.2023.10.021.

References

[1] Flores-Mireles AL, Walker JN, Caparon M, Hultgren SJ. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nat Rev Microbiol 2015;13(5):269–84.
[2] Foxman B. The epidemiology of urinary tract infection. Nat Rev Urol 2010;7(12):653–60.
[3] Rodríguez-Baño J, López-Prieto MD, Portillo MM, Retamar P, Natera C, Nuño E, et al. Epidemiology and clinical features of community-acquired, healthcare-associated and nosocomial bloodstream infections in tertiary-care and community hospitals. Clin Microbiol Infect 2010;16(9):1408–13.
[4] Wagenlehner FM, Tandogdu Z, Johansen TEB. An update on classification and management of urosepsis. Curr Opin Urol 2017;27(2):133–7.
[5] Gleckman R, Hibert D. Afebrile bacteremia: a phenomenon in geriatric patients. Jama 1982;248(12):1478–81.
[6] Horcajada J, Shaw E, Padilla B, Pintado V, Calbo E, Benito N, et al. Healthcare-associated, community-acquired and hospital-acquired bacteraemic urinary tract infections in hospitalized patients: a prospective multicentre cohort study in the era of antimicrobial resistance. Clin Microbiol Infect 2013;19(10):962–8.
[7] Shaw E, Benito N, Rodríguez-Baño J, Padilla B, Pintado V, Calbo E, et al. Risk factors for severe sepsis in community-onset bacteraemic urinary tract infection: impact of antimicrobial resistance in a large hospitalised cohort. J Infect 2015;70(3):247–54.
[8] Nicolle LE, Bradley S, Colgan R, Rice JC, Schaeffer A, Hooton TM. Infectious Diseases Society of America guidelines for the diagnosis and treatment of asymptomatic bacteriuria in adults. Clin Infect Dis 2005:643–54.
[9] Grabe M, Bjerklund-Johansen T, Botto H, Çek M, Naber K, Tenke P, et al. Guidelines on urological infections. Eur Assoc Urol 2015;182:237–57.
[10] Karakonstantis S, Kalemaki D. Blood culture useful only in selected patients with urinary tract infections - a literature review. Infect Dis (Lond) 2018;50(8):584–92. https://doi.org/10.1080/23744235.2018.1447682
[11] Hooton TM. Uncomplicated urinary tract infection. N Engl J Med 2012;366(11):1028–37.
[12] Demilie T, Beyene G, Melaku S, Tsegaye W. Diagnostic accuracy of rapid urine dipstick test to predict urinary tract infection among pregnant women in Felege Hiwot Referral Hospital, Bahir Dar, North West Ethiopia. BMC Res Notes 2014;7(1):1–5.
[13] Shimoni Z, Glick J, Hermush V, Froom P. Sensitivity of the dipstick in detecting bacteremic urinary tract infections in elderly hospitalized patients. PLoS One 2017;12(10):e0187381. https://doi.org/10.1371/journal.pone.0187381
[14] Foudraine DE, Bauer MP, Russcher A, Kusters E, Cobbaert CM, van der Beek MT, et al. Use of automated urine microscopy analysis in clinical diagnosis of urinary tract infection: defining an optimal diagnostic score in an academic medical center population. J Clin Microbiol 2018;56(6). https://doi.org/10.1128/JCM.02030-17
[15] Kim D, Oh SC, Liu C, Kim Y, Park Y, Jeong SH. Prediction of urine culture results by automated urinalysis with digital flow morphology analysis. Sci Rep 2021;11(1):1–8.
[16] Ozkan IA, Koklu M, Sert IU. Diagnosis of urinary tract infection based on artificial intelligence methods. Comput Methods Prog Biomed 2018;166:51–9. https://doi.org/10.1016/j.cmpb.2018.10.007

[17] Taylor RA, Moore CL, Cheung KH, Brandt C. Predicting urinary tract infections in the emergency department with machine learning. PLoS One 2018;13(3):e0194085. https://doi.org/10.1371/journal.pone.0194085
[18] Burton RJ, Albur M, Eberl M, Cuff SM. Using artificial intelligence to reduce diagnostic workload without compromising detection of urinary tract infections. BMC Med Inf Decis Mak 2019;19(1):171. https://doi.org/10.1186/s12911-019-0878-9
[19] Rajkomar A, Dean J, Kohane I. Machine learning in medicine. New Engl J Med 2019;380(14):1347–58.
[20] Chiew CJ, Liu N, Wong TH, Sim YE, Abdullah HR. Utilizing machine learning methods for preoperative prediction of postsurgical mortality and intensive care unit admission. Ann Surg 2020;272(6):1133.
[21] National Healthcare Safety Network. Patient Safety Component Manual 2019. CDC; 2019.
[22] Raith EP, Udy AA, Bailey M, McGloughlin S, MacIsaac C, Bellomo R, et al. Prognostic accuracy of the SOFA score, SIRS criteria, and qSOFA score for in-hospital mortality among adults with suspected infection admitted to the intensive care unit. Jama 2017;317(3):290–300.
[23] Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162(1):W1–73.
[24] Sarker IH. Ai-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems. SN Comput Sci 2022;3(2):158.
[25] Arik SÖ, Pfister T. Tabnet: Attentive interpretable tabular learning. Proc AAAI Conf Artif Intell 2021:6679–87.
[26] Regazzoni F, Dede L, Quarteroni A. Machine learning for fast and reliable solution of time-dependent differential equations. J Comput Phys 2019;397:108852.
[27] Akulich F, Anahideh H, Sheyyab M, Ambre D. Explainable predictive modeling for limited spectral data. Chemom Intell Lab Syst 2022;225:104572.
[28] Kim HY, Lampertico P, Nam JY, Lee H-C, Kim SU, Sinn DH, et al. An artificial intelligence model to predict hepatocellular carcinoma risk in Korean and Caucasian patients with chronic hepatitis B. J Hepatol 2022;76(2):311–8.
[29] Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression. John Wiley & Sons; 2013.
[30] Chawla NV. Data mining for imbalanced datasets: an overview. Data Min Knowl Discov Handb 2009:875–86.
[31] Parta M, Hudson BY, Le TP, Ittmann M, Musher DM, Stager C. IRIS iQ200 workstation as a screen for performing urine culture. Diagn Microbiol Infect Dis 2013;75(1):5–8.
[32] Stürenburg E, Kramer J, Schön G, Cachovan G, Sobottka I. Detection of significant bacteriuria by use of the iQ200 automated urine microscope. J Clin Microbiol 2014;52(8):2855–60.
[33] Baxt WG. Complexity, chaos and human physiology: the justification for nonlinear neural computational analysis. Cancer Lett 1994;77(2–3):85–93.
[34] Mody L, Juthani-Mehta M. Urinary tract infections in older women: a clinical review. JAMA 2014;311(8):844–54.
[35] Dreger NM, Degener S, Ahmad-Nejad P, Wöbker G, Roth S. Urosepsis—etiology, diagnosis, and treatment. Dtsch Ärzteblatt Int 2015;112(49):837.
[36] Chen T., Guestrin C. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining2016, p. 785–794.
[37] Simerville JA, Maxted WC, Pahira JJ. Urinalysis: a comprehensive review. Am Fam Physician 2005;71(6):1153–62.
[38] Chaudhari PP, Monuteaux MC, Shah P, Bachur RG. The importance of urine concentration on the diagnostic performance of the urinalysis for pediatric urinary tract infection. Ann Emerg Med 2017;70(1):63–71. e8.
[39] Choi MH, Kim D, Park Y, Jeong SH. Impact of urinary tract infection-causative microorganisms on the progression to bloodstream infection: a propensity score-matched analysis. J Infect 2022;85(5):513–8. https://doi.org/10.1016/j.jinf.2022.08.039