Comparing machine learning models for rule-out prediction of sexually transmitted infections in male patients: a retrospective study using urinalysis and symptom data

Tuncer Bahçeci[1]*, Burak Güler[1], Hakan Anıl[2], Murat Telli[3]

[1]Department of Urology, Faculty of Medicine, Aydın Adnan Menderes University, Aydın, Turkey
[2]Department of Urology, Adana City Training and Research Hospital, Adana, Turkey
[3]Department of Medical Microbiology, Faculty of Medicine, Aydın Adnan Menderes University, Aydın, Turkey

***Corresponding author:**
Tuncer Bahçeci
Department of Urology
Faculty of Medicine
Aydın Adnan Menderes University
Aydın, Turkey
E-mail:
tuncerbahceci@gmail.com

Submitted: 9 November 2025; **Accepted:** 13 December 2025
Online publication: 30 December 2025

Arch Med Sci
DOI: https://doi.org/10.5114/aoms/215558

Abstract

Introduction: Sexually transmitted infections (STIs) are a major public health issue, and prompt diagnosis is crucial for controlling infections. Multiplex PCR-based nucleic acid amplification tests (NAATs) are the gold standard for diagnosis, especially for identifying pathogens involved in gonococcal and nongonococcal urethritis. However, their frequent use in low-risk or asymptomatic men increases laboratory workload and healthcare expenses.
Material and methods: This study employed machine learning algorithms to develop and validate a rule-out model designed to reduce unnecessary NAAT requests. Data from 2020 to 2024 (n = 455) were used for model training, and data from 2025 (n = 75) served as an external temporal test set. Model thresholds were optimized to ensure high sensitivity (≥ 95%) and a high negative predictive value (NPV ≥ 90%), emphasizing patient safety and reducing the risk of false negatives. The selected rule-out threshold (0.079) represented the lowest predicted-probability cut-point achieving these safety criteria and was applied to the independent test cohort. Data were collected from our institutional database. Five supervised models – XGBoost, logistic regression (LR), support vector machine (SVM), k-nearest neighbors (KNN), and random forest – were built and assessed.
Results: XGBoost achieved the best rule-out performance (AUC = 0.86; sensitivity = 0.98; NPV = 0.91) while safely deferring NAAT testing in 44.2% of patients. Urinary WBC, leukocyte esterase, and nitrite were the strongest positive predictors, whereas absence of inflammation and asymptomatic presentation favored test deferral.
Conclusions: This model can assist clinicians in identifying male patients likely to test negative on NAATs, supporting diagnostic stewardship through cost-effective, data-driven decision-making.

Key words: sexually transmitted diseases, urethritis, nucleic acid amplification, urinalysis, machine learning, diagnostic tool.

Introduction

Sexually transmitted infections (STIs) continue to be a significant global public health problem, with over one million new cases reported each

day worldwide [1]. In men, urethritis is a common symptom, especially among those visiting urology or sexual health clinics [2]. Typical signs include urethral discharge, pain during urination, and itching, although symptoms can range from none to atypical manifestations such as pelvic pain or infertility [2–4].

Urethritis is classified as gonococcal or non-gonococcal (NGU). *Neisseria gonorrhoeae* causes gonococcal urethritis, accounting for 5–20% of cases, whereas NGU is more prevalent and associated with diverse pathogens [5]. *Chlamydia trachomatis* and *Mycoplasma genitalium* are the leading causes, while *Trichomonas vaginalis* is less frequent in high-income countries [6, 7]. The pathogenic roles of *Ureaplasma* spp. and *Mycoplasma hominis* remain uncertain, as these organisms are often detected in healthy individuals [8, 9]. The above-mentioned microorganisms can occur in patients both singly and in various combinations (mixed infections) [10]. Therefore, the European STI Guidelines Editorial Board (2018) and British Association for Sexual Health and HIV (BASHH; 2021) have both discouraged routine testing for these pathogens, emphasizing the risks of unnecessary antibiotic exposure and increased healthcare costs [11, 12]. Other microorganisms, including *Gardnerella vaginalis*, *Streptococcus* spp., and HSV-2, have also been inconsistently linked to urethritis, with limited evidence supporting a causal role [6].

According to the 2024 European Association of Urology (EAU) guidelines, the diagnosis of urethritis in symptomatic men requires evidence of inflammation, defined as ≥ 5 polymorphonuclear leukocytes (PMNLs) per high-power field (HPF) in urethral smears or ≥ 10 PMNLs/HPF or a positive leukocyte esterase (LE) test in first-void urine [13]. Men meeting these criteria should undergo nucleic acid amplification testing for *C. trachomatis*, *N. gonorrhoeae*, and *M. genitalium*, as nucleic acid amplification tests (NAATs) are considered the diagnostic gold standard due to their high sensitivity and specificity [13].

Recent advances in machine learning (ML) provide new opportunities to optimize STI diagnostics. ML models have been used to predict bacteriuria in the emergency department, early hypothermia in sepsis patients, bloodstream infections, and STI risk using clinical and laboratory data, improving test prioritization and diagnostic efficiency [14–17]. However, most STI-focused ML studies aim to identify infected individuals rather than low-risk patients who could safely defer NAAT testing. Reliable rule-out tools based on routine point-of-care data could streamline diagnostics, reduce unnecessary NAAT use, and support antimicrobial stewardship.

To our knowledge, no prior study has developed an ML model specifically designed to rule out STIs in men using routine urinalysis and symptom data. Therefore, we aimed to develop and validate an interpretable ML-based rule-out model that identifies male patients with a low likelihood of NAAT-detectable infection, emphasizing high sensitivity and negative predictive value to ensure safe test deferral and promote stewardship-aligned, cost-effective STI care.

Material and methods

Study design and population

This retrospective, single-center study involved male patients who underwent multiplex PCR-based NAAT testing for suspected sexually transmitted infections – including *Ureaplasma* spp., *Mycoplasma hominis*, *Mycoplasma genitalium*, *Chlamydia trachomatis*, *Neisseria gonorrhoeae*, Group B Streptococcus, *HSV-2*, *Trichomonas vaginalis*, and *Gardnerella vaginalis* – between January 2020 and May 2025. Male patients who underwent multiplex PCR-based NAAT testing for suspected sexually transmitted infections were eligible for inclusion. Testing was performed using the Anatolia Geneworks STD Multiplex Real-Time Panel (Anatolia, Istanbul, Turkey) following the manufacturer's instructions.

Patients with incomplete urinalysis or symptom data, missing laboratory parameters, duplicate tests (n = 149), age younger than 18 (n = 3), or older than 65 (n = 13) were excluded from the analysis. After applying these criteria, a total of 532 patients were included in the final analysis.

Clinical variables

Demographic and clinical data, including age, urinalysis parameters – leukocyte esterase (LE), nitrite, erythrocyte (RBC), and leukocyte counts (WBC) – and symptom clusters (dysuria, urethral discharge, urethral discomfort, lower urinary tract symptoms, genitourinary inflammatory symptoms, and sexual/reproductive dysfunction) were extracted from electronic medical records. The NAAT result was coded as a binary reference outcome (positive = 1; negative = 0).

Symptoms were categorized into six binary groups: urethral discharge; dysuria; urethral discomfort; sexual and reproductive dysfunction (painful ejaculation, erectile dysfunction, infertility); genitourinary inflammatory symptoms (testicular or scrotal pain, epididymitis, chronic pelvic pain); and lower urinary tract symptoms (frequency, urgency, hesitancy, weak stream). Asymptomatic patients were coded as 0 for all symptom groups.

Urinalysis variables

Four routine urinalysis markers were binarized as follows: leukocyte esterase (LE) ≥ 1+ = 1 (positive); nitrite positive = 1; white blood cells (WBC) > 5/high-power field (HPF) = 1; and red blood cells (RBC) > 3/HPF = 1. The NAAT result served as the binary outcome variable (1 = positive; 0 = negative).

Feature preparation and handling of missing data

Patients without urinalysis data were excluded from model development. Continuous variables were standardized using a *z*-score transformation, while categorical and binary variables were used without additional scaling. The dataset was divided temporally into training and testing cohorts to simulate real-world deployment. The training set comprised patients from 2020 to 2024, while those from 2025 formed the independent temporal test set. This method ensured that model performance was assessed on future, unseen data, thereby preventing information leakage and better reflecting the conditions of clinical implementation.

Model development

Five supervised learning algorithms were evaluated: 1. Logistic Regression (LR). 2. Random Forest (RF). 3. Gradient Boosting Machine (GBM – XGBoost). 4. Support Vector Machine (SVM, radial basis kernel). 5. Naive Bayes (NB).

Hyperparameters were optimized using grid search with five-fold stratified cross-validation on the training set. Model discrimination was assessed using the area under the receiver operating characteristic curve (AUC-ROC) and area under the precision–recall curve (AUC-PR).

Rule-out threshold and performance metrics

To simulate a clinical rule-out application, model probability thresholds were calibrated to achieve a sensitivity of ≥ 95% on the validation folds. At this fixed safety target, specificity, negative predictive value (NPV), the proportion of potentially avoidable NAATs (true negatives + false positives/total), and missed positives (false-negative rate) were calculated on the independent test set. The model achieving the highest NPV and AUC while maintaining ≥ 95% sensitivity was selected as the optimal rule-out model.

Threshold selection and evaluation strategy

To operationalize a safety-first rule-out strategy, the probability threshold was selected as the lowest cut-point that achieved a sensitivity of ≥ 95% and preferably ≥ 90% NPV on the validation folds. This conservative "no-miss" calibration approach prioritizes the avoidance of false-negative classifications, aligning with principles of diagnostic stewardship. ROC-AUC was reported only to describe global discrimination ability; however, sensitivity and NPV served as the primary performance metrics for clinical applicability. The final threshold (0.079) was determined based on this criterion and applied to the independent temporal test set.

Model interpretation

For interpretability, SHAP (Shapley Additive Explanations) analysis was applied to the best-performing model to quantify the contribution of each variable to prediction outcomes. Predictor importance was visualized via mean absolute SHAP values.

Ethics approval

This study protocol was approved by the Adnan Menderes University Institutional Ethics Committee (Approval No. 2025/193). Because of its retrospective and anonymized design, informed consent was waived. The study was conducted in accordance with the ethical standards of the institutional research committee and the 1964 Declaration of Helsinki and its subsequent amendments.

Statistical analysis

All analyses were performed using Python 3.13 (within a virtual environment) on macOS. The following packages were used: xgboost==1.7.6, shap==0.44.0, numpy==1.26.4, along with pandas, matplotlib, seaborn, and scikit-learn. All analysis scripts were run in a reproducible environment and stored in a structured directory (NAAT SET/naat_outputs/). Statistical significance was set at $p < 0.05$ (two-tailed). Continuous variables are presented as mean ± SD, and categorical variables as frequencies.

Results

Patient characteristics

A total of 532 male patients who underwent NAAT testing and had complete urinalysis data were included in the final analysis. The mean age of the study population was 39.0 ±12.3 years, with no significant age difference between NAAT-negative and NAAT-positive patients (38.93 ±12.34 vs. 39.05 ±12.27, $p = 0.918$).

Overall, 34.8% ($n = 185$) of patients had a positive NAAT result, while 65.2% ($n = 347$) tested neg-

ative. Including cases with multiple detections, the total number of detections was 234. *Ureaplasma* spp. accounted for half of all detections (*n* = 118; 50.4%), followed by *Mycoplasma hominis* (*n* = 35; 15.0%), *Mycoplasma genitalium* (*n* = 25; 10.7%), *Chlamydia trachomatis* (*n* = 26; 11.1%), and *Neisseria gonorrhoeae* (*n* = 9; 3.8%). Other pathogens included Group B *Streptococcus* (*n* = 15; 6.4%), *HSV*-2 (*n* = 4; 1.7%), *Trichomonas vaginalis* (*n* = 1; 0.4%), and *Gardnerella vaginalis* (*n* = 1; 0.4%).

Leukocyte esterase (LE) and pyuria (WBC) were significantly more common in NAAT-positive individuals (LE: 21.6% vs. 10.4%, $p < 0.001$; WBC: 28.6% vs. 13.5%, $p < 0.001$). Nitrite and microscopic hematuria (RBC) positivity showed no significant association with NAAT outcomes ($p = 0.252$ and $p = 0.800$, respectively).

Regarding clinical presentation, 71.4% of patients reported genitourinary symptoms, and 28.6% (*n* = 152) were asymptomatic. Dysuria (20.7%), urethral discharge (20.9%), and urethral discomfort (22.0%) were the most frequently reported symptoms. Among these, only urethral discomfort was significantly higher in NAAT-positive patients (27.0% vs. 19.3%, $p = 0.041$). Other symptom groups, including lower urinary tract symptoms (LUTS) and sexual/reproductive symptoms, did not differ significantly between groups ($p > 0.05$).

A detailed breakdown of clinical and urinalysis characteristics by NAAT status is provided in Table I.

Machine-learning model results

A comprehensive machine-learning pipeline was implemented to develop a rule-out model for identifying patients at low likelihood of NAAT positivity. Models tested included logistic regression, random forest, support vector machine (SVM), naïve Bayes, and XGBoost.

Data from January 2020 to December 2024 (*n* = 455) were used for model training, and data from January to May 2025 (*n* = 75) served as an external temporal test set. Model thresholds were optimized to ensure high sensitivity (≥ 95%) and a high negative predictive value (NPV ≥ 90%), pri-

Table I. Crosstab and group comparison results between variables and NAAT results

Variable	Category	NAAT = 0 *n* (%)	NAAT = 1 *n* (%)	Total *n* (%)	*P*-value
Age [years]	–	38.93 ±12.34	39.05 ±12.27	–	0.918[t]
Leukocyte esterase (LE)	0	311 (89.6)	145 (78.4)	456 (85.7)	< 0.001[c]
	1	36 (10.4)	40 (21.6)	76 (14.3)	
Nitrite	0	341 (98.3)	184 (99.5)	525 (98.7)	0.252[c]
	1	6 (1.7)	1 (0.5)	7 (1.3)	
RBC	0	321 (92.5)	170 (91.9)	491 (92.3)	0.800[c]
	1	26 (7.5)	15 (8.1)	41 (7.7)	
WBC	0	300 (86.5)	132 (71.4)	432 (81.2)	< 0.001[c]
	1	47 (13.5)	53 (28.6)	99 (18.6)	
Symptom present	0	101 (29.1)	51 (27.6)	152 (28.6)	0.708[c]
	1	246 (70.9)	134 (72.4)	380 (71.4)	
Dysuria	0	275 (79.3)	147 (79.5)	422 (79.3)	0.955[c]
	1	72 (20.7)	38 (20.5)	110 (20.7)	
Discharge	0	282 (81.3)	139 (75.1)	421 (79.1)	0.097[c]
	1	65 (18.7)	46 (24.9)	111 (20.9)	
Urethral discomfort	0	280 (80.7)	135 (73.0)	415 (78.0)	0.041[c]
	1	67 (19.3)	50 (27.0)	117 (22.0)	
Sexual and reproductive dysfunction	0	297 (85.6)	157 (85.3)	454 (85.5)	0.934[c]
	1	50 (11.4)	27 (14.7)	77 (14.5)	
Genitourinary inflammatory symptoms	0	301 (86.7)	160 (86.5)	461 (86.7)	0.934[c]
	1	46 (13.3)	25 (13.5)	71 (13.3)	
LUTS-related symptoms	0	305 (87.9)	172 (93.0)	477 (89.7)	0.067[c]
	1	42 (12.1)	13 (7.0)	55 (10.3)	
Asymptomatic	0	246 (70.9)	134 (72.4)	380 (71.4)	0.708[c]
	1	101 (29.1)	51 (27.6)	152 (28.6)	

[c] χ^2 test, [t] Student's t-test. NAAT – nucleic acid amplification test, LUTS – lower urinary tract symptoms, RBC – red blood cells, WBC – white blood cells (leukocytes).

Table II. Comparative evaluation of machine learning algorithms for diagnostic rule-out modeling in NAAT testing

Model	ROC-AUC	Sensitivity	Specificity	NPV	Accuracy	Rule-Out Rate	Threshold
XGBoost (Gradient Boosting)	0.86	0.982	0.045	0.914	0.502	0.442	0.079
Random Forest	0.83	0.979	0.042	0.909	0.499	0.439	0.080
Logistic Regression	0.79	0.972	0.038	0.900	0.487	0.435	0.082
SVM (RBF)	0.77	0.970	0.036	0.897	0.485	0.432	0.083
Naïve Bayes	0.71	0.965	0.031	0.890	0.476	0.427	0.086

oritizing patient safety and minimizing the risk of false negatives. The selected rule-out threshold (0.079) represented the lowest predicted-probability cut-point achieving these safety criteria and was applied to the independent test cohort.

Among all tested machine-learning algorithms, XGBoost provided the most favorable rule-out performance, characterized by a sensitivity of 0.98, a negative predictive value of 0.91, and a 44.2% reduction in NAAT testing, with specificity at 0.045 and an overall accuracy of approximately 50%. A comprehensive comparison of algorithm performance across sensitivity, specificity, NPV, and rule-out rate is presented in Table II.

Random forest and logistic regression showed similar discrimination performance, while naïve Bayes had the lowest predictive ability. ROC curves and global performance metrics are shown in Figure 1, and rule-out threshold–performance trade-offs are depicted in Figure 2. Interpretability analysis using SHAP demonstrated that urinary inflammatory markers – particularly WBC positivity, leukocyte esterase, and nitrite positivity – along with age, dysuria, and urethral discomfort were primary contributors to NAAT positivity. Conversely, the absence of urinary inflammation and asymptomatic presentation were strongly predictive of NAAT negativity. These feature importance patterns and individual feature contributions are illustrated in the mean absolute SHAP plot and the SHAP beeswarm visualization (Figures 3 and 4).

Discussion

Sexually transmitted infections (STIs) often manifest as urethritis in men – particularly those visiting urology or sexual health clinics – with symptoms ranging from urethral discharge, dysuria, and itching to atypical signs such as pelvic pain or infertility [2–4]. Inflammation of the genitourinary tract, indicated by leukocyturia,

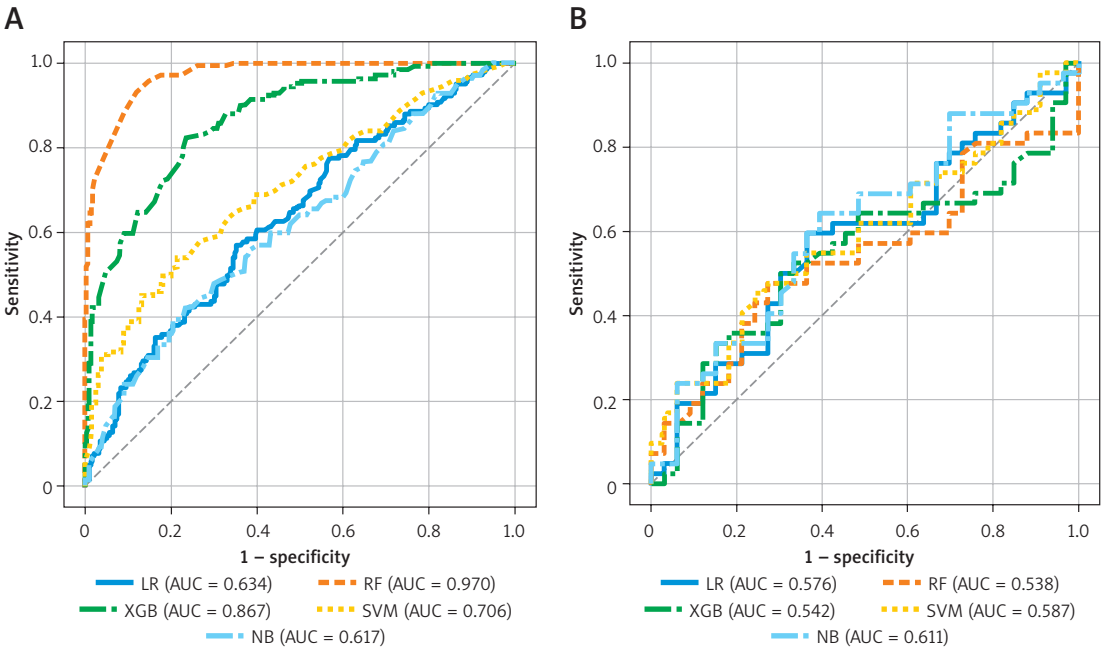

Figure 1. ROC curves for model discrimination in the training (**A**) and temporal external test (**B**) cohorts. The figure shows the ROC performance of five supervised ML algorithms: Logistic Regression (LR), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine with radial basis function kernel (SVM-RBF), and Naïve Bayes (NB). ROC curves illustrate overall discrimination ability, with the corresponding area under the ROC curve (AUC) values displayed in the legend for each model. These ROC curves are provided primarily to demonstrate overall ranking and discrimination performance. Decision thresholds were chosen based on rule-out criteria (sensitivity ≥95% and high negative predictive value), rather than ROC-derived optimal cut-points

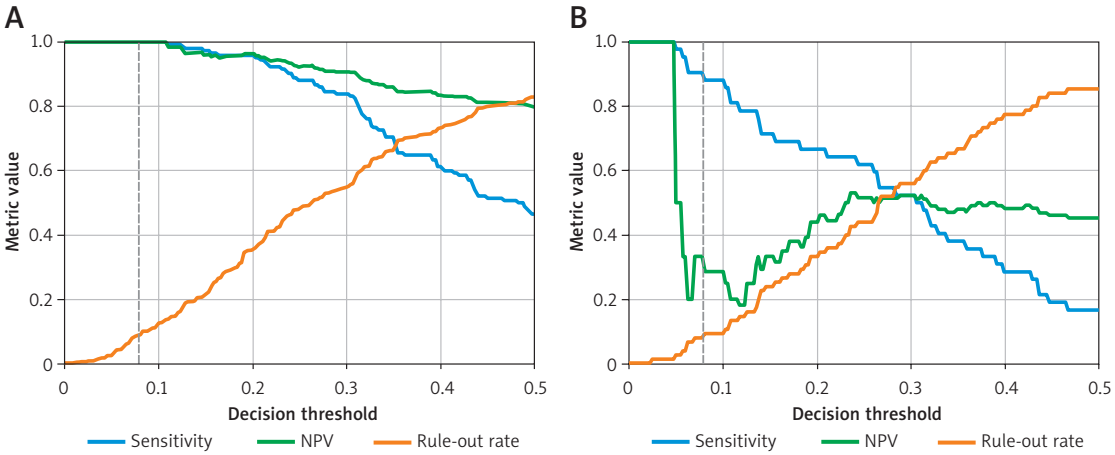

Figure 2. Rule-out operating curves demonstrating threshold–performance trade-offs. Threshold–performance curves showing sensitivity, NPV, and avoided-testing rate. The selected rule-out threshold (dashed line) reflects the lowest cut-off maintaining ≥ 95% sensitivity with high NPV, enabling safe reduction in NAAT utilization

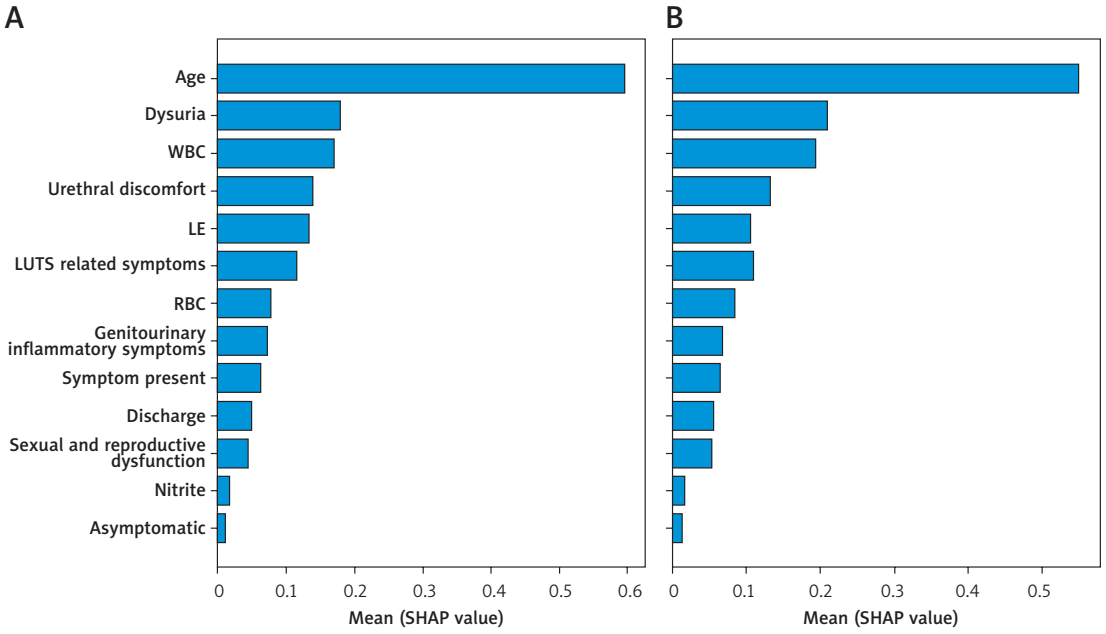

Figure 3. Overall feature importance based on mean absolute SHAP values

WBC – white blood cell count, LE – leukocyte esterase, RBC – red blood cell count.

positive leukocyte esterase, or microhematuria, can help distinguish infectious from noninfectious cases, though these markers are not perfectly sensitive or specific, especially in sexually active men with overlapping urinary and genital symptoms [18, 19]. In this context, pairing routine urinalysis results with symptom profiles using machine learning (ML) may provide a more accurate estimate of infection likelihood than traditional single-variable methods. Therefore, this study aimed to develop an ML-based rule-out model using low-cost urinalysis parameters and common symptom clusters – such as dysuria, discharge, and urethral discomfort – to identify male patients with a high probability of negative NAAT results, thereby reducing unnecessary testing.

Five ML algorithms were used to predict and rule out NAAT positivity among male patients tested for genital infections. Unlike traditional statistical methods, ML models can detect complex nonlinear relationships and handle diverse, interconnected medical data. In urology, these algorithms – including logistic regression, random forest, support vector machine, k-nearest neighbors, and naïve Bayes – have been effectively applied in various areas, such as predicting postoperative outcomes after radical cystectomy [20], estimating prostate cancer risk in PI-RADS 3 lesions [21], forecasting urethral stricture after transurethral prostate resection [22], creating predictive models for the risk of infertility in men [23], and modeling infertility or varicocele severity [24]. These

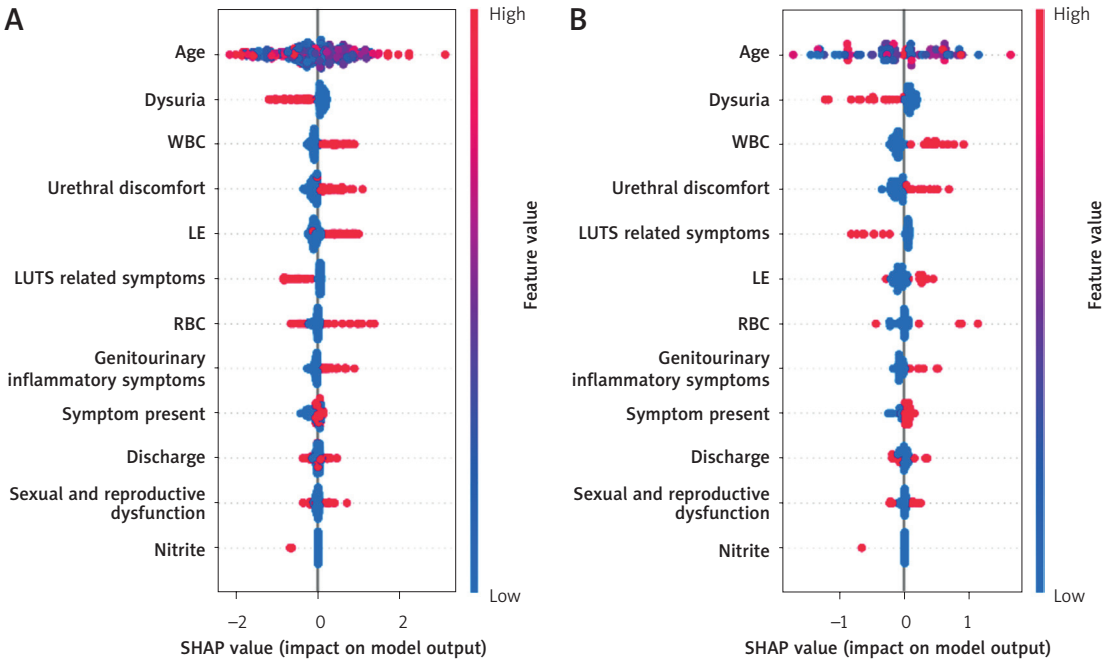

Figure 4. SHAP beeswarm plot showing individual feature contributions. This plot displays the distribution of feature impacts across individual predictions in the gradient boosting model. Each dot represents a patient's contribution, with color indicating feature value and horizontal position reflecting its effect on the model output. Inflammatory urinalysis markers and symptom variables exhibited the most substantial directional influence, supporting the biological plausibility of the model's decision process in sexually transmitted infection triage

WBC – white blood cell count, LE – leukocyte esterase, RBC – red blood cell count.

wide-ranging applications highlight how ML can enhance diagnostic accuracy, support personalized decision-making, and improve cost-efficiency in urological care.

Comparable studies have shown the potential of ML to improve infectious disease diagnostics. In their 2022 web-based risk calculator, Xu *et al.* developed an interpretable tool to predict HIV and three major STIs using demographic and behavioral variables, achieving AUCs of about 0.70–0.84 [17]. These results highlight the feasibility of deploying ML-driven triage systems for early risk stratification and targeted testing in sexual health settings. Similarly, in a large cohort of men who have sex with men, Bao *et al.* demonstrated that ML approaches consistently outperformed multivariable logistic regression in predicting HIV and other major STIs [25]. Gradient-boosting models achieved the highest performance, with AUCs of 0.763 for HIV, 0.858 for syphilis, 0.755 for gonorrhea, and 0.680 for chlamydia, outperforming logistic regression across all outcomes [25]. Overall, previous ML-based STI studies have mainly focused on diagnostic prediction – finding infected individuals – while our work shifts the focus toward safe rule-out and diagnostic stewardship, enabling the selective use of NAAT. This approach contrasts with traditional case detection models by emphasizing the safe delay of molecular diagnostics rather than simply maximizing the detection of positive cases. By combining symptom clusters and urinalysis biomarkers, our model safely reduces reflex NAAT testing while maintaining diagnostic accuracy, reinforcing the importance of interpretable ML tools in supporting diagnostic stewardship and resource optimization.

In urinary diagnostics, Del Ben *et al.* reported an interpretable decision-tree model that excluded negative urine cultures with 94.5% sensitivity, reducing laboratory workload by 16% without compromising safety [16]. Similarly, Farashi *et al.* demonstrated that ensemble learning (XGBoost, decision tree, LightGBM) achieved an AUC of 0.885 for UTI prediction, with leukocyte esterase and WBC count emerging as prominent features [26]. These studies support the application of ML for microbiologic triage, but mainly focus on diagnostic prediction, while our approach emphasizes a safety-focused "rule-out" threshold to reduce unnecessary use of NAATs.

Additionally, our work aligns with recent advances in AI-assisted STI triage. Soe et al. integrated symptom and image-based features to predict 12 common STIs, achieving AUCs from 0.81 to 0.95, although performance was lower for urinary tract infections (AUC = 0.72) [27]. While their approach relied solely on symptoms, our model incorporates objective urinalysis biomarkers to support safe NAAT deferral. With an AUC of 0.86 and the ability to avoid NAAT testing in 44.2% of patients while maintaining 98% sensitivity and 91% NPV, our framework offers a clinically viable,

stewardship-oriented strategy for outpatient STI evaluation.

Rule-out ML frameworks are also emerging in other infectious contexts. In emergency medicine, Schinkel *et al.* demonstrated that blood cultures could be safely withheld in about 30% of cases using an ML-based decision tool [28]. This approach aligns with our strategy, which focuses on minimizing false negatives while reducing unnecessary molecular testing. In comparison, our model achieved a 44.2% reduction in NAAT utilization with very high sensitivity (~98%) and a strong NPV (~91%), indicating that rule-out algorithms may offer particularly significant benefits in urogenital infection pathways.

Taken together, our results position this model as an interpretable, data-efficient, and clinically aligned tool for the selective use of NAAT. By utilizing routine urinalysis and symptom-based risk stratification instead of specialized biomarkers, the model enables targeted NAAT deployment, reduces unnecessary molecular testing, and supports scalable, cost-effective STI care. Notably, this approach enhances diagnostic stewardship by promoting ML-based triage rather than reflex testing, which is particularly important in health systems where the increasing volume of NAATs presents significant financial and operational challenges.

The proposed model also has the potential to be integrated into electronic laboratory ordering systems, where real-time feedback could identify patients with a low predicted likelihood of STI and suggest deferring NAAT testing. Beyond STIs, this framework could be adapted for other infectious syndromes, such as urinary tract infections or bacterial vaginosis, where rule-out strategies are clinically acceptable. Such implementation would be a significant step toward sustainable, data-driven infection management across healthcare settings.

Another strength of this study is its transparent and interpretable modeling approach. Instead of relying on black-box architectures, we used well-established algorithms such as logistic regression, random forest, and support vector machine, which help identify the most influential predictors of NAAT positivity. Variables such as leukocyte esterase, urinary WBC, and genitourinary inflammatory symptoms consistently proved to be the most informative, supporting the biological plausibility of the model's predictions and boosting its potential clinical acceptance.

This study has several limitations. First, the overall ROC–AUC performance ranged from 0.58 to 0.63, aligning with a rule-out goal rather than a diagnostic classification goal. The model is not designed for rule-in purposes and should not be used to confirm infection; its primary strength lies in achieving high sensitivity and NPV to exclude low-risk patients from NAAT testing safely. Despite the excellent rule-out performance, the model was developed using retrospective data from a single center and may need recalibration in settings with different patient populations, testing platforms, or pathogen prevalence. Since NPV and sensitivity depend on prevalence, threshold adjustments might be necessary in regions with significantly different STI epidemiology. Before clinical use, prospective external validation in independent cohorts is essential, along with monitoring false-negative cases to ensure the test's safety, and preferably should be performed in larger and multicenter populations to confirm the model's generalizability. Additionally, because the number of pathogen-positive cases was relatively small, pathogen-specific performance, microbiological subtyping (e.g., *Chlamydia trachomatis*, *N. gonorrhoeae*, *Ureaplasma/Mycoplasma* spp.), and mixed-infection analyses were not performed separately at this stage. Pathogen-specific models could be developed once larger multicenter datasets are available. If necessary, semi-supervised model updating or threshold fine-tuning may enhance performance in real-world applications. Future work will include external multicenter validation and dynamic threshold adjustment to support safe clinical deployment.

In conclusion, this study demonstrates that machine learning models utilizing routine urinalysis parameters and symptom profiles can accurately identify male patients at low risk for sexually transmitted infections. By confidently excluding unnecessary NAAT testing, such models may enhance diagnostic stewardship, reduce healthcare costs, and support improved clinical decision-making. With prospective validation and seamless integration into clinical workflows, this approach has the potential to be a scalable and practical tool for improving STI diagnostics in routine care.

Acknowledgments

We thank Mecostech Software Technologies Inc., Istanbul, Turkey, for technical assistance and support during model development and visualization processes.

Funding

No external funding.

Ethical approval

This study protocol was approved by the Adnan Menderes University Institutional Ethics Committee (Approval No. 2025/193). The study was conducted in accordance with the ethical standards of the institutional research committee and with

the 1964 Helsinki Declaration and its later amendments.

Conflict of interest

The authors declare no conflict of interest.

References

1. Global progress report on HIV, viral hepatitis and sexually transmitted infections, 2021. Accountability for the global health sector strategies 2016–2021: actions for impact. Geneva: World Health Organization; 2021.
2. Moi H, Blee K, Horner PJ. Management of non-gonococcal urethritis. BMC Infect Dis 2015; 15: 294.
3. Brunner H, Weidner W, Schiefer HG. Studies on the role of Ureaplasma urealyticum and Mycoplasma hominis in prostatitis. J Infect Dis 1983; 147: 807-13.
4. Paira DA, Molina G, Tissera AD, Olivera C, Molina RI, Motrich RD. Results from a large cross-sectional study assessing Chlamydia trachomatis, Ureaplasma spp. and Mycoplasma hominis urogenital infections in patients with primary infertility. Sci Rep 2021; 11: 113655.
5. Martin DH. Urethritis in males. In: Sexually Transmitted Diseases. Holmes KK, Sparling PF, Stamm WE (eds.). McGraw-Hill, New York 2008; 1107-26.
6. Bradshaw CS, Tabrizi SN, Read TR, et al. Etiologies of nongonococcal urethritis: bacteria, viruses, and the association with orogenital exposure. J Infect Dis 2006; 193: 336-45.
7. Workowski KA, Bachmann LH, Chan PA, et al. Sexually transmitted infections treatment guidelines, 2021. MMWR Recomm Rep 2021; 70: 1-187.
8. Martin DH. Genital mycoplasmas: Mycoplasma genitalium, Mycoplasma hominis, and Ureaplasma species. In: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. Bennett JE, Dolin R, Blaser MJ (eds.). Elsevier, Philadelphia 2020; 2340-3.
9. Mena L, Wang X, Mroczkowski TF, Martin DH. Mycoplasma genitalium infections in asymptomatic men and men with urethritis attending a sexually transmitted diseases clinic in New Orleans. Clin Infect Dis 2002; 35: 1167-73.
10. Yasynetskyi M, Banyra O, Nikitin O, et al. Mixed sexually transmitted infections in infertile couples: empirical treatment and influence on semen quality. Recent Adv Antiinfect Drug Discov 2021; 16: 227-36.
11. Horner P, Donders G, Cusini M, Gomberg M, Jensen JS, Unemo M. Should we be testing for urogenital Mycoplasma hominis, Ureaplasma parvum and Ureaplasma urealyticum in men and women? A position statement from the European STI Guidelines Editorial Board. J Eur Acad Dermatol Venereol 2018; 32: 1845-51.
12. British Association for Sexual Health and HIV (BASHH). Position statement on the inappropriate use of multiplex testing platforms and suboptimal antibiotic treatment regimens for bacterial sexually transmitted infections. London: BASHH; 2021. Available from: https://www.bashh.org/news/1430/. Accessed 04 Nov 2025.
13. Kranz J, Bartoletti R, Bruyère F, et al. European Association of Urology guidelines on urological infections: summary of the 2024 guidelines. Eur Urol 2024; 86: 27-4.
14. Sheele JM, Campbell RL, Jones DD. Machine learning to predict bacteriuria in the emergency department. Sci Rep 2025; 15: 31087.
15. Li J, Cheng Y, Tang W, Jing Y, Chen Y, Tu S. Machine learning prediction of early hypothermia in sepsis patients. Arch Med Sci 2025; Epub 2025 Oct 29. doi: 10.5114/aoms/209620.
16. Del Ben F, Da Col G, Cobârzan D, et al. A fully interpretable machine learning model for increasing the effectiveness of urine screening. Am J Clin Pathol 2023; 160: 620-32.
17. Xu X, Yu Z, Ge Z, et al. Web-based risk prediction tool for an individual's risk of HIV and sexually transmitted infections using machine learning algorithms: development and external validation study. J Med Internet Res 2022; 24: e37850.
18. Stamm WE, Hooton TM. Management of urinary tract infections in adults. N Engl J Med 1993; 329: 1328-34.
19. Lammers RL, Gibson S, Kovacs D, Sears W, Strachan G. Comparison of test characteristics of urine dipstick and urinalysis at various test cutoff points. Ann Emerg Med 2001; 38: 505-12.
20. Wang G, Lam KM, Deng Z, Choi KS. Prediction of mortality after radical cystectomy for bladder cancer by machine learning techniques. Comput Biol Med 2015; 63: 124-32.
21. Altıntaş E, Şahin A, Erol S, et al. Navigating the gray zone: machine learning can differentiate malignancy in PI-RADS 3 lesions. Urol Oncol 2025; 43: 195.e11-e20.
22. Altıntaş E, Şahin A, Babayev H, et al. Machine learning algorithm predicts urethral stricture following transurethral prostate resection. World J Urol 2024; 42: 324.
23. Koç S, Tomak L, Karabulut E. A predictive model for the risk of infertility in men using machine learning algorithms. J Urol Surg 2022; 9: 265-71.
24. Kayra MV, Şahin A, Toksöz S, et al. Machine learning-based classification of varicocele grading: a promising approach for diagnosis and treatment optimization. Andrology 2025; 13: 1451-61.
25. Bao Y, Medland NA, Fairley CK, et al. Predicting the diagnosis of HIV and sexually transmitted infections among men who have sex with men using machine learning approaches. J Infect 2021; 82: 48-59.
26. Farashi S, Momtaz HE. Prediction of urinary tract infection using machine learning methods: a study for finding the most-informative variables. BMC Med Inform Decis Mak 2025; 25: 13.
27. Soe NN, Towns JM, Latt PM, et al. Accuracy of symptom checker for the diagnosis of sexually transmitted infections using machine learning and Bayesian network algorithms. BMC Infect Dis 2024; 24: 1408.
28. Schinkel M, Boerman AW, Bennis FC, et al. Diagnostic stewardship for blood cultures in the emergency department: a multicenter validation and prospective evaluation of a machine learning prediction tool. EBioMedicine 2022; 82: 104176.