International Urology and Nephrology
https://doi.org/10.1007/s11255-026-05308-9

NEPHROLOGY – ORIGINAL PAPER

Machine learning models for early detection of urinary tract infections in kidney transplant patients

Alejandro Camargo-Salamanca[1] · Liceth Viviana Castañeda-Silva[2] · Carlos Puentes-Morales[3] · Alejandro Duitama-Leal[3] · Andrés Cardona-Mendoza[4] · Alejandro Ramos-Casallas[4] · Sandra J. Perdomo[4] · Consuelo Romero-Sánchez[4] · Julia Andrea Gomez-Montero[5] · Andrea Garcia-Lopez[5] · Fernando Girón-Luque[5,6]

Received: 24 September 2025 / Accepted: 21 July 2026

Abstract

Background Renal transplantation is the preferred treatment for end-stage chronic kidney disease but requires lifelong immunosuppression, increasing the risk of infections such as urinary tract infection (UTI). UTI in kidney transplant recipients can lead to serious complications, including acute kidney injury, reduced graft survival, and increased mortality. Machine learning can enhance risk detection accuracy and support proactive management of complications. This study aims to develop a machine learning-based approach for predicting urinary tract infections in kidney transplant recipients, facilitating the early identification of high-risk individuals.

Methods Retrospective cohort of 29,340 anonymized records from a Colombian transplant center to develop and compare machine learning models for predicting UTI in post-kidney transplant patients. Fifteen machine learning models were evaluated using 41 sociodemographic and clinical variables. Model performance was assessed using metrics such as accuracy and area under the ROC curve (AUC).

Results Both models are good, with Random Forest achieving 93.57% accuracy, 94.43% AUC, 92.5% sensitivity, and 95.2% specificity. XGBoost achieved 91.14% accuracy, 91.65% AUC, 90.8% sensitivity, and 92.6% specificity. Feature importance analysis revealed that key predictors included Body mass index (BMI), use of mofetil mycophenolate or sodium mycophenolate, donor type, hypertension, diabetes mellitus, use of steroids, re-transplantation, serum levels of tacrolimus and sex of the recipient.

Conclusion This study demonstrates the potential of machine learning models to predict UTI in kidney transplant patients, offering a tool for early intervention and personalized care. Implementing these models in clinical settings could improve patient outcomes and reduce healthcare costs by preventing severe complications associated.

Keywords Kidney transplant · Urinary tract infection · Machine learning · Risk factors · Predictive value of tests

✉ Andrea Garcia-Lopez
aegarcia@colombianadetrasplantes.com

[1] Nephrology Department, Colombiana de Trasplantes, Av. Carrera 30 # 47 A – 47., Bogotá DC, Colombia

[2] Virtualization and Artificial Intelligence Advanced Solutions Laboratory (SavIA-Lab), Universidad El Bosque, Bogotá, Colombia

[3] Virtualization and Artificial Intelligence Advanced Solutions Laboratory (SavIA-Lab), SIGNOS Group, Science Faculty, Universidad El Bosque, Bogotá, Colombia

[4] Virtualization and Artificial Intelligence Advanced Solutions Laboratory (SavIA-Lab), Cellular and Molecular Immunology Group (INMUBO), Universidad El Bosque, Bogotá, Colombia

[5] Research Department, Colombiana de Trasplantes, Bogotá, Colombia

[6] Transplant Surgery, Colombiana de Trasplantes, Bogotá, Colombia

Introduction

Kidney transplantation is the best therapeutic option for treating end-stage chronic kidney disease (CKD) [1]. To prevent rejection, patients must take immunosuppressive medications; however, the use of immunosuppression increases the risk of infection [2]. Among the infections affecting transplant patients, urinary tract infection (UTI) is the most frequent and it is associated with acute kidney injury (AKI), decreased graft survival, and increased mortality [3, 4]. Additionally, they lead to increased treatment costs, reduced quality of life, and greater time devoted to caregiving by family members [5].

Multiple ML-based strategies have been developed to support UTI identification, most commonly leveraging routinely available clinical and laboratory data [6, 7] Frequently used predictors include demographic characteristics (e.g., age and sex), symptoms such as dysuria, urinary frequency, suprapubic pain, and foul-smelling urine, and standard urinalysis finding particularly leukocyte esterase, nitrites, pyuria, hematuria, bacteriuria, and epithelial cells. Some models also incorporate prior urine culture results and comorbidities such as diabetes. In contrast, several biomarkers (e.g., NGAL, IL-1β, MMP9, CXCL8) have been explored mainly in research settings and are not part of routine UTI assessment in most transplant programs**.** Overall, this heterogeneity of candidate predictors highlights the potential for ML models to integrate routinely collected data to improve early risk prediction and timely recognition of UTI.

However, despite the high burden and clinical consequences of UTI after kidney transplantation, early recognition remains challenging in routine care, particularly because immunosuppression can attenuate typical symptoms and because clinical decision-making often relies on heterogeneous information from symptoms, urinalysis, prior microbiology, and comorbidities. In this setting, delays in identifying patients at increased risk may translate into potentially avoidable morbidity and graft-related complications. Moreover, most published ML models have been developed in specific healthcare systems with distinct patient profiles, antimicrobial resistance patterns, laboratory workflows, and follow-up practices; as a result, their performance and calibration may not generalize to other regions without local validation and adaptation [6, 7]. Consequently, there is a need for context-specific, early-stage prediction models that leverage variables routinely captured in real-world practice and can be implemented within transplant programs. Therefore, this study aims to develop and compare machine learning models to enable early identification of UTI in kidney transplant recipients, leveraging routinely available clinical variables from a Latin American population.

Methodology

Study design

In this retrospective cohort study, we developed and compared machine learning models to predict symptomatic urinary tract infection (UTI) within the first post-transplant year among kidney transplant recipients treated at Colombiana de Trasplantes. The analytic dataset comprised sociodemographic and clinical information collected as part of routine care. A total of 51 variables were available; 41 were used as candidate predictors, while 10 corresponded to identification/administrative fields and were not used for model training. The primary endpoint was a clinically documented symptomatic UTI requiring assessment and treatment, encompassing both lower UTI (e.g., cystitis) and upper UTI (e.g., pyelonephritis).

Patients and methods

We used routinely collected data from adult kidney transplant recipients followed at Colombiana de Trasplantes between January 2017 and March 2023. Eligible patients were ≥18 years old, had a registered kidney transplant, and had follow-up at our center. Patients with <1 year of follow-up were excluded.

An initial extraction from the institutional data warehouse yielded 29,340 anonymized records, which reflect clinical documentation entries and follow-up records and therefore may include multiple records per patient across time. After selection criteria were applied and deduplication using unique patient identifiers, these records corresponded to 959 unique kidney transplant recipients with UTI. Because the dataset was imbalanced, we applied class balancing for model development, resulting in a final patient-level analytic dataset of 1,918 observations (959 UTI; 959 non-UTI) used for training, validation, and testing.

UTI definition

The outcome was defined as ≥1 symptomatic UTI within the first post-transplant year. Symptomatic UTI was established when the medical record documented urinary symptoms and/or systemic signs compatible with infection (e.g., dysuria, urgency/frequency, suprapubic discomfort, flank pain, fever) with supportive paraclinical evidence when available (urinalysis and/or urine culture), operationalized as a clinically diagnosed and treated episode. To reduce misclassification, episodes consistent with colonization/asymptomatic bacteriuria (positive culture without compatible symptoms/signs and/or explicitly documented as asymptomatic and not treated) were not counted as outcome events, whereas

episodes with compatible symptoms/signs and antimicrobial treatment for UTI, with or without culture confirmation, were classified as symptomatic UTI events for the primary endpoint.

Given the limitations of retrospective clinical records to reliably discriminate among syndromic subtypes (e.g., urethritis, cystitis, and pyelonephritis), we modeled a composite symptomatic UTI endpoint rather than subtype-specific diagnoses.

Imputation

Two imputation techniques were explored: Multiple Imputation by Chained Equations (MICE) and mean imputation.

MICE is a widely used approach that models each variable conditionally on the others to generate multiple plausible values for missing data. This iterative method effectively captures relationships between variables, making it a statistically robust technique, particularly when data are *Missing at Random* (MAR). However, MICE can be computationally demanding and may introduce variability if not properly fine-tuned.

Mean imputation is a simpler and more computationally efficient method, where missing values are replaced with the mean of the observed values for that variable. While this technique preserves the overall mean of the dataset, it may underestimate variance and distort correlations between variables. Despite these limitations, mean imputation is often preferred when computational efficiency is a priority and when statistical tests confirm that it maintains the original data distribution effectively.

Machine learning modeling

Because the outcome distribution in the original cohort was imbalanced, we used random under-sampling (underbalancing) of the majority class during model development. Specifically, we retained all patients with UTI within the first post-transplant year and randomly selected an equal number of patients without UTI to match the minority class. This resulted in a final patient-level balanced analytic dataset of 1918 observations (959 UTI; 959 non-UTI), which was used for model training, validation, and testing.

A comparison of 15 classification machine learning models was conducted to predict urinary tract infection (UTI) in kidney transplant patients. These models include Light Gradient Boosting Machine, Extra Trees Classifier, Random Forest Classifier, Extreme Gradient Boosting, Decision Tree Classifier, Gradient Boosting Classifier, AdaBoost Classifier, K Neighbors Classifier, Logistic Regression, Quadratic Discriminant Analysis, Naive Bayes, Ridge Classifier, Linear Discriminant Analysis, SVM - Linear Kernel, and Dummy Classifier.

Feature selection

In this study, two variable selection methods were explored: SelectKBest and Featurewise selection.

SelectKBest allows the selection of the most relevant variables for prediction by performing individual statistical tests between each variable and the target outcome. Numerical variables are typically evaluated using ANOVA tests (F-score), and categorical variables using Chi-square tests. Variables with higher statistical scores (indicating stronger predictive power) are retained, simplifying and improving the predictive model.

Featurewise selection involves independently analyzing each variable or feature in the dataset to assess its individual predictive capability. By evaluating variables separately based on their unique contributions, this method identifies and retains only those that significantly enhance the model's predictive accuracy, resulting in a simpler and more robust model.

To select the most relevant variables in predicting UTI in kidney transplant patients, the Featurewise variable selection technique was used. This technique involved an iterative process in which the individual characteristics of each variable were evaluated in relation to the study's objective. Importance metrics were calculated for each variable, such as mutual information and correlation coefficients, to determine their contribution to the predictive model. Additionally, the variables selected by these two algorithms were discussed with experts from Colombiana de Trasplantes to validate their relevance and pertinence. To simplify the model and improve its interpretability, variables with lower relevance were removed. Subsequently, the Variance Inflation Factor (VIF) was implemented to identify multicollinearity; those with the highest VIF scores (>11) were excluded, ensuring that the resulting dataset was robust and capable of enhancing the accuracy of the prediction model (Fig 1).

Model training

The dataset was divided into two subsets: 80% was allocated for model training and 20% for validation. To estimate the model's performance on unseen data and reduce overfitting, a 10-fold cross-validation technique was implemented on the training subset, which internally used 90% of the data for training and 10% for testing in each fold (Fig 1). The remaining 20% of the data was reserved for final validation to assess the model's generalization capability.

For model training, the database was divided into 70% for training, 15% for validation, and 15% for testing. To estimate the model's performance on unseen data and avoid overfitting, a 5-fold cross-validation technique was implemented (Fig 1).

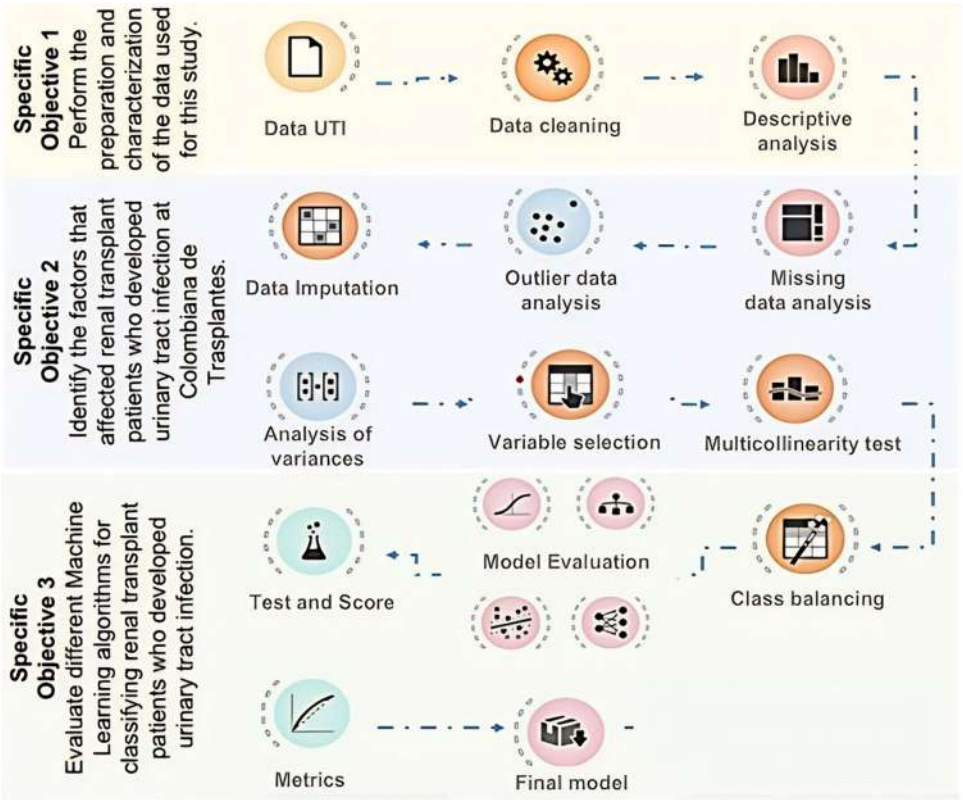

Fig. 1 Methodological framework illustration. The figure depicts the workflow designed to achieve the study's specific objectives. Objective 1 focuses on data preparation and characterization, including the extraction of data related to urinary tract infections (UTI), data cleaning, and descriptive analysis. Objective 2 aims to identify factors influencing renal transplant patients who developed UTI through processes such as data imputation, outlier analysis, missing data handling, analysis of variances, variable selection, and multicollinearity testing. Objective 3 involves evaluating machine learning algorithms for classifying renal transplant patients with UTI, incorporating class balancing, model evaluation, testing and scoring, and final model metric assessment

Finally, the performance of the trained models was evaluated using metrics from the confusion matrix, specificity, sensitivity, F1-score, precision, and receiver operating characteristic (ROC) curves. This was one of the main criteria for model selection, as it provides a more thorough evaluation of the model's ability to discriminate between cases and controls. By analyzing the relationship between the true positive rate (TPR) and the false positive rate (FPR), it facilitated the identification of which model best fits the data.

For each implemented model, the metrics were calculated using the test set. The results obtained among the different models were compared to determine which model performed best in detecting urinary tract infections in hospitalized kidney transplant patients.

Model evaluation

To assess the performance of the trained models, we used standard classification metrics derived from the confusion matrix, including accuracy, sensitivity, specificity, F1-score, and area under the receiver operating characteristic curve (AUC). ROC curves were employed to evaluate discriminative performance by comparing the true positive rate (TPR) to the false positive rate (FPR). In addition, calibration curves were generated to assess the agreement between predicted probabilities and observed outcomes, ensuring the reliability of the model's risk estimates. To further evaluate clinical applicability, decision curve analysis (DCA) was conducted to quantify the net benefit of each model across different probability thresholds. Together, these methods allowed for a comprehensive evaluation and informed selection of the most effective and clinically relevant model.

Ethical approval

The principles of the Helsinki Declaration and Colombian regulations established in Resolution 8430 of 1993, which sets the scientific, technical, and administrative standards for health research and the protection of clinical data, will be adhered to. Furthermore, the current proposal is framed within the first global report on artificial intelligence (AI) applied to health in 2021. It is classified as minimal-risk research since a retrospective documentary method is employed, and there are no modifications to the biological, physiological, psychological, or social variables of the study participants, thus informed consent is not required. The study has been approved by the ethics committee of Universidad El Bosque (CIE 2022-092).

Results

Exploratory analysis

Among patients who developed UTI, the mean age was 42.3 years and 60% were women. Most transplants were from deceased donors (63.7%). Comorbidity burden was substantial: 80.8% had hypertension and 17.6% had diabetes mellitus. The mean body mass index was 24.1 kg/m^2. Dysuria was the most frequently reported symptom (60.7%). Escherichia coli was the most common isolated pathogen (53.2%), and meropenem was the antibiotic most frequently administered (40.5%). Hospitalization was required in 7.9% of cases**.** (Supplementary file Table S1 and S2).

Preprocessing

The dataset was first split into training and validation sets, with 80% of the data allocated to training and 20% to validation. The number of records per class is 767 in the training dataset. In the validation dataset, each class has 192 records. A stratified split approach was used to ensure that the class proportions remained balanced in the training and validation sets. This method preserves the original class distribution in each subset, preventing biases in the model's learning process.

Before evaluating the models, data preprocessing was conducted. This included an analysis of missing data, revealing a 35% rate of missing values. The variables with the highest percentage of missing data were admission date, hospitalization, length of hospitalization, and acute rejection date. The latter three were considered Missing Not At Random (MNAR), meaning the absence of data is related to events or factors inherent to the variable. For example, in the case of acute rejection, if there is no information, it is because acute rejection did not occur (Fig. S1).

No outliers or duplicate data were found. Variables were encoded using LabelEncoder to facilitate the imputation of missing data and the evaluation of machine learning models.

Two imputation methods were tested: MICE and mean imputation, with the latter being ultimately selected. After imputing the data using two techniques (MICE and mean imputation), two tests were conducted to maintain the original variance of the data: one parametric (Student's t-test) yielded t-statistic: 0.0 with p-value: 1.0, and one non-parametric (Mann–Whitney U test) resulted in U-statistic: 169,164,759,460.5 with p-value: 1.0 for the mean imputation dataset. For the MICE imputed dataset, U-statistic: 153,794,462,224.5 with p-value: 1.0 was obtained. The higher U-statistic indicates that the distribution of the mean-imputed dataset is more similar to the original distribution, suggesting that mean imputation preserved the data structure better compared to MICE. For this reason, mean imputation was ultimately selected.

Variable selection

Once the database was inputted, selecting the most relevant variables for predicting UTI in kidney transplant patients was conducted using SelectKBest and Featurewise techniques (Fig. S1).

The variables selected by these two algorithms were discussed with experts from Colombiana de Trasplantes to validate their relevance and appropriateness. Subsequently, the Variance Inflation Factor (VIF) was implemented to identify multicollinearity by assessing the degree of redundancy or correlation among independent variables. Variables with the highest VIF values (greater than 11) were excluded, thereby improving the interpretability and reliability of the resulting model coefficients.

The variables that remained under these criteria were: body mass index, mycophenolate mofetil, mycophenolate sodium, prednisone, deflazacort, tacrolimus level, sex, hypertension, diabetes mellitus, re-transplantation, and donor type, ensuring that the resulting dataset was robust and capable of improving the accuracy of the prediction model.

Model evaluation

Among the fifteen machine learning algorithms evaluated, the Random Forest Classifier achieved the highest performance. During tenfold cross-validation, it yielded an average accuracy of 93.57%, an AUC of 94.43%, a recall of 93.57%, and a precision of 94.32% (Table 1). The low standard deviation of these metrics indicates high model consistency and minimal overfitting across folds.

Performance metrics per fold are detailed in Table 2, which shows consistently high values across all

Table 1 Performance metrics and standard deviations for cross-validation with test data

Metric	Mean	Standard deviation
Accuracy	0.9357	0.0310
AUC	0.9443	0.0338
Recall	0.9357	0.0310
Precision	0.9432	0.0254
F1-score	0.9353	0.0314
Specificity	0.9529	0.0424

Table 2 Cross-validation prediction with test data

Fold	Accuracy	AUC	Sensitivity (Recall)	Precision	F1-score	Specificity
0	0.9537	0.9563	0.9537	0.9576	0.9536	0.9589
1	0.9074	0.9451	0.9074	0.9219	0.9066	0.9828
2	0.9074	0.9078	0.9074	0.9219	0.9066	0.9082
3	0.9439	0.9464	0.9439	0.9464	0.9438	0.9489
4	0.9346	0.9285	0.9346	0.9421	0.9343	0.9224
5	0.8972	0.8955	0.8972	0.9146	0.8960	0.8938
6	0.8972	0.9043	0.8972	0.9086	0.8964	0.9114
7	0.9907	1.0000	0.9907	0.9908	0.9907	1.0093
8	0.9720	0.9862	0.9720	0.9735	0.9719	1.0004
9	0.9533	0.9733	0.9533	0.9547	0.9532	0.9933

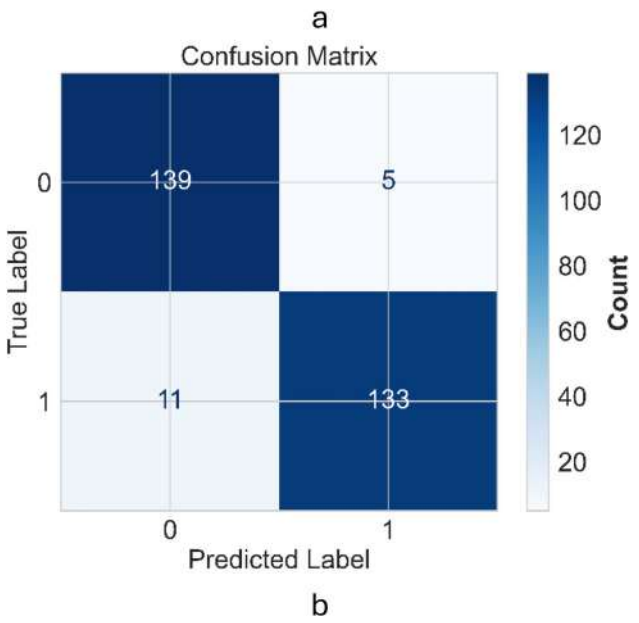

Fig. 2 Confusion matrices of the random forest and XGBoost model. **a** Random Forest Model: The confusion matrix indicates 142 true negatives, 128 true positives, 16 false negatives, and 2 false positive. **b** XGBoost Model: The confusion matrix shows 139 true negatives, 133 true positives, 11 false negatives, and 5 false positives

indicators. Confusion matrices for both the Random Forest and XGBoost models are presented in Fig. 2, allowing for a direct comparison of their classification performance on the external validation set. The Random Forest model achieved a higher number of true positives and true negatives, reflecting a more balanced classification capacity and fewer misclassifications compared to XGBoost.

The external validation set, representing 15% of the original dataset, was used to assess the generalization capability of the best-performing models. As shown in Table 3, Random Forest maintained robust predictive power with an AUC of 92.43%, accuracy of 91.67%, sensitivity of 93.10%, and specificity of 94.05%. In comparison, XGBoost performed slightly lower across all metrics, though still within acceptable clinical ranges.

Table 3 Performance metrics of random forest and XGBoost models on the external validation set

Metric	Random forest	XGBoost
Accuracy	91.67%	90.48%
AUC	92.43%	91.65%
Sensitivity	93.10%	91.38%
Precision	92.50%	89.74%
F1-score	92.80%	90.56%
Specificity	94.05%	92.86%

Comparison of accuracy, AUC, sensitivity, precision, F1-score, and specificity for both models evaluated on unseen data

Figure 3 presents the ROC curves for both Random Forest and XGBoost models in the external validation set. The Random Forest model consistently outperformed XGBoost in terms of area under the curve. To evaluate model calibration, Fig. 4 displays calibration plots comparing predicted and observed probabilities. The Random Forest model demonstrated better alignment, indicating more reliable risk estimates. Additionally, clinical utility was explored using decision curve analysis (DCA), shown in Fig. 5. Across a range of decision thresholds, the Random Forest model offered greater net benefit than XGBoost, reinforcing its practical value for identifying patients at risk of UTI.

Lastly, the relative importance of predictors is presented in Fig. 6, which shows the top contributing variables to UTI prediction. These include obesity, use of mycophenolate mofetil or sodium, donor type, hypertension, diabetes mellitus, prednisone or deflazacort use, re-transplantation, tacrolimus level, and recipient sex.

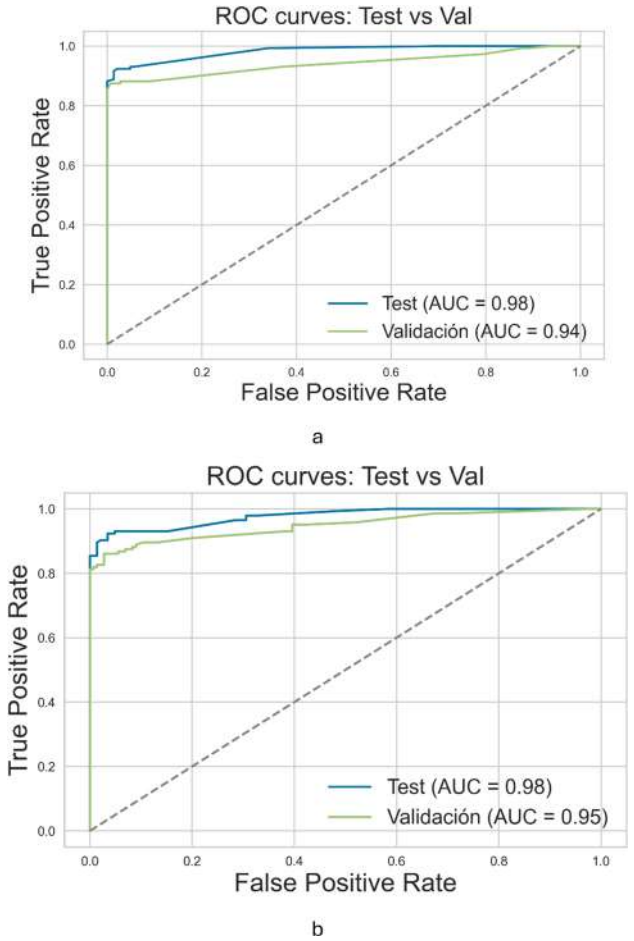

Fig. 3 ROC curves of the random forest and XGBoost model. **a** Random Forest Model: The ROC curve demonstrates an AUC of 94.43%. **b** XGBoost Model: The ROC curve demonstrates an AUC of 91.65%

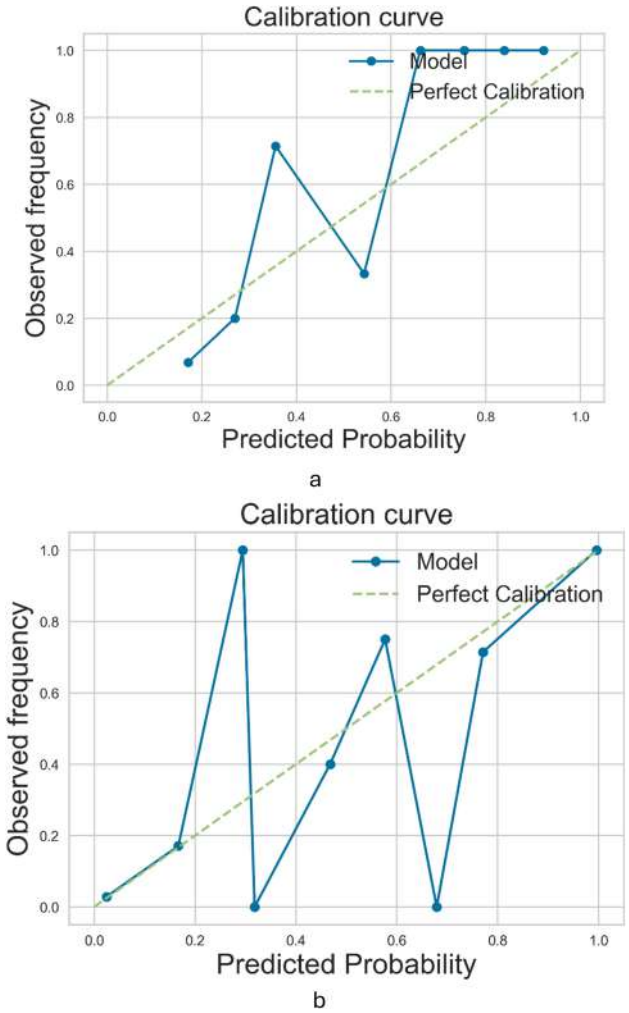

Fig. 4 Calibration curves of the random forest and XGBoost model. **a** Random Forest Model: The calibration curve closely follows the ideal diagonal, suggesting strong agreement between predicted and observed probabilities. **b** XGBoost Model: The curve shows slight deviation from ideal calibration, particularly at higher predicted probabilities

Discussion

The first comparison of machine learning models to predict urinary tract infection (UTI) in kidney transplant patients within a Latin American population was conducted. The results obtained in this study open new opportunities for the integration of machine learning techniques into the clinical management of transplant patients. Both models demonstrated strong performance in predicting urinary tract infections among kidney transplant recipients. While the Random Forest model slightly outperformed XGBoost in terms of discrimination, calibration, and overall predictive accuracy, the results from both approaches indicate robust generalizability and clinical applicability. These findings suggest that either model could be effectively implemented in similar clinical contexts to support early identification and management of UTIs. Additionally, the most significant variables related to the occurrence of UTI in transplant patients were identified. In order of importance: body mass index (BMI), use of mycophenolate mofetil or mycophenolate sodium, donor type, hypertension, diabetes mellitus, use of prednisolone or deflazacort, requirement for re-transplantation, serum tacrolimus levels, and recipient sex.

Regarding BMI, we found that obesity significantly increases the risk of developing UTI in kidney transplant patients. These results are consistent with those recently described by Sobral Antonelli et al. [8], reporting that transplant patients with increased abdominal circumference, visceral fat area, and total fat mass experienced more urinary infections ($p=0.014$, $p=0.020$, and $p=0.018$, respectively). Key mechanisms in the immunological alteration associated with obesity include hormonal changes, increased concentrations of insulin and leptin, which have direct consequences on immunometabolism and inflammation, a

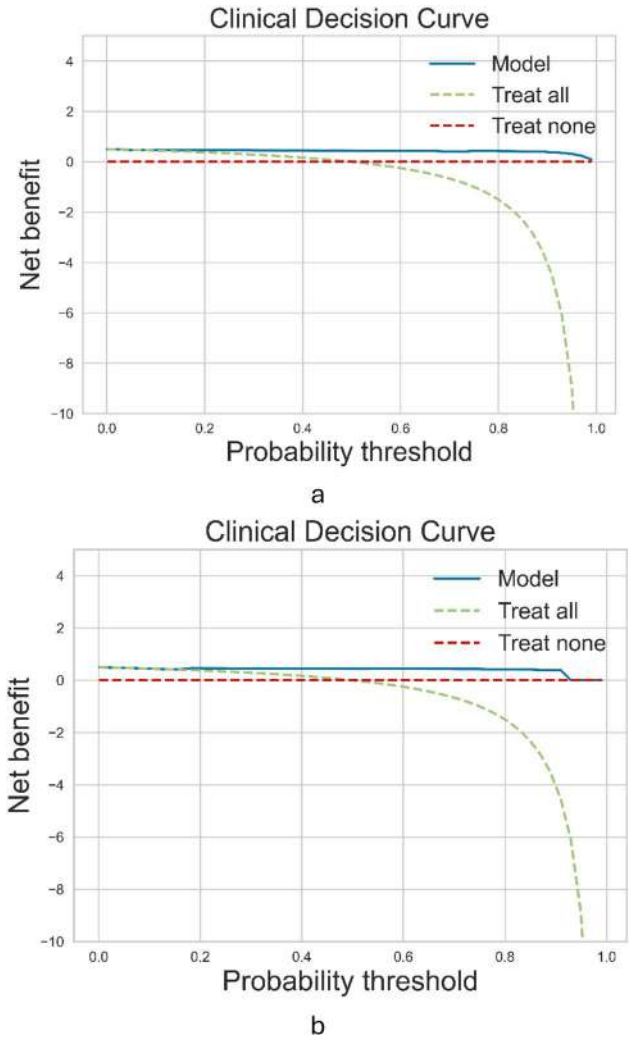

Fig. 5 Decision Curve Analysis (DCA) of the final models. **a** Random Forest Model: Displays a consistently higher net benefit across a wide range of threshold probabilities. **b** XGBoost Model: Shows good clinical utility, though slightly lower net benefit compared to Random Forest

positive regulatory effect on glucose absorption and oxidative metabolism, mainly in T lymphocytes, and an increase in the number of inflammatory cells (CD8 + lymphocytes, CD4Th1 and CD4Th17 lymphocytes, and M1 macrophages) [9]. Additionally, modifications in the composition and quantity of extracellular vesicles derived from adipocytes regulate immune activation and systemic metabolism, along with alterations in the biosynthesis of oxylipins formed from polyunsaturated fatty acids, which are crucial for controlling inflammation [9].

Regarding the use of immunosuppressants (including corticosteroids) may be explained by the risk of infection by suppressing key components of the immune system and weakening natural barriers in the urinary tract [10, 11]. In our study, the use of mycophenolate mofetil was more important than mycophenolate sodium for predicting UTI. Regarding the difference between mycophenolate mofetil and mycophenolate sodium, mycophenolate sodium was initially designed to reduce the frequency of gastrointestinal symptoms, but not to decrease the risk of UTI [12]. In any case, the use of immunosuppressants remains a fundamental risk factor, as it is modifiable and allows for risk reduction through the implementation of appropriate preventive strategies [13].

The type of donor is also relevant [14]. It is important to consider that living donors are evaluated beforehand for bacteriuria prior to the intervention, and it is likely that the recipient of a living donor requires less immunosuppression. More research is needed in this case to reach a conclusive justification. Some studies have established that the presence of hypertension or diabetes mellitus increases the risk of UTI [14]. Diabetes generates multiple alterations in the immune system response, a previous study suggests presence of dysfunctional adipose tissue, liver dysfunction, and skeletal muscle dysfunction collectively contribute to the increased circulatory levels of proinflammatory factors [13]. Therefore, strict control of diabetes would possibly decrease the risk of presenting UTI. Regarding hypertension, we found it to be an important factor in presenting UTI, some authors propose that these patients have alterations in their intestinal microbiota profiles [15].

When comparing our results with previous studies, multiple studies demonstrate the significant role of female sex in predicting UTI in kidney transplant patients [14]. An explanation for the higher frequency in women is the shorter urethra length and changes in the urethral and vaginal microbiota associated with the use of immunosuppressants [16, 17]. Our study is consistent with most of the existing literature. Therefore, one of the most relevant variables in the clinical context of our study is re-transplantation. Bodro et al. has shown that surgical manipulation of the urinary tract and UTI caused by MDR organisms have a higher risk of UTI [18]. Thus, our findings are consistent with previous evidence regarding the risks associated with re-transplantation as a risk factor for UTI.

Our study aligns with previous research on machine learning for UTI prediction but presents distinct methodological strengths. Taylor et al. and Burton et al. used XGBoost on large retrospective cohorts, achieving high specificity (94.9% and 60.93%, respectively) but lower sensitivity (61.7% and 95.2%) compared to our models. Ozkan et al. tested multiple models, reporting higher accuracy for ANN (98.3%) but with a small sample size (n = 59), which limits generalizability. Heckerling et al. combined ANN with genetic algorithms, obtaining lower sensitivity (82.1%) and specificity (74.4%), relying primarily on symptom-based predictors rather than transplant-specific risk factors [6, 7]. These differences highlight the impact

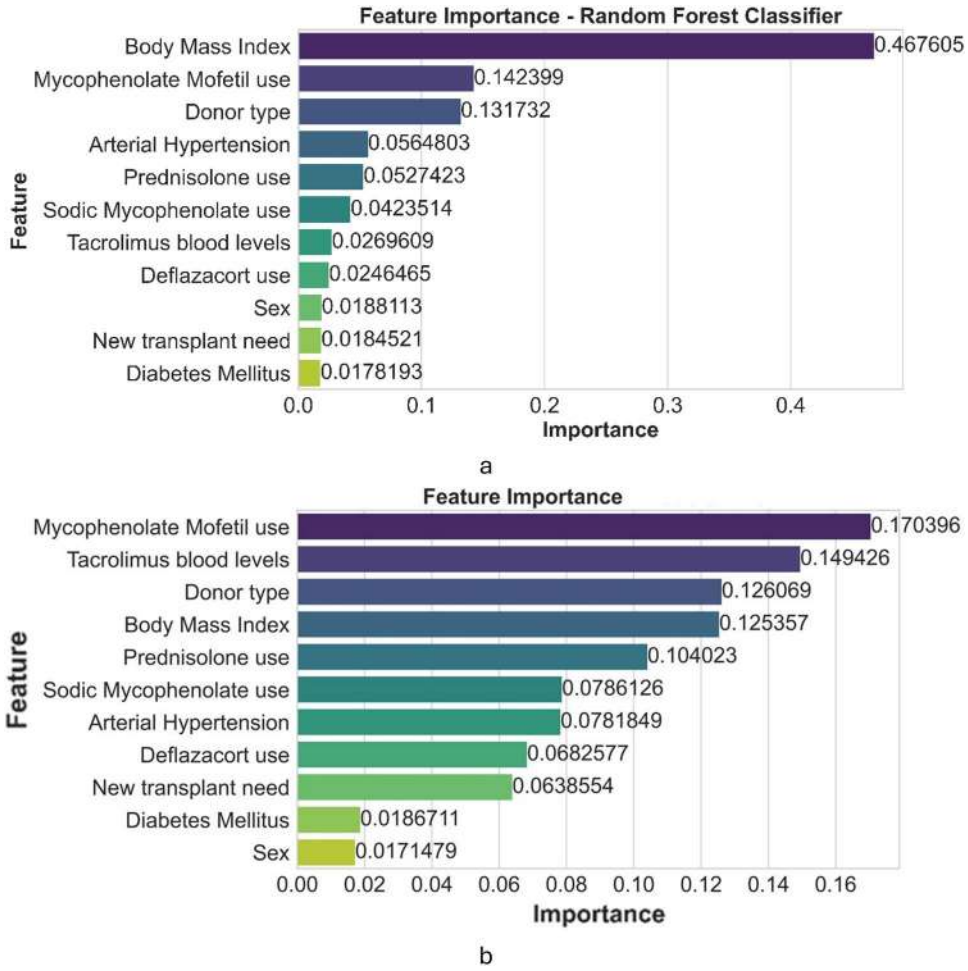

Fig. 6 Feature importance of the random forest and XGBoost model. **a** Random Forest Model: The most relevant variables include body mass index, mycophenolate mofetil or sodium, donor type, hypertension, diabetes mellitus, prednisone or deflazacort use, re-transplantation, serum tacrolimus level, and sex. **b** XGBoost Model: The most influential predictors are similar, although with slight variations in the relative importance assigned to each variable

of study population, dataset size, and predictor selection on model performance.

To ensure the effective integration of our predictive models into clinical practice, it is essential that its outputs are both interpretable and actionable for healthcare providers. This study highlights key predictive variables, such as BMI, immunosuppressant use, and donor type, which are well-established clinical factors and can be seamlessly incorporated into existing risk assessment workflows. Additionally, the model's predictions could be integrated into electronic health records (EHR) to enable real-time risk stratification, supporting early interventions such as closer monitoring or prophylactic antibiotic strategies for high-risk patients. At our center, we have implemented visualization tools to present predictive insights in a clear and structured manner, and these have proven to be valuable for clinical decision-making. We consider the expansion of such tools to be a crucial step toward enhancing the practical application of machine learning models in patient care.

However, it is important to consider the limitations associated with this study. The main limitation is the use of retrospective data, which can introduce biases related to the quality and integrity of clinical records. Furthermore, although the models demonstrated robust performance with the data used, it is necessary to validate it in different populations and clinical settings to ensure its generalizability and effectiveness. The study's strengths lie in its high predictive performance. It employs rigorous methodologies, including robust feature selection and data balancing, ensuring reliable and clinically relevant results. Additionally, expert validation of key predictors highlights the model's potential for improving patient outcomes and integrating machine learning into clinical practice. Finally, the implementation of these models in clinical practice requires interdisciplinary collaboration among physicians, data engineers, and health administrators. It is crucial to develop training strategies and usage protocols that facilitate the adoption of this technology, ensuring that the benefits of machine learning translate into tangible improvements in care quality and patient outcomes.

Conclusion

This study demonstrates the potential of machine learning in improving early detection of UTIs in kidney transplant recipients through a strict and systematic methodology. By applying ML models to a well-defined set of clinical and laboratory variables, we developed a structured approach for risk assessment that could support timely interventions and personalized patient care. While the results are promising, external validation in diverse populations is needed to confirm the model's applicability. Future efforts should focus on real-time integration into electronic health records and the development of decision-support tools to facilitate clinical adoption, ultimately improving post-transplant management.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11255-026-05308-9.

Author contributions Conceptualization: LVCS, CPM, ADL, ACM, ARC, SJP, CRS, AGL,FGL—Study Design: LVCS, CPM, ADL, ACM, ARC, SJP, CRS—Data Acquisition: JAGM, AGL, ACS, FGL—Data Analysis: LVCS, CPM, ADL, ACM, ARC, SJP, CRS, JAGM, AGL, ACS, FGL—Results Interpretation: LVCS, CPM, ADL, ACM, ARC, SJP, CRS, AGM, AGL, ACS, FGL—Manuscript Writing: LVCS, AGM, AGL, ACS, FGL—Critical Review of the Manuscript: LVCS, CRS, JAGM, AGL, ACS, FGL—Final Approval of the Manuscript: LVCS, CPM, ADL, ACM, ARC, SJP, CRS, AGM, AGL, ACS, FGL—Supervision: CPM, ADL, ACM, ARC, SJP, CRS, ACS, FGL.

Funding This study was funded by Colombiana de Trasplantes.

Data availability The research data supporting the results of this manuscript are derived from a non-public patient database. Due to ethical and private concerns, the data cannot be made publicly available. Data access is subject to approval and requires a signed data use agreement. However, researchers may request access to the data by contacting the corresponding author at [aegarcia@colombianadetrasplantes.com.] (mailto:aegarcia@colombianadetrasplantes.com).

Declarations

Conflict of interest The authors declare no financial or non-financial competing interests directly or indirectly related to the work submitted for publication.

Ethical approval and consent to participate The study adhered to the Declaration of Helsinki and Colombian Resolution 8430 of 1993. It was classified as risk-free research, not requiring informed consent, and was approved by the ethics committee of Universidad El Bosque (CIE 2022–092).

Consent for publication Not applicable.

References

1. Rosselli D, Rueda JD, Diaz C (2015). Cost-effectiveness of kidney transplantation compared with chronic dialysis in end-stage renal disease. Saudi Journal of Kidney Diseases and Transplantation [Internet]. 26(4): 733. https://doi.org/10.4103/1319-2442.160175. Available from: http://www.sjkdt.org/text.asp?2015/26/4/733/160175
2. Agrawal A, Ison MG, Danziger-Isakov L (2022) Long-term infectious complications of kidney transplantation. Clin J Am Soc Nephrol 17(2):286–295. https://doi.org/10.2215/CJN.15971020
3. Lorenz EC, Cosio FG (2010) The impact of urinary tract infections in renal transplant recipients. Kidney Int 78(8):719–721. https://doi.org/10.1038/ki.2010.219
4. Camargo-Salamanca A, Garcia-Lopez A, Patino-Jaramillo N, Giron-Luque F (2020) Acute kidney injury in hospitalized kidney transplant recipients. Transplant Proc. https://doi.org/10.1016/j.transproceed.2019.12.046
5. Naik AS, Dharnidharka VR, Schnitzler MA, Brennan DC, Segev DL, Axelrod D et al (2016) Clinical and economic consequences of first-year urinary tract infections, sepsis, and pneumonia in contemporary kidney transplantation practice. Transplant Int 29(2):241–252. https://doi.org/10.1111/tri.12711
6. Goździkiewicz N, Zwolińska D, Polak-Jonkisz D (2022) The use of artificial intelligence algorithms in the diagnosis of urinary tract infections—a literature review. J Clin Med. https://doi.org/10.3390/jcm11102734
7. Ozkan IA, Koklu M, Sert IU (2018) Diagnosis of urinary tract infection based on artificial intelligence methods. Comput Methods Programs Biomed 166:51–59. https://doi.org/10.1016/j.cmpb.2018.10.007
8. Antonelli TS, Mantovani MDS, de Carvalho NC, Archangelo TE, Minicucci MF, Ferreira Filho SP et al (2024) Body fat predicts urinary tract infection in kidney transplant recipients: a prospective cohort study. J Nephrol. https://doi.org/10.1007/s40620-024-01922-x
9. Shaikh SR, Beck MA, Alwarawrah Y, MacIver NJ (2024) Emerging mechanisms of obesity-associated immune dysfunction. Nat Rev Endocrinol 20(3):136–148. https://doi.org/10.1038/s41574-023-00932-2
10. Klein RD, Hultgren SJ (2020) Urinary tract infections: microbial pathogenesis, host–pathogen interactions and new treatment strategies. Nat Rev Microbiol 18(4):211–226. https://doi.org/10.1038/s41579-020-0324-0
11. Perez-Carrasco V, Soriano-Lerma A, Soriano M, Gutiérrez-Fernández J, Garcia-Salcedo JA (2021) Urinary microbiome: yin and yang of the urinary tract. Front Cell Infect Microbiol. https://doi.org/10.3389/fcimb.2021.617002
12. Zolezzi M (2005) Mycophenolate sodium versus mycophenolate mofetil: a review of their comparative features. Transplantation 16(2):140–145
13. Costa LM, Galato D (2024). Identification of factors associated with medication adherence in renal transplant patients: an integrative literature review. Brazilian Journal of Transplantation [Internet]. [cited 2024 Nov 24] 26(1): e0123 https://doi.org/10.53855/BJT.V26I1.484_ENG. Available from: https://www.scielo.br/j/bjt/a/35Q8zC75DYdFSNqBM7NDVkg/?lang=en
14. Hosseinpour M, Pezeshgi A, Mahdiabadi MZ, Sabzghabaei F, Hajishah H, Mahdavynia S (2023) Prevalence and risk factors of urinary tract infection in kidney recipients: a meta-analysis study. BMC Nephrol 24(1):1–13. https://doi.org/10.1186/s12882-023-03338-4
15. Liu F, Ling Z, Xiao Y, Yang Q, Wang B, Zheng L et al (2017) Alterations of urinary microbiota in type 2 diabetes mellitus with hypertension and/or hyperlipidemia. Front Physiol 8:238491. https://doi.org/10.3389/FPHYS.2017.00126/BIBTEX
16. Rotjanapan P, Jaroensukrungruang A, Pisitkun P, Ngamjanyaporn P, Manonai J, Sawaswong V, et al (2021). Vaginal microbiota affects urinary tract infection risk in women with systemic lupus erythematosus: a pilot cross-sectional study from Thailand. Lupus Sci Med [Internet]. [cited 2024 Nov 24] 8(1). https://doi.org/10.

1136/LUPUS-2021-000551. Available from: https://lupus.bmj.com/content/8/1/e000551
17. Deltourbe L, Lacerda Mariano L, Hreha TN, Hunstad DA, Ingersoll MA (2022) The impact of biological sex on diseases of the urinary tract. Mucosal Immunol 15(5):857–866. https://doi.org/10.1038/S41385-022-00549-0
18. Bodro M, Sanclemente G, Lipperheide I, Allali M, Marco F, Bosch J et al (2015) Impact of antibiotic resistance on the development of recurrent and relapsing symptomatic urinary tract infection in kidney recipients. Am J Transplant 15(4):1021–1027. https://doi.org/10.1111/ajt.13075