

Machine Learning Models for the Early Detection of Urinary Infections in Post-Kidney Transplantation Patients

Liceth Castañeda¹, Carlos Puentes^{2,3}, Alejandro Duitama^{2,3}, Andrés Cardona^{2,4}, Consuelo Romero^{2,4}, Alejandro Ramos^{2,4}, Sandra Perdomo^{2,4}, Alejandro Camargo⁵, Andrea Garcia⁵, Julia Gomez⁵, Fernando Girón⁵.

1. Master in Applied Statistics and Data Science, Bogotá, Colombia.
2. Artificial Intelligence Laboratory, Universidad El Bosque, Bogotá, Colombia.
3. SIGNOS Group, Bogotá, Colombia.
4. Cellular and Molecular Immunology Group, Universidad El Bosque, Bogotá, Colombia.
5. Colombian Transplant Research Group, Bogotá, Colombia.

Introduction:

Urinary tract infections (UTIs) are among the most common bacterial infections, posing significant health burdens globally.

Early and accurate prediction of UTIs can significantly enhance patient outcomes, reduce healthcare costs, and minimize the risk of complications, including antibiotic resistance.

This study presents a comparison of machine learning models to predict UTIs in patients with kidney transplantation within a Colombian population.

Methods:

Retrospective cohort of 911 kidney transplant recipients

29340 records were classified into two groups UTI Yes/No

Specific data imputation techniques were applied

Under-sampling was used to address data imbalance

Variable selection was performed using featurewiz technique

Comparisons between 15 models were performed

using confusion matrix, sensitivity, specificity, and AUC

All models underwent cross-validation on normalized data

- Highest performance models were the Random Forest Classifier (AUC 95.6%) and the Extra Trees Classifier (AUC 94.2%).
- The Random Forest Classifier model showed the best metrics, confirmed by the confusion matrix, where sensitivity and specificity for test data are 88% and 89% respectively.

Influential variables in this prediction included

- Age
- induction therapy
- follow-up
- body mass index
- donor type
- Mycophenolate Mofetil
- Diabetes
- Tacrolimus
- Mycophenolate Sodium
- Prednisolone
- Tacrolimus levels
- Everolimus
- Deflazacort

Results:

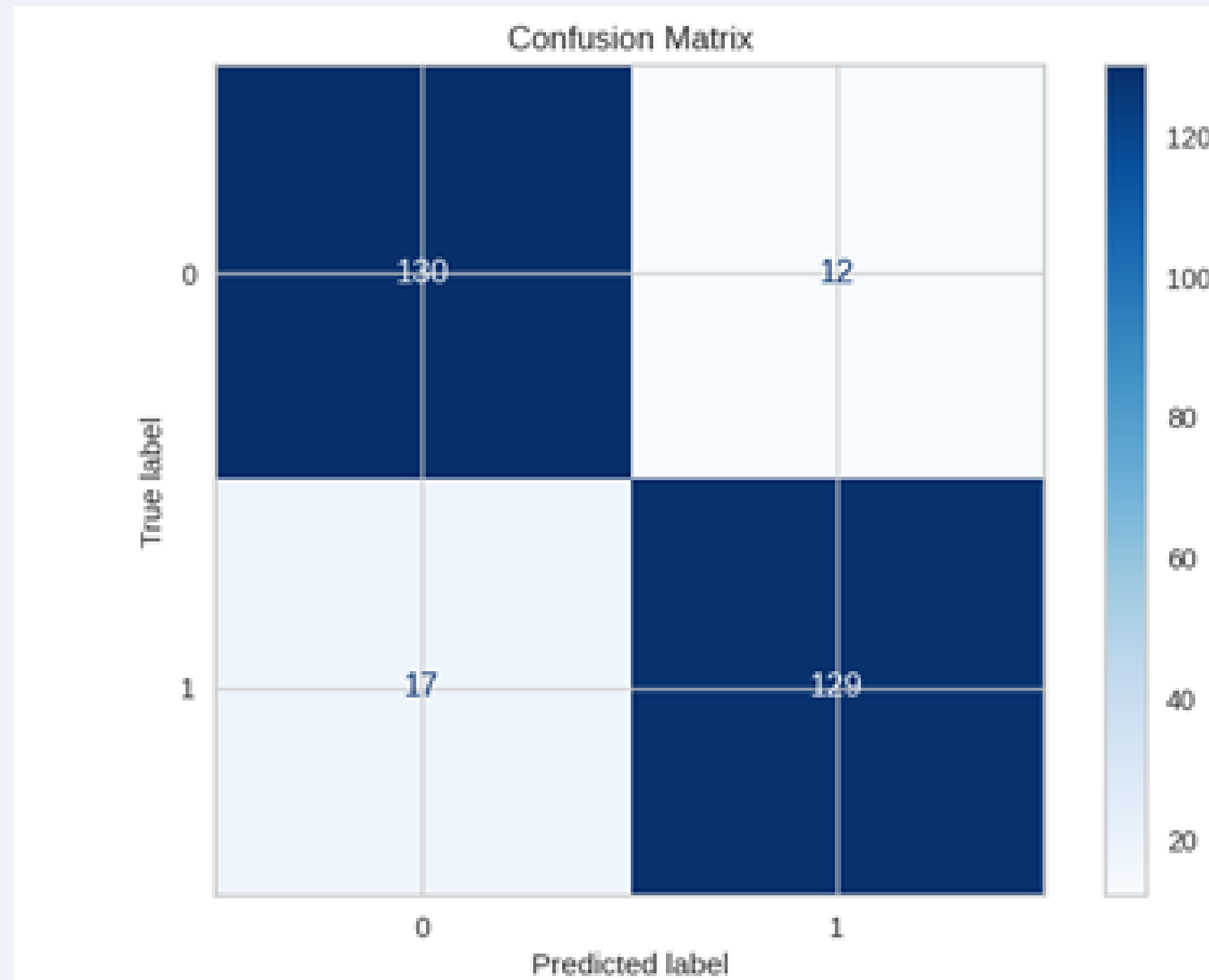


Figure 1. Ada Boost model confusion matrix.

Metric	Value
Accuracy	0.8993
Precision	0.9149
Recall	0.8836
F1-score	0.899

Conclusions:

- The model excels in predicting UTIs, proving its potential as a valuable clinical tool.
- It demonstrates strong accuracy, enabling early identification and intervention.
- This could improve patient outcomes and reduce the burden of UTIs.
- Further validation across diverse populations and clinical settings is essential to ensure its effectiveness and applicability.

