Machine Learning Analysis of Automated Fluorescence Flow Cytometry Data for Fast and Efficient Microbiological Analysis of Urine Samples

Tim Graf

Supervisor(s): PD Dr. Alban Ramette
Institution(s): University of Bern, Institute for Infectious Diseases (IFIK)

Introduction

Urinary tract infections (UTIs) are among the most common infections, and urine microbiological cultures are the gold standard for diagnosis. As culturing is time-consuming, it cannot guide initial therapy, leading to unnecessary antibiotic treatments. Urine flow cytometry (UFC) can analyze urine samples within minutes and can be used as a screening tool to rule out negative cultures. Recent UFC analyzers have shown some ability to predict the Gram type of the bacteria in urine, which can help guide initial antibiotic treatment, if the prediction is reliable.

Materials and Methods

Retrospective data from microbiological analysis and UFC analysis of 3,148 urine samples (1,775 female, 1,373 male) sent to the microbiology lab at IFIK for routine diagnosis of urinary tract infection by urine culture was used. The primary aim was to develop thresholds for bacteria and leukocyte counts from UFC analysis to rule out negative urine cultures, with examination of different thresholds depending on age and sex. Additionally, extreme gradient boosting (XGBoost) models were used to incorporate more UF-5000 measurements into the decision process in order to better predict negative urine cultures.
In a secondary aim, evaluation of the Sysmex UF-5000 Gram typing performance was conducted, and machine learning algorithms were trained that aim to improve Gram typing performance from the bacteria scattergram produced by the UF-5000.

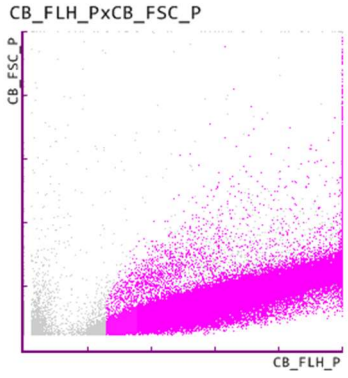

Fig. 1 Example bacterial scattergram produced by the SYSMEX UF-5000

Outcome measures included the area under the receiver operating characteristic curve (AUC), sensitivity (SE), specificity (SP), negative predictive value (NPV), and positive predictive value (PPV).

Results

Regarding the primary aim, thresholding of leukocyte and bacterial counts to rule out negative cultures yielded an overall AUC of 0.755 with SE 91.0% and SP 44.2%. Specific thresholds for women showed AUC of 0.702 with SE 92.5% and SP 31.5%, while the threshold for men yielded AUC of 0.823 with SE 88.1% and SP 55.6%. For patients older than 55 years, AUC of 0.735 with SE 91.9% and SP 37.3% was achieved, while thresholds for patients of age 55 or younger yielded AUC of 0.749 with SE 91.7% and SP 49.3%. Predictions with XGBoost yielded AUC of 0.787 with SE 86.1% and SP 56.8%.
Concerning the secondary aim, the UF-5000 detected Gram negative bacteria, Gram positive bacteria and Mixed Gram respectively with SE 72.7%, 50.0%, 45.5% and SP 84.0%, 85.1%, 94.6%. The machine learning approach using the data from the scattergram yielded SE 81.8%, 80.6%, 45.5% and SP 78.7%, 74.7%, 92.3% for Gram negative, Gram positive and Mixed Gram bacteria, respectively.

Discussion

These results are in line with the findings of similar studies performed with the Sysmex UF-5000. Bacterial counts from the UF-5000 have a higher predictive power on the growth result of urine culture than leukocyte counts, and thresholds for women are generally higher. Gram type prediction from both the UF-5000-interpreted data directly and the raw scattergrams yielded better results for Gram negative bacteria than Gram positive bacteria. Prediction of cultures with mixed Gram type was shown to be most difficult. While the Gram type predictions might help guide initial treatment, they are not accurate enough to replace urine cultures.

Acknowledgements

The contribution of Dr. Benjamin Schoenbeck, Ulrike Pauliks and Cara-Dorothea Schneider from SYSMEX for helping with the export of the data and answering technical questions regarding the UF-5000 is gratefully appreciated, as is the contribution from Dr. Peter Keller for the fruitful discussions and ideas.