

Promote utilization of the Vivli AMR register to more researchers and drive advances in AMR field: Use of AI to model WHO priority pathogen and antibiotics from Africa using ATLAS dataset

Drileonzi Caesar^{1*}, Asianzu Chris Benjamin¹, Yokoko Christine¹, Asiku Nasibu¹, Luyima Denis¹

¹Muni Labs, Arua, Uganda

*Team Leader

ABSTRACT

This study explores the application of machine learning techniques to predict antibiotic resistance in *Staphylococcus aureus* within African settings, where comprehensive data collection and routine testing are often limited. Due to constraints in data size and availability, the research focuses on the resistance of *S. aureus* to three critical antibiotics. We developed predictive models using three machine learning methods, XGBoost, Random Forest, and Support Vector Machines (SVM). Among these, the SVM model exhibited superior performance, achieving accuracy rates between 0.9975 and 0.9911, surpassing both Random Forest (0.7713 and 0.9212), XGBoost (0.318 and 0.3405). To support healthcare professionals and researchers, we also created an interactive R Shiny application that leverages the SVM models to predict antibiotic resistance based on key variables such as gender, age group, and sample source. The app, available online, serves as a valuable tool for predicting resistance patterns in resource-limited settings. The study's findings are particularly relevant to Africa, where the lack of infrastructure for routine antibiotic resistance testing presents significant challenges. By providing accurate predictive models, this research offers crucial insights for healthcare providers, enabling more informed clinical decisions and improved patient outcomes. Additionally, the study underscores the importance of investing in laboratory infrastructure and standardized data collection practices to fully harness the potential of AI in combating antimicrobial resistance (AMR). While the research highlights the benefits of AI-driven models, it also acknowledges limitations related to data availability and sample size. The integration of these advanced tools into public health policy has the potential to significantly enhance AMR monitoring and management across the continent.

INTRODUCTION

Our research focused on leveraging machine learning techniques, including XGBoost, Random Forest, and Support Vector Machine, to uncover insights from the ATLAS (1) dataset from Africa. The aim was to enhance public health practices and improve the health system on the continent. By analyzing the antimicrobial resistance (AMR) dataset using these machine learning methods, we sought to identify correlations, patterns, and trends that could inform targeted interventions. We wanted to develop AI models using the Vivli ATLAS dataset, but we could not validate them with data from St. Joseph's Hospital Maracha in Uganda because of lack of data. The research prioritized pathogens identified by the WHO's Global

Antimicrobial Resistance Surveillance System (GLASS) (2), including *Escherichia coli*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Staphylococcus aureus*, *Streptococcus pneumoniae*, and *Salmonella* spp. The antibiotics of focus were ceftriaxone, metronidazole, ciprofloxacin, and amoxicillin. However, we were not able to work with all the bacteria and antibiotics due to small sample sizes in ATLAS dataset. Therefore, we changed our focus to modelling *Staphylococcus aureus* to Erythromycin, Levofloxacin, and Clindamycin. *S. aureus* has been on WHO critical pathogen lists for 2017 and 2024 (3), these three antibiotics are used extensively in Africa, and they are becoming resistant to *S. aureus*.

OBJECTIVES

Objective 1: Develop AI Models for predicting the resistance patterns of *S. aureus* to Erythromycin, Levofloxacin, and Clindamycin antibiotics.

Objective 2: Develop interactive Shiny Applications (dashboards) for predicting the resistance of *S. aureus* to three antibiotics, Erythromycin, Levofloxacin, and Clindamycin.

METHODS

Predicting Antibiotic Resistance in *Staphylococcus aureus* Using Machine Learning Models

The goal of this analysis is to predict antibiotic resistance in *S. aureus*, particularly focusing on resistance to Levofloxacin, Erythromycin, and Clindamycin. The dataset includes various predictor variables such, gender, age group, specimen sources, and inpatient/outpatient status. Three machine learning models—XGBoost, Random Forest, and Support Vector Machine (SVM)—were employed to achieve this objective. Each model's performance was evaluated using test dataset and metrics such as Accuracy, Sensitivity, Specificity. The cleaning and prediction was done using R version 4.4.1 (4)

Data Cleaning and Imputation

- The dataset 'Atlas.csv' was subsetted to focus on specific African countries.
- Missing values in categorical columns ('Erythromycin_I', 'Clindamycin_I', 'Levofloxacin_I', and 'In...Out.Patient') were imputed using the mode (most frequent value) for each column.
- Additional imputation was applied to the 'Gender' column, also using the mode.

Data Preparation

- The cleaned dataset was filtered to include only samples identified as *S. aureus*.
- Separate datasets were created for each antibiotic: 'Erythromycin', 'Clindamycin', and 'Levofloxacin'.
- Categorical variables were converted to factors and one-hot encoded for modeling.

Modeling with XGBoost, Random Forest, and Support Vector Machine (SVM)

- For each antibiotic, the dataset was split into training (80%) and test (20%) sets.

- The XGBoost model was configured for multi-class classification, and the training data was used to fit the model.
- Predictions were made on the test set, and a confusion matrix was generated to evaluate model performance.
- The Antibiotic modeling was repeated with Random Forest, and Support Vector Machine (SVM)

Model Evaluation

- The confusion matrix was visualized to better understand the model's accuracy in predicting antibiotic resistance.
- This process was repeated separately for each of the three antibiotics.

Development of Shiny Dashboard

- We developed a Shiny Dashboard using SVM models. The app, datasets, and the R scripts are uploaded to GitHub at <https://github.com/Munilabs2024/AMR-Data-Challenge>

RESULT

In our research, we initially aimed to model a broad range of pathogens and antibiotics across Africa. However, due to the constraints posed by the size and availability of the dataset, we focused specifically on *S. aureus* and its resistance to three critical antibiotics. We employed three machine learning techniques, XGBoost, Random Forest, SVM to develop predictive models for antibiotic resistance. Among these methods, the SVM model outperformed both Random Forest and XGBoost, achieving accuracy rates between 0.9975 and 0.9911 (as shown in Tables 1 to 3). Notably, the SVM model demonstrated superior performance in predicting resistance patterns for the antibiotic compared to the others, using a set of carefully selected variables. This research holds particular significance for regions like Africa, where infrastructure for routine antibiotic resistance testing is often inadequate, and comprehensive data collection is frequently lacking. By leveraging machine learning techniques, we can provide valuable insights and predictive capabilities in settings where traditional laboratory resources are scarce.

In addition to the predictive models, we developed an interactive R Shiny application designed to assist healthcare professionals and researchers in predicting antibiotic resistance patterns. The app utilizes key variables such as gender, age group, and the source of the samples to generate resistance predictions using SVM models (Figure 1). Figure 1A illustrates the instruction page within the app, which provides step by step guidance on how to use the tool effectively. Figure 1B showcases a sample prediction, where users can select the variables of interest to visualize the predicted resistance outcomes. The R Shiny app is available online, offering an accessible platform for users to explore and predict resistance patterns. The link to the Shiny Dashboard is: <http://167.71.249.72:3838/Muni/>

IMPACT

OF

WORK

African researchers now have access to SVM AI-driven models specifically designed to predict

the resistance of three key antibiotics against *S. aureus*. These advanced models are a significant tool for healthcare providers across Africa, enabling them to make more informed clinical decisions. By accurately predicting antibiotic resistance patterns, clinicians can choose the most effective treatments, leading to improved patient outcomes, reduced treatment failures, and the conservation of vital healthcare resources. For policymakers, this study provides critical insights into the application of AI in combating antimicrobial resistance (AMR). The ability to predict resistance trends allows for the development of targeted strategies tailored to the specific needs of different countries. This could include revising treatment guidelines to better reflect emerging resistance patterns, launching public awareness campaigns to promote responsible antibiotic use, and implementing robust surveillance programs.

The research highlights the crucial role of investment in laboratory infrastructure to enhance AMR monitoring capabilities. By strengthening laboratory networks, African nations can improve the accuracy and timeliness of resistance data, which is essential for effective decision-making. Moreover, the integration of AI-driven models and Shiny App dashboards into public health policy has the potential to transform Africa's health systems. These tools not only enhance the continent's ability to respond to AMR but also position Africa as a leader in the global fight against this growing threat. The adoption of such innovative technologies demonstrates Africa's commitment to leveraging cutting-edge solutions to address complex public health challenges, ultimately contributing to the global effort to curb the spread of antimicrobial resistance.

CONCLUSION AND LIMITATIONS

The study demonstrates the potential of AI in improving AMR testing and antibiotic prescribing in Africa. While AI can enhance patient outcomes and support policy decisions, the research is limited by the availability of comprehensive AMR data, small sample sizes, and inconsistent data collection practices across Africa. The study calls for standardized data collection, expanded lab capabilities, and increased investment in healthcare infrastructure to fully realize AI's benefits in combating AMR on the continent.

TABLE OF FIGURES

Table 1. Performance of AI Methods in Predicting Erythromycin Susceptibility of *S. aureus*

AI model	Sensitivity (Resistant Class)	Specificity (Resistant Class)	Accuracy
XGBBOOST	0.29091	0.64952	0.324
Random Forest	0.18987	0.95707	0.7713
Support Vector Machine	0.9758	1.0000	0.9911

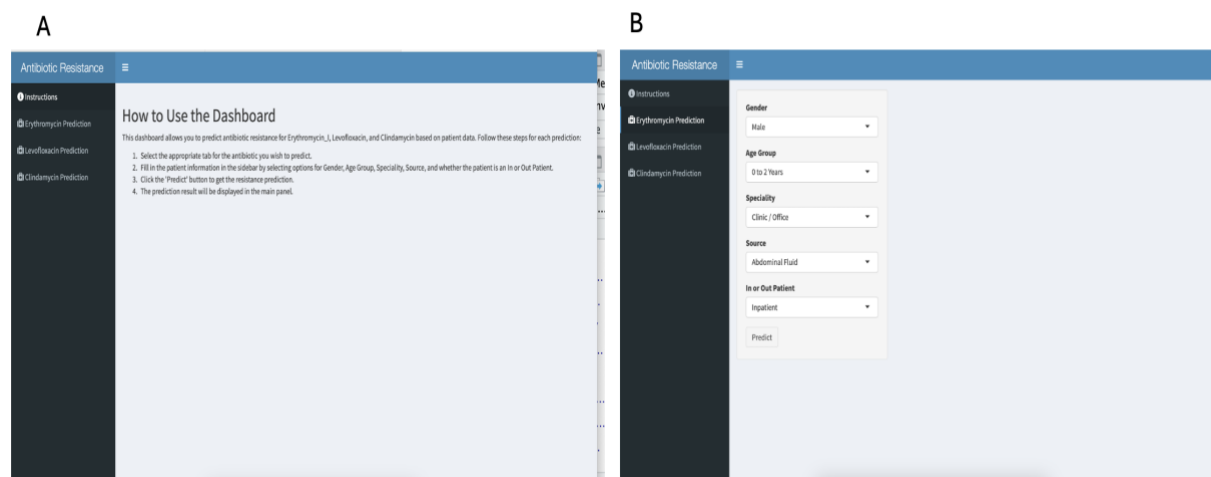
Table 2. Performance of AI Methods in Predicting Levofloxacin Susceptibility of *S. aureus*

AI Model	Sensitivity (Resistant Class)	Specificity (Resistant Class)	Accuracy
XGBBOOST	0.31628	0.65794	0.3198
Random Forest	0.24651	0.89878	0.7132
Support Vector Machine	1.0000	0.9965	0.9937

Table 3. Performance of AI Methods in Predicting Clindamycin Susceptibility of *S. aureus*

AI Model	Sensitivity (Resistant Class)	Specificity (Resistant Class)	Accuracy
XGBBOOST	0.30508	0.66484	0.3405
Random Forest	0.084746	0.991758	0.9212
Support Vector Machine	1.00000	0.99863	0.9975

Figure 1. Antibiotic Resistance Prediction Dashboard, A shows the instruction page, while B illustrates how to use the dashboard.



REFERENCES

1. <https://amr.vivli.org>
2. <https://www.who.int/initiatives/glass>
3. <https://www.who.int/publications/i/item/9789240093461>
4. R Core Team (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.URL <https://www.R-project.org/>.