ENHANCING REPORTING AND PREDICTION OF ADVERSE EFFECTS FOLLOWING MATERNAL AND CHILD BCG IMMUNIZATION IN RWANDA

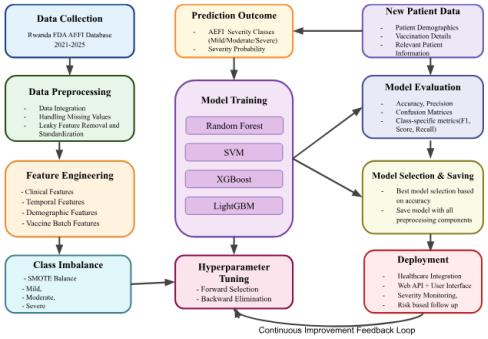
Nicole Igiraneza Ishimwe, Christine Niyizamwiyitira, Aimable Musafili, Michael Mugisha, Lazare Ntirenganya, Emmanuel Nkurunziza, Augustine Muhirwa, Angelique Uwamahoro

Introduction

Despite Rwanda's high BCG vaccination coverage (>90%), there is limited systematic analysis of AEFI severity patterns and predictive capabilities. The current passive reporting system lacks predictive analytics that could help healthcare providers anticipate and better manage potential severe adverse effects.

GAP: Current Passive Reporting System vs. Proposed Predictive Model

Methodology



Methodology Workflow

Methodology Summary

- **Data Collection:** Rwanda FDA's AEFI database (2021-2025) containing 5,830 reported cases
- **Data Preprocessing:** Cleaning, feature standardization, missing value imputation
- **Feature Engineering:** Created new features from patient demographics, vaccination details, and reaction types
- **Model Training:** Random Forest, SVM, XGBoost, and LightGBM models were trained and evaluated
- **Class Imbalance:** SMOTE technique to address significant

SMOTE Application

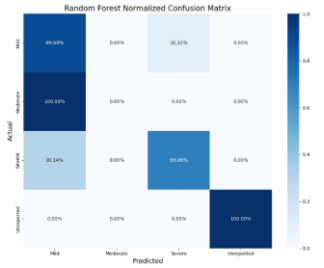
Original encoded class distribution:
- Class 0 (Mild): 3022 samples
- Class 1 (Moderate): 6 samples
- Class 2 (Severe): 1170 samples
- Class 3 (Unreported): 466 samples

After SMOTE application:
- Class 0 (Mild): 3022 samples
- Class 1 (Moderate): 3022 samples
- Class 2 (Severe): 3022 samples
- Class 3 (Unreported): 3022 samples

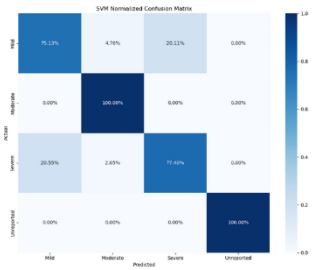
Results

Model Comparison

Model	CV Accuracy	Test Accuracy	Training time
Random Forest	94.33%	85.68%	15.3 sec
XGBoost	94.29%	84.48%	12.1 sec
LightGBM	94.37%	84.56%	4.8 sec
SVM	87.06%	78.22%	23.7 sec



Random Forest Confusion Matrix

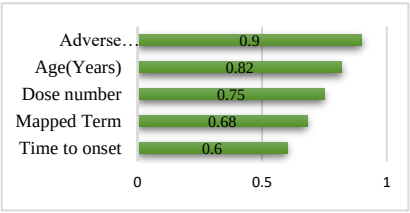


SVM Confusion Matrix

Key Observations: Random Forest has the highest overall accuracy of 85.68%, LightGBM has the best trade-off between accuracy and speed, and SVM has the Lower overall accuracy but better at identifying rare moderate cases

Feature Engineering

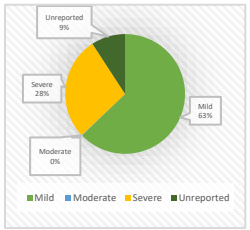
Key predictive features identified using Random Forest importance analysis:



Key Observations: Removed leaky features ("Serious" and "Seriousness criteria"), and then Applied SMOTE to balance training data (3022 samples per class), then 12 final features selected via Random Forest importance

Dataset Overview

Rwanda FDA's AEFI Database (2021-2025)
5,830 reported cases



Severity Distribution (Moderate cases: only 7 out of 5,830 = 0.1%, with 62.8 % of Mild, 27.9% of Severe, and 9.2 % of cases that were not reported)

Conclusions

Clinical Integration

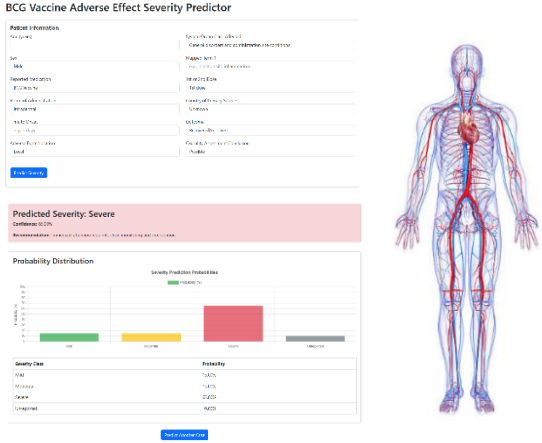
The prediction model seamlessly integrates with Rwanda's healthcare system to provide:
- Severity risk prediction at vaccination
- Automatic monitoring recommendations and evidence follow-up protocols
- Mobile interface for healthcare workers

Research Impact

- Machine learning can significantly improve prediction of AEFI severity following BCG vaccination
- Key predictors: adverse event location, patient age, dose number and reaction type
- The model provides a valuable addition to Rwanda's proactive vaccine safety monitoring system

Future Work

- Incorporate genetics/immunology data
- Expand to other vaccines (HPV, Measles)
- Develop mobile app for community health workers
- Integrate with active surveillance systems



Web based Prediction Tool Demo

Acknowledgements

The authors thank Carnegie Mellon University, Africa - Rwanda, for funding this study through the AFRETEC Network. We also express our gratitude to the Rwanda Food and Drugs Authority for their support and for providing the data used in this research.

Contact: nii@alumni.cmu.edu

CARNEGIE MELLON UNIVERISTY AFRICA & UNIVERSITY OF RWANDA