Predicting Medication Adherence Among Chronic Disease Patients in Ghana Using Machine Learning: A Data-Driven Approach

Samuel Danso *
Ghana Communication Technology University, Accra, Ghana

Isaac Nyantakyi
Ghana Communication Technology University, Accra, Ghana

Afriyie Karikari Bempah
Ghana Communication Technology University, Accra, Ghana

Enimil Kweku Boateng
Ghana Communication Technology University, Accra, Ghana

Francis Arku
Ghana Communication Technology University, Accra, Ghana

Samuel Bonsu-Duah
Ghana Communication Technology University, Accra, Ghana

Diana Danso
Ghana Communication Technology University, Accra, Ghana

Abstract

This study presents a context-specific machine learning framework for predicting medication adherence among chronic disease patients in Ghana using structured electronic medical record data. Existing research predominantly relies on generic modeling pipelines from high-income countries, often overlooking the unique socio-economic, demographic, and health system factors present in low- and middle-income contexts. To address this, we introduce a novel predictive pipeline that integrates domain-informed interaction features with a custom stacked ensemble architecture, whose meta-learner is rigorously tuned for local optimization.

The methodology applies multiple supervised algorithms—logistic regression, random forest, support vector machine, XGBoost, and multilayer perceptron—alongside robust feature engineering, cross-validated hyperparameter tuning, and systematic evaluation of class imbalance. Notably, while synthetic oversampling techniques such as SMOTE were explored to enhance minority class detection, the best performance was achieved on the original imbalanced dataset, as indicated by an F1-score of 0.9299 and ROC-AUC of 0.9336 for the final stacked ensemble model. Interpretability was prioritized through Shapley Additive Explanations (SHAP), revealing insurance status, age, medication cost, and comorbidity burden as critical predictors.

A systematic ablation study validated the importance of SHAP identified features, enabling deployment-tailored model variants: a 7-feature model maximizing performance (87.6% accuracy), a 5-feature model balancing efficiency (86.2% accuracy), and a streamlined 3-feature model suitable for low-resource environments (84.5% accuracy). This explainability-driven optimization approach delineates the essential predictive core while supporting flexible deployment across resource settings.

These findings demonstrate that combining model-level in novation, explainability, and transparent evaluation of class imbalance produces a scalable, clinically actionable framework for risk stratification and proactive intervention in resource constrained healthcare systems. The approach offers evidence based guidance for balancing complexity, interpretability, and practical constraints in LMIC settings.