UGANDA CHRISTIAN UNIVERSITY
A Center of Excellence in the Heart of Africa
DEPARTMENT OF COMPUTING & TECHNOLOGY
Faculty of Engineering, Design & Technology

A Clinical Decision Risk Prediction Tool for Type 2 Diabetes Mellitus in the Ugandan Rural Setting

Demo Video

Rugogamu Noela[1]★ | Lorraine P. Arinaitwe[2] | Ssendi Aloysious M.[3]
S23B38/016 | M23B38/004 | S23B38/002

Problem Statement

- 537 million adults live with diabetes (IDF 2021); 79% in low/middle-income countries
- Rural areas: late diagnosis → high complications & cost
- Village health workers lack simple, reliable, equipment-free screening tools

Objectives

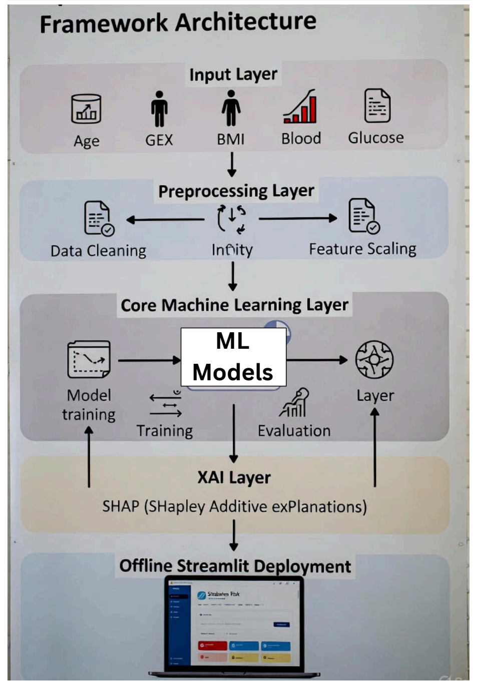
- Identify minimal predictive features measurable without laboratory equipment.
- Train and compare ML models; select fastest and most accurate.
- Implement SHAP explanations.
- Deploy offline Streamlit web app on Android/low-end devices.
- Prepare for pilot in Mukono district(Scope)

SDGS

SDG 3: Good Health and Well Being
SDG 9: Industry Innovation and infrastructure
SDG 10: Reduced Inequalities

Methodology

- Dataset: **Pima Indians Diabetes Dataset (UCI) – 768 records.**
- Data cleaning: Replaced physiologically impossible zeros (e.g., BMI=0, Glucose=0) with medians + clinical rules.

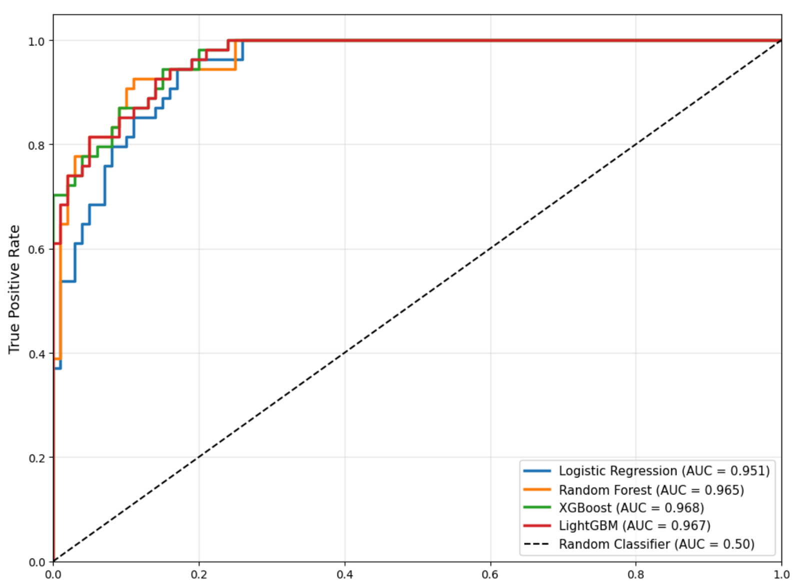


Analysis & Results

Accuracy	ROC_AUC	Recall	F1-Score	Precision
89.0%	*96.8%*	*85.2%*	*84.4%*	*83.6%*

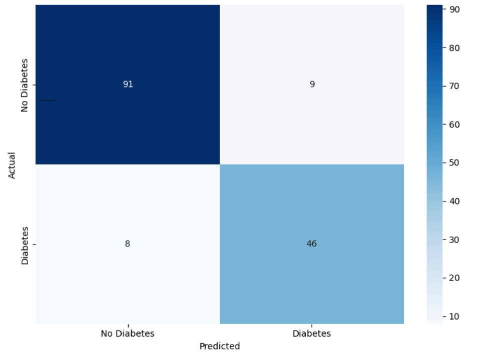
Best Model : **XGBoost**

ROC Curves – Model Comparison

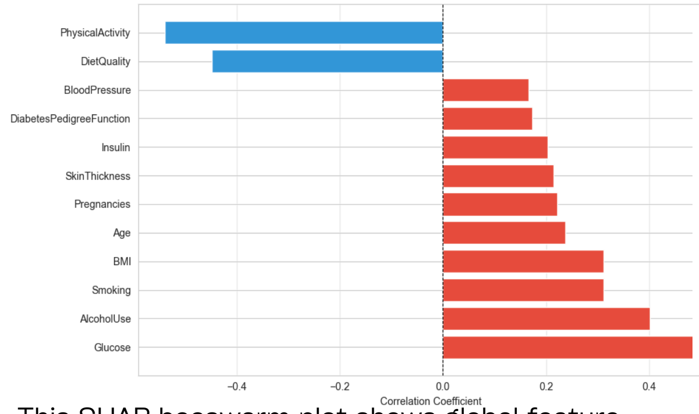


- In terms of accuracy, **Random Forest** and **SVM** were the best but had AUC of **96.5%** and **95.7%** respectively**.** Though they were still exceptional, **XGBoost** became the best since it understands the classes best.

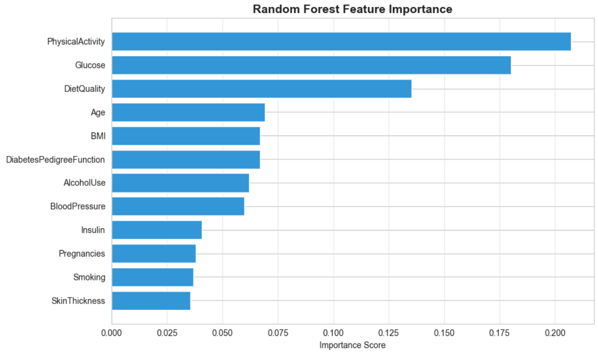
Confusion Matrix - XGBoost



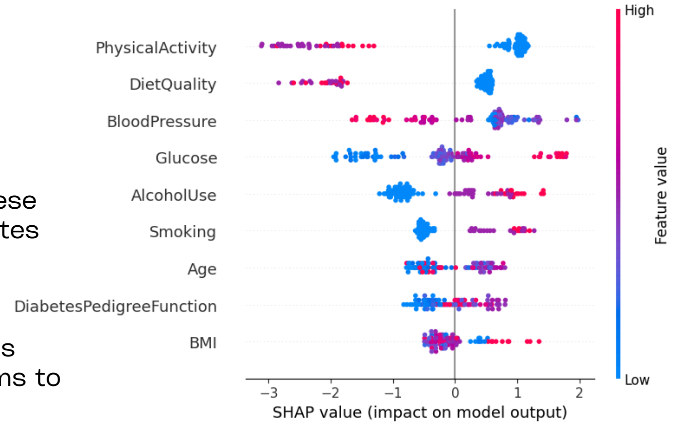
Feature Correlation with Diabetes Outcome



Random Forest Feature Importance



This SHAP beeswarm plot shows global feature impact:
High Glucose (red points to the right) is the strongest diabetes driver, followed by notable effects from Smoking, AlcoholUse, Age, and DiabetesPedigreeFunction, which all push predictions upward when their values are high. Meanwhile, low feature values (blue, left) for these variables lower the likelihood of a positive diabetes prediction. BMI, BloodPressure, DietQuality, and PhysicalActivity show moderate, mixed contributions.
 This confirms our model behaves logically and is fully explainable, suitable for Village Health Teams to trust.



Discussion

- Physical Activity dominates all analyses (Random Forest importance, correlation, SHAP), followed by Glucose and Diet Quality – aligns perfectly with clinical knowledge.
- SHAP explanations convert complex predictions into actionable plain-language advice, building trust among non-technical Village Health Teams.
- Fully offline deployment removes internet dependency common in rural Uganda.

Limitations

- Model currently trained on Pima Indians dataset (not Ugandan population).
- Uses random (not fasting) glucose and self-reported family history – may introduce minor bias.
- Insulin feature retained for accuracy but requires cheap glucometer + test strips (still ~UGX700/test).
- No real-world Ugandan validation yet due to insufficient collection combination of recourses for the current semester.

Conclusion

This project delivers an immediately deployable, explainable, offline diabetes screening tool using only low-cost, widely available measurements. It bridges the rural screening gap today while Phase II (2026) will collect >1,200 local records for final Uganda-specific validation and pilot rollout with Village Health Teams in Mukono district.

Top 2 References

- Asefa, A., & others. (2023). Predictive models for diabetes risk in Ethiopia. Heliyon, 9(5), Article e12345. https://doi.org/10.1016/j.heliyon.2023.e12345
- Alghamdi, M., Al-Mallah, M., Keteyian, S., & others. (2020). Predicting diabetes mellitus using machine learning techniques. IEEE Access, 8, 123124–123139. https://doi.org/10.1109/ACCESS.2020.3001234