

Prostate Cancer Risk Prediction: A Clinical Decision Support Dashboard for South African Patients.

Tom Balay, Blessing Magabane

Department of Computer Science, SAPCS, CSIR



Department of Computer Science
Faculty of Engineering, Built Environment and Information Technology
Capstone Project — MIT 808
Course Coordinators:
Dr. Vukosi Marivate (vukosi.marivate@cs.up.ac.za)
Abiodun Modupe (abiodun.modupe@cs.up.ac.za)

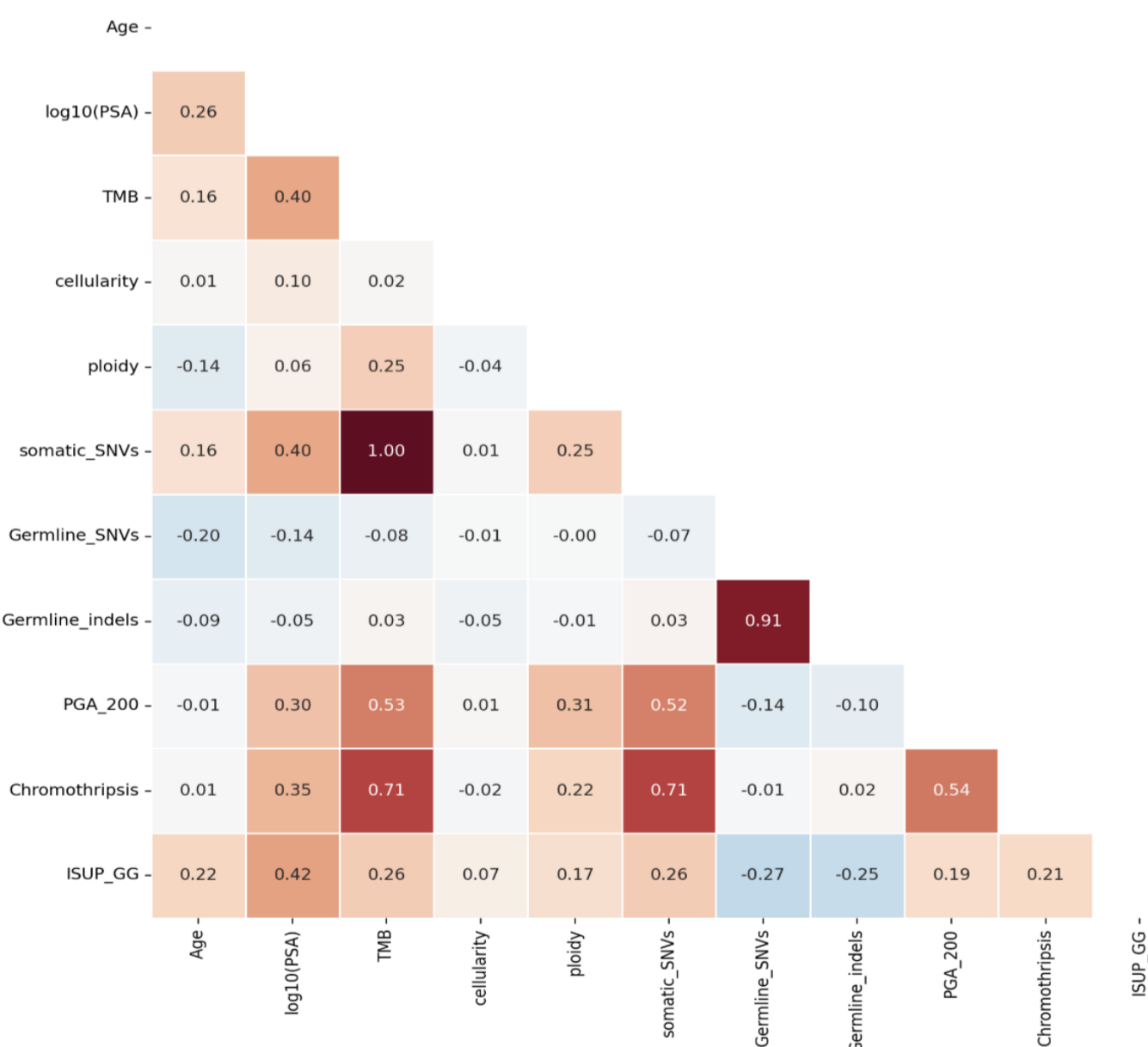
Background

- African men face higher prostate cancer incidence and aggressiveness, yet genomic studies are >90% European-ancestry
- The SAPCS cohort (Jaratlerdsiri *et al.*, Nature 2022) is the first large African whole-genome sequencing dataset — but no decision-support tool exists for it yet
- We built an interactive clinical dashboard with a built-in classifier trained exclusively on South African patients

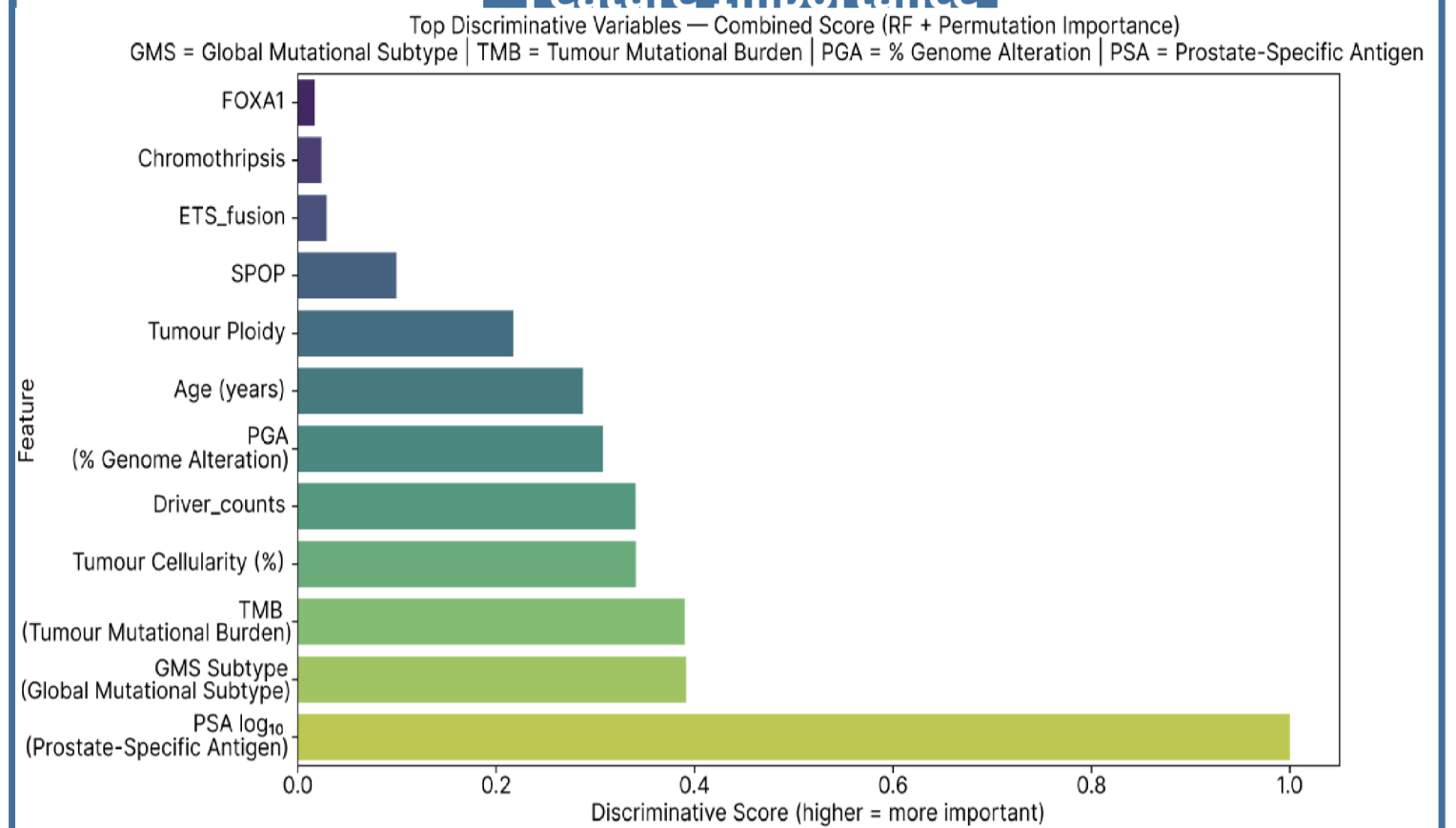
Methods

- N = 115 labelled patients (123 SA patients; 8 excluded — missing ISUP_GG)
- Features from whole-genome sequencing: Age, log(PSA), TMB, Cellularity, Ploidy, PGA_200, Chromothripsis, Driver counts, GMS subtype
- Binary target: High-Risk (ISUP GG 4–5, n=84) vs Low-Risk (ISUP 1–2–3, n=29)
- Random Forest + SMOTENC (class imbalance correction) / Stratified K-Fold CV
- Threshold optimised on Precision-Recall curve to hit Recall $\geq 85\%$

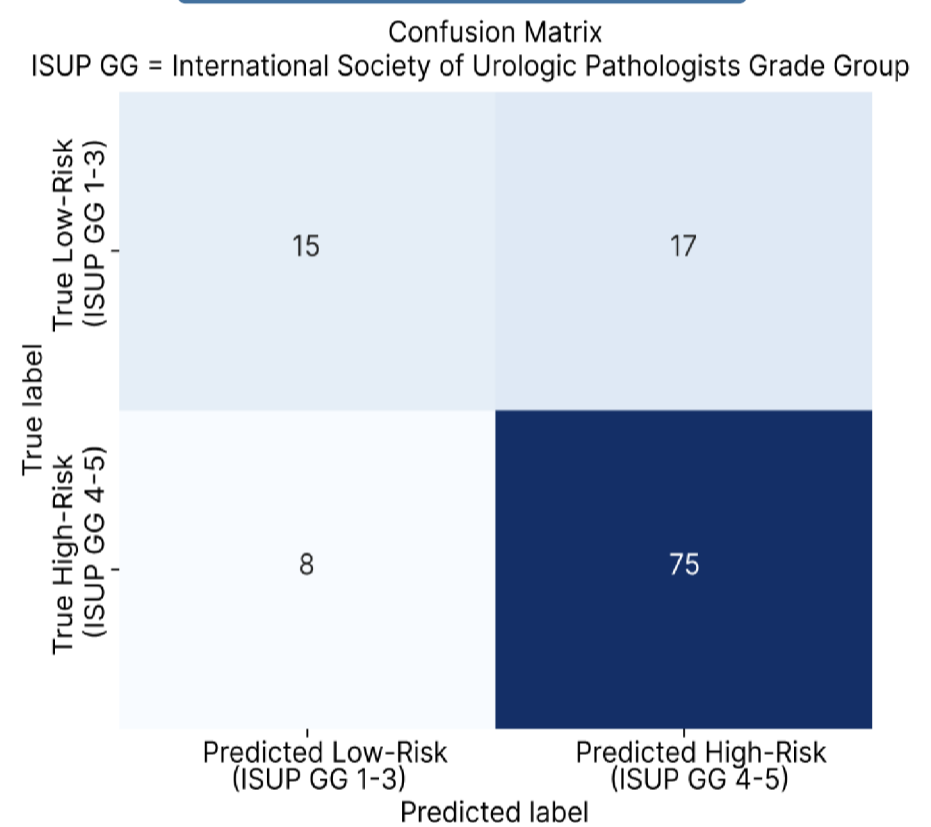
EDA



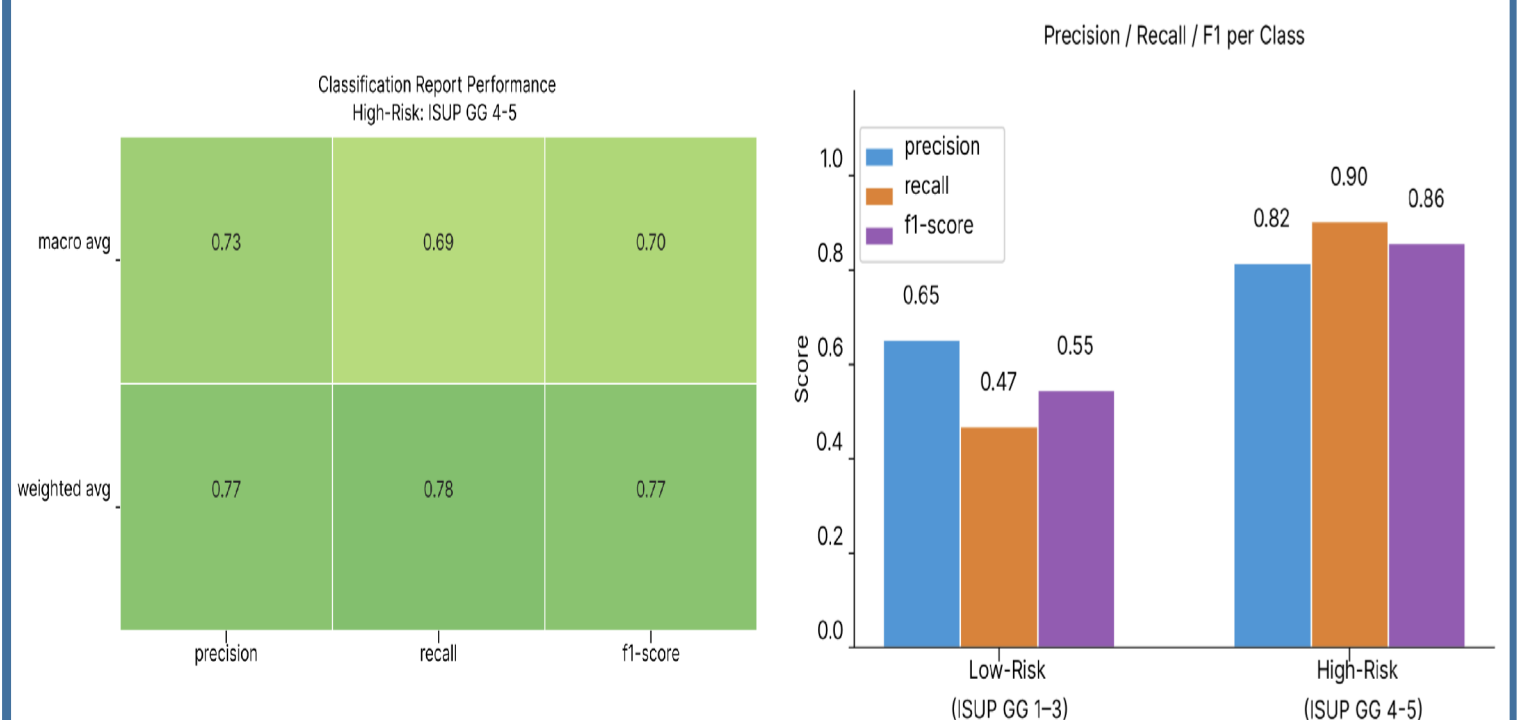
Random Forest Feature Importance



Confusion Matrix



Model Performance



Discussion

- The Random Forest achieved strong overall performance (weighted F1 = 0.77, precision = 0.77) with a high-risk recall of 86.7%, meeting our primary clinical threshold. SHAP analysis confirmed that the top predictors — PSA (log), TMB, Age, Driver counts, GMS, Cellularity.
- Class imbalance persists despite SMOTE, limiting confidence in minority-class predictions. Binary stratification (low vs high ISUP GG) sacrifices the nuance of intermediate grades — a multiclass approach becomes viable with larger samples.
- Future work priorities are cohort expansion to strengthen minority-class representation, and integration of per-patient SHAP waterfall charts directly into the dashboard for point-of-care
- The dashboard lets clinicians add new patients and retrain in real time, staying current as more SA data arrives — unlike static models

Scan me

