ML approaches for improved T2D risk stratification in the Qatar Biobank

Ammira Akil
Sidra Medicine's Metabolic and Mendelian Clinical Genomic Research Program, Doha, Qatar

Type 2 diabetes (T2D) represents a significant health concern, particularly in Middle Eastern populations, where its prevalence is exceedingly high. Existing polygenic risk scores (PRS) based on European-derived cohorts show limited effectiveness within these diverse populations. This study aimed to enhance T2D risk stratification through multi-trait polygenic ensemble modeling using data from the Qatar Biobank, comprising over 14,000 participants. We conducted genome-wide association studies on T2D and 11 related metabolic traits, developing population-specific PRS that improved predictive performance, achieving an AUC of 0.85. Stratifying individuals by T2D-PRS revealed a sixfold increase in disease risk among the top decile. This optimization was validated externally in a diverse UK Biobank cohort, demonstrating an AUC of 0.82. Our findings underscore the importance of integrating correlated metabolic traits within ensemble machine-learning frameworks to improve risk prediction and transferability across ancestries. This research highlights the potential of tailored genetic models to enhance clinical strategies for managing T2D in underrepresented populations, paving the way for more effective precision medicine.