nature medicine

Supplementary information

https://doi.org/10.1038/s41591-025-03849-7

# Multimodal AI correlates of glucose spikes in people with normal glucose regulation, pre-diabetes and type 2 diabetes

In the format provided by the authors and unedited

## Supplementary Results

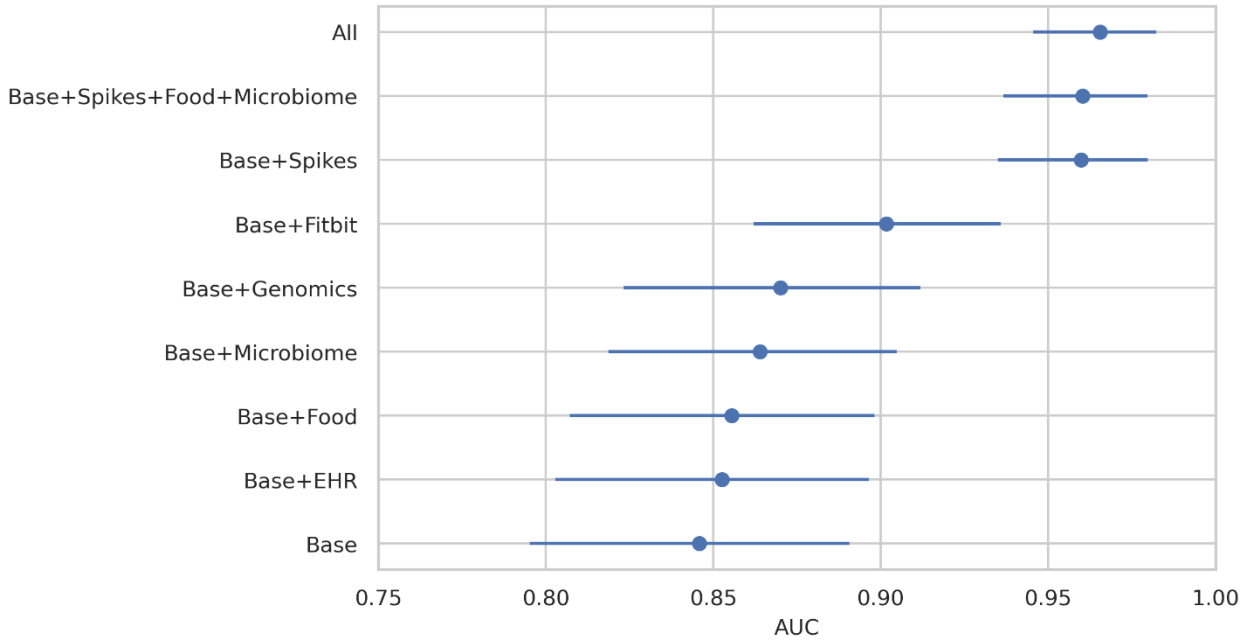
### Evaluation of the multimodal model on binary classification

The primary assessment of the trained multimodal models focused on their performance in the binary classification task of distinguishing between PROGRESS T2D and normoglycemic individuals. The best performance was obtained when variables from all data modalities were used ('All'), with an AUC of 0.97 (confidence interval, CI: 0.95-0.98), an improvement in AUC of 0.12 (CI: 0.08-0.17, p-value<0.001), compared to the base model using only demographic and anthropometric variables, that had an AUC of 0.85 (CI: 0.80-0.89). Overall, the inclusion of variables derived from wearable devices such as CGM and Fitbit, led to statistically significant enhancements compared to the base model. Specifically, there was an improvement in AUC of 0.11 (CI: 0.07-0.16, p-value<0.001) for glucose spike metrics, and 0.06 (CI: 0.02-0.09, p-value=0.003) for Fitbit variables. However, there were no statistically significant improvements observed when adding microbiome variables (p-value=0.164), genomics variables (p-value=0.058), food intake variables (p-value=0.265), or EHR variables (p-value=0.441) to the base variables as input features. (Supplementary Figure 1)

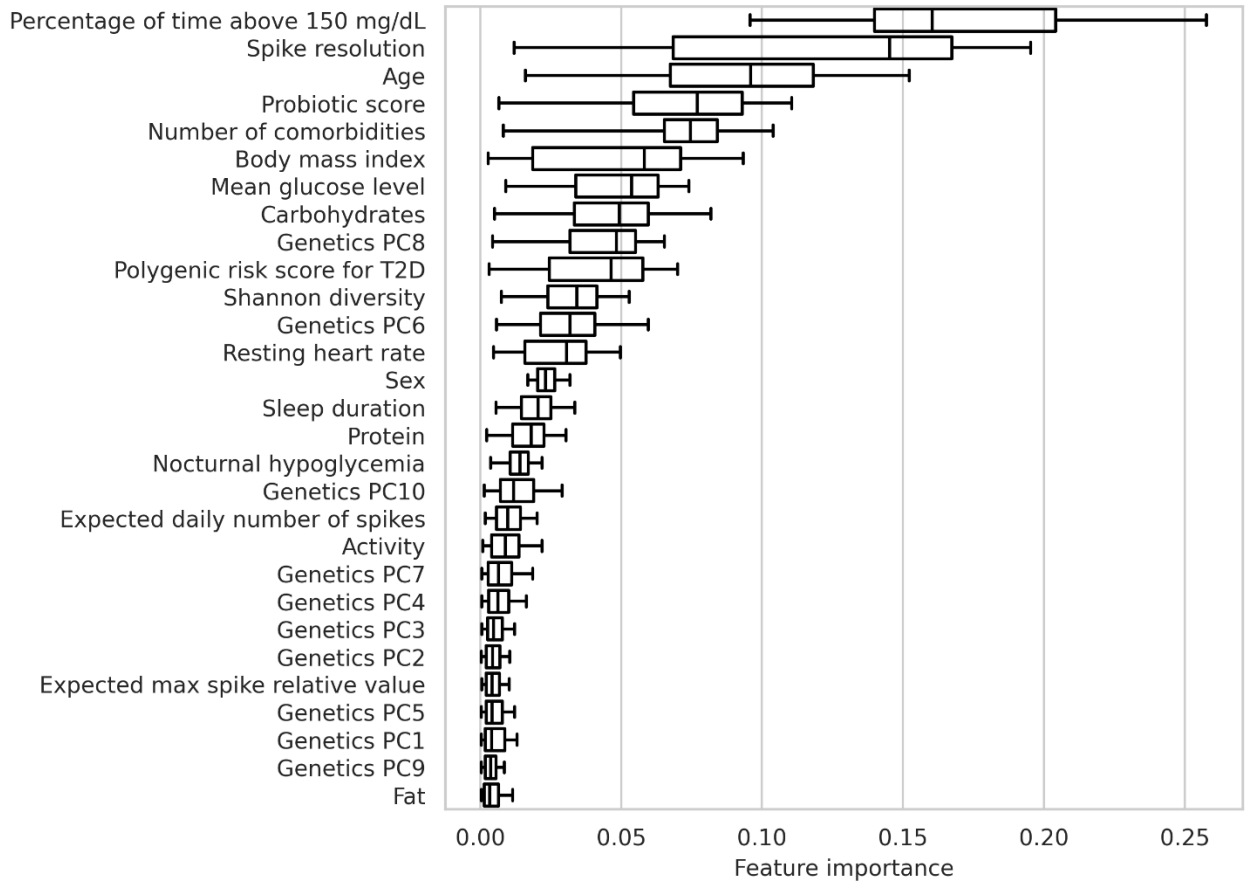
To evaluate the model's generalizability, we assessed the performance of the model trained on PROGRESS with base features, glucose spike features, food features and microbiome features as inputs on the HPP dataset based on the measured HbA1c value. We considered HPP participants with HbA1c levels in the normoglycemia range (HbA1c < 5.7 %, n=1548) and the T2D range (HbA1c ≥ 6.5 %, n=27). The model achieved an AUC = 0.90 (CI: 0.82-0.97). We also evaluated the model's area under the precision-recall curve (AUPRC), obtaining an AUPRC = 0.60 (CI: 0.40-

0.77), which is substantially higher than the baseline prevalence of the minority class (1.7%), indicating strong predictive performance in distinguishing positive cases despite the class imbalance.

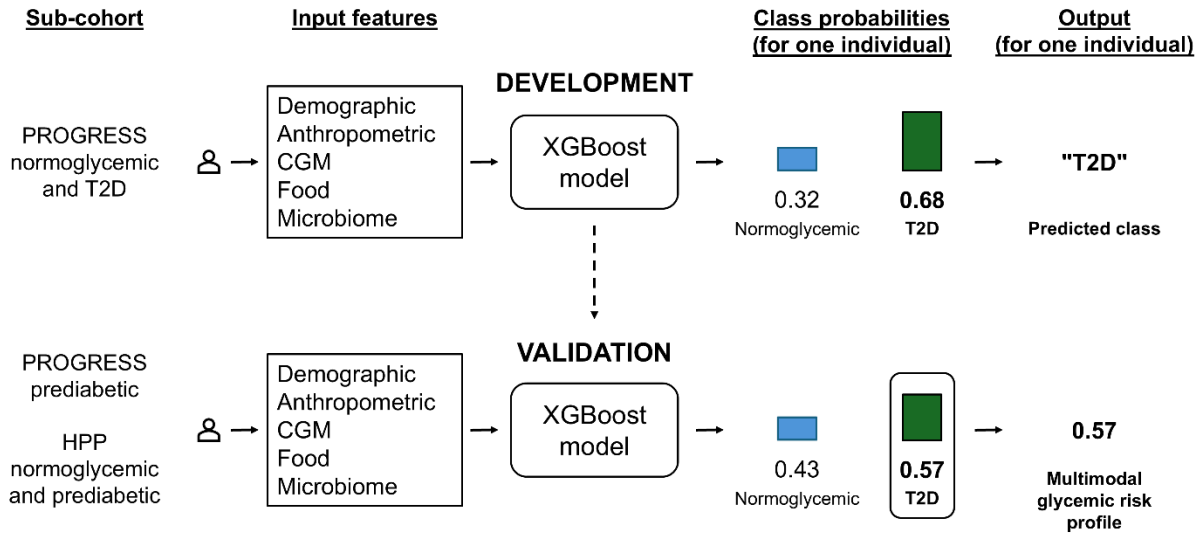
## Supplementary Figures




**Supplementary Figure 1 | AUC evaluation for the binary classification task of separating T2D and normoglycemic individuals in PROGRESS.** *Evaluation of the AUC metric for the binary classification models with different sets of input features. The horizontal lines represent 95% confidence intervals estimated with a bootstrap method with 10000 iterations. AUC, area under the curve; EHR, electronic health record.*




**Supplementary Figure 2 | Feature importance evaluation for the model using all features in PROGRESS.** *Feature importance scores for the model using all features, evaluated on the PROGRESS dataset for the binary task of separating T2D and normoglycemic individuals. The boxes represent the interquartile range, the vertical lines are the median values, the whiskers show the 5th and 95th percentiles. T2D, type 2 diabetes; PC, principal component.*




**Supplementary Figure 3 | Multimodal model's development and validation.** *Development was performed on the PROGRESS normoglycemic and T2D participants, validation was performed on the PROGRESS prediabetic participants and the HPP normoglycemic and prediabetic participants. T2D, type 2 diabetes; CGM, continuous glucose monitoring.*