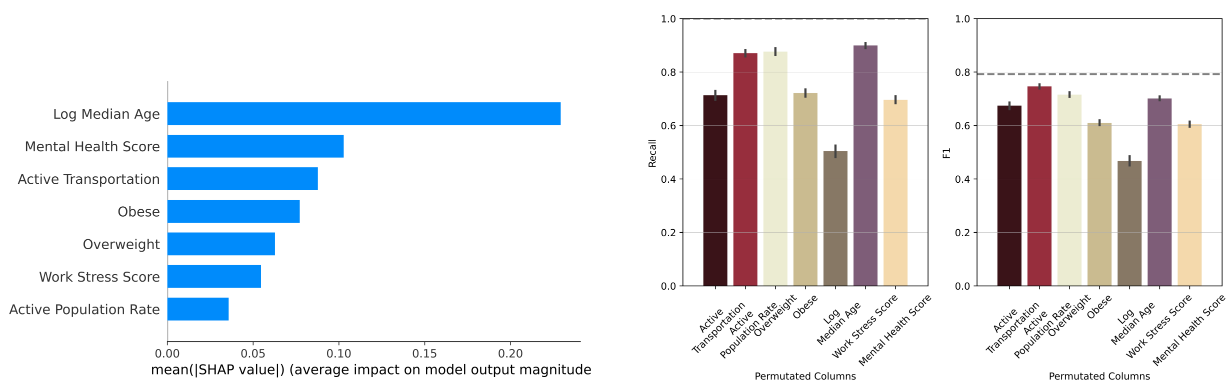
SUPPLEMENTARY INFORMATION

Mapping Neighborhood-Level Drivers of Type 2 Diabetes: A Predictive-Causal Approach for Precision Public Health

Contents

1 SUPPLEMENTARY NOTE 1: SENSITIVITY ANALYSIS FOR THE EXTENDED MODELS



(a) SHAP Analysis

(b) Sensitivity Analysis

Supplementary Figure 1. SHAP and Sensitivity Analysis of the extended SVM model on the Toronto CMA dataset with seven variables (Test dataset). (a) shows that Mental Health Score emerges as a highly important feature, with strong overlap in the Sensitivity Analysis. While not perfectly aligned, both analyses rank Mental Health Score as the second most important feature, whereas Work Stress Score appears second least important. (b) indicates that among the two added features, Mental Health Score has the second largest impact, while Work Stress Score shows the lowest effect on the model's sensitivity. The figure presents the most reliable extracted result from multiple runs.

1.1 Extended SHAP and Sensitivity Analysis for the SVM Model (Test Dataset)

Supplementary Figure 1 presents the outcomes of the extended analysis for the SVM model on the Toronto CMA dataset, using seven variables in the Test set: (a) shows that `Mental Health Score` has high SHAP values, suggesting a strong influence on diabetes risk. This is consistent with the SHAP formula,

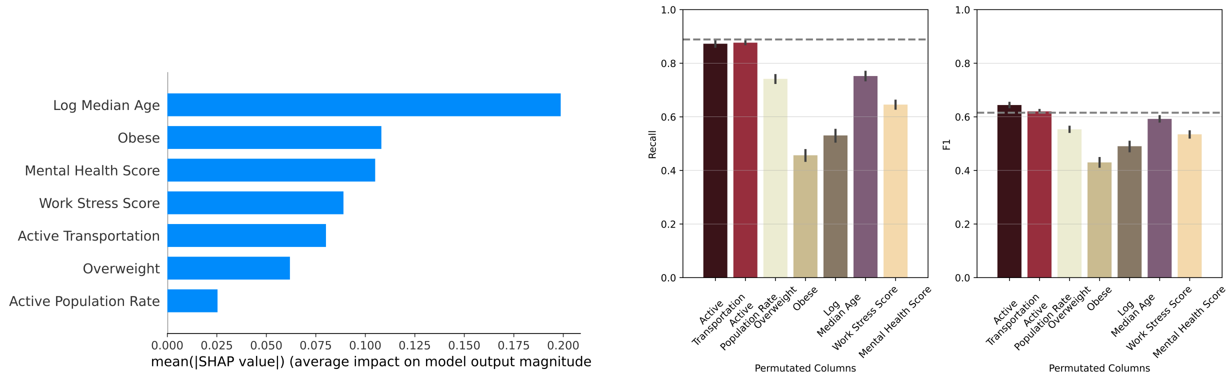
$$\phi_i(\mathbf{x}) = \mathbb{E}\big[f(S \cup \{i\}, \mathbf{x}) - f(S, \mathbf{x})\big],$$

which calculates the local effect of each feature by comparing predictions with and without that feature across all subsets S. Mental well-being may contribute to or protect against diabetes risk through stress-related metabolic changes or lifestyle patterns; (b) confirms that permuting Mental Health Score leads to the largest drop in recall and F1, reinforcing its central role in identifying higher-prevalence neighborhoods. Work Stress Score shows a lesser impact, possibly because it captures a narrower range of psychosocial influences compared to overall mental well-being.

These outcomes support the idea that mental health interventions could help reduce diabetes prevalence in at-risk communities, since psychosocial factors appear critical at the neighborhood level. The figure reflects the most stable result among multiple runs, highlighting that both local contributions (SHAP) and global feature sensitivity (permutation test) point to the importance of addressing psychological and behavioral factors in diabetes risk estimation.

1.2 Extended SHAP and Sensitivity Analysis for the SVM Model (External Dataset)

Supplementary Figure 2 presents a parallel analysis for the external dataset: (a) again identifies Mental Health Score as a key predictor, ranking it second only to Log Median Age. This result points to a possible paradox in real-world settings. Certain communities may report strong social ties or cultural practices that elevate self-reported mental well-being, yet those same areas might lack safe venues for physical activity or face limited access to healthy foods, pushing diabetes risk upward; (b) shows that permuting Mental Health Score causes a marked drop in recall and F1, while Work Stress Score has the smallest impact. These observations imply that improving mental health alone may not be enough if local economic or infrastructural barriers remain, suggesting a complex interaction between psychosocial and environmental factors in shaping neighborhood-level diabetes outcomes.



(a) SHAP Analysis

(b) Sensitivity Analysis

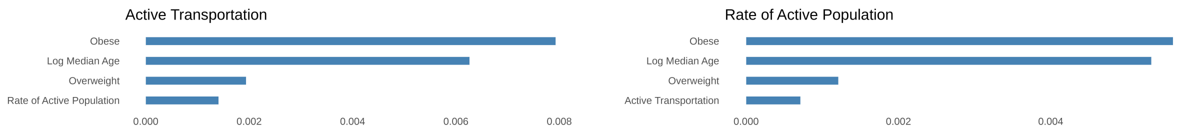
Supplementary Figure 2. SHAP and Sensitivity Analysis of the extended SVM model on the Toronto CMA dataset with seven variables (External dataset). (a) demonstrates that Mental Health Score emerges as a highly important feature, consistently ranking as the second most influential variable while Work Stress Score ranks as the second least important. (b) shows that among the two added features, Mental Health Score has the second largest impact on model sensitivity, whereas Work Stress Score exhibits the smallest effect. The figure presents the most reliable result obtained from multiple runs.

2 SUPPLEMENTARY NOTE 2: ITERATIVE SHAP VISUALIZATION FOR SVM FEATURE SETS

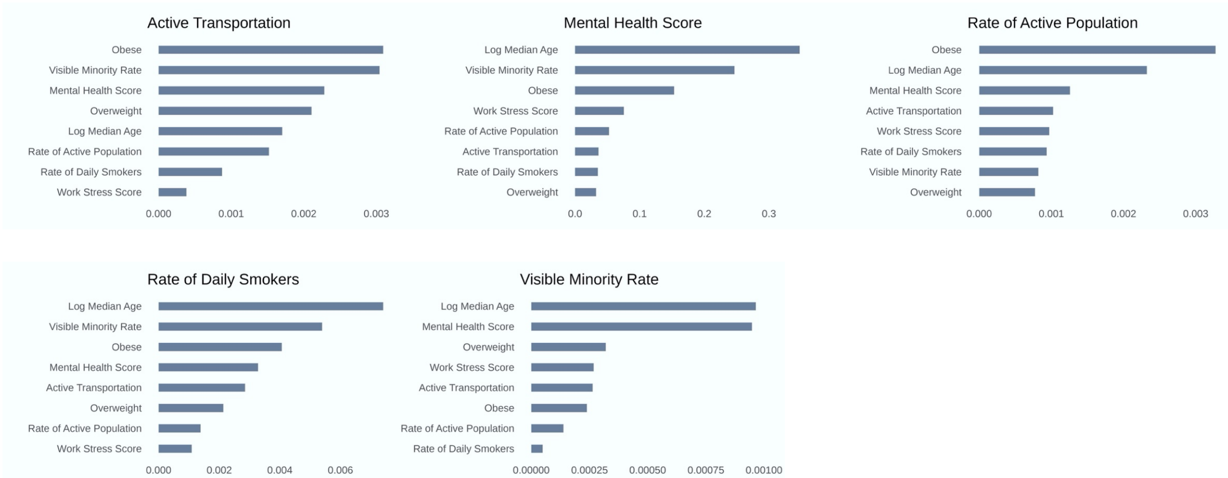
2.1 Iterative SHAP Visualization for SVM Feature Sets

Supplementary Figure 3 presents the comparison of the SVM model's base feature set (with five variables) and an extended set (with nine variables): Figure 3a includes Active Transportation and Rate of Active Population as potential treatment variables, whileFigure 3b adds elements such as Mental Health Score, Work Stress Score, and Visible Minority Rate. These additions provide a broader view of the social and behavioral influences that might shape neighborhood-level diabetes risk. Each panel reports SHAP values for selected treatment features, with the remaining ones considered covariates, clarifying how each variable contributes to the model's predictions under different configurations. `Log Median Age`, `Obesity`, and `Overweight` are excluded from the treatment sets to focus on psychosocial and mobility-related factors.

This iterative approach clarifies how adding new variables can shift the apparent importance of certain predictors. `Active Transportation` consistently shows strong contributions in both panels, though the introduction of `Mental Health Score` and `Visible Minority Rate` modifies the relative influence of `Rate of Active Population`. These insights go beyond the simple ranking of features, indicating that psychosocial, demographic, and behavioral factors interact in complex ways when shaping diabetes risk. The figure thus provides a richer perspective on how incremental expansions of the feature space can refine understanding of community-level health patterns, suggesting that targeted interventions may need to address multiple domains simultaneously.



(a) The SVM feature set with five features.

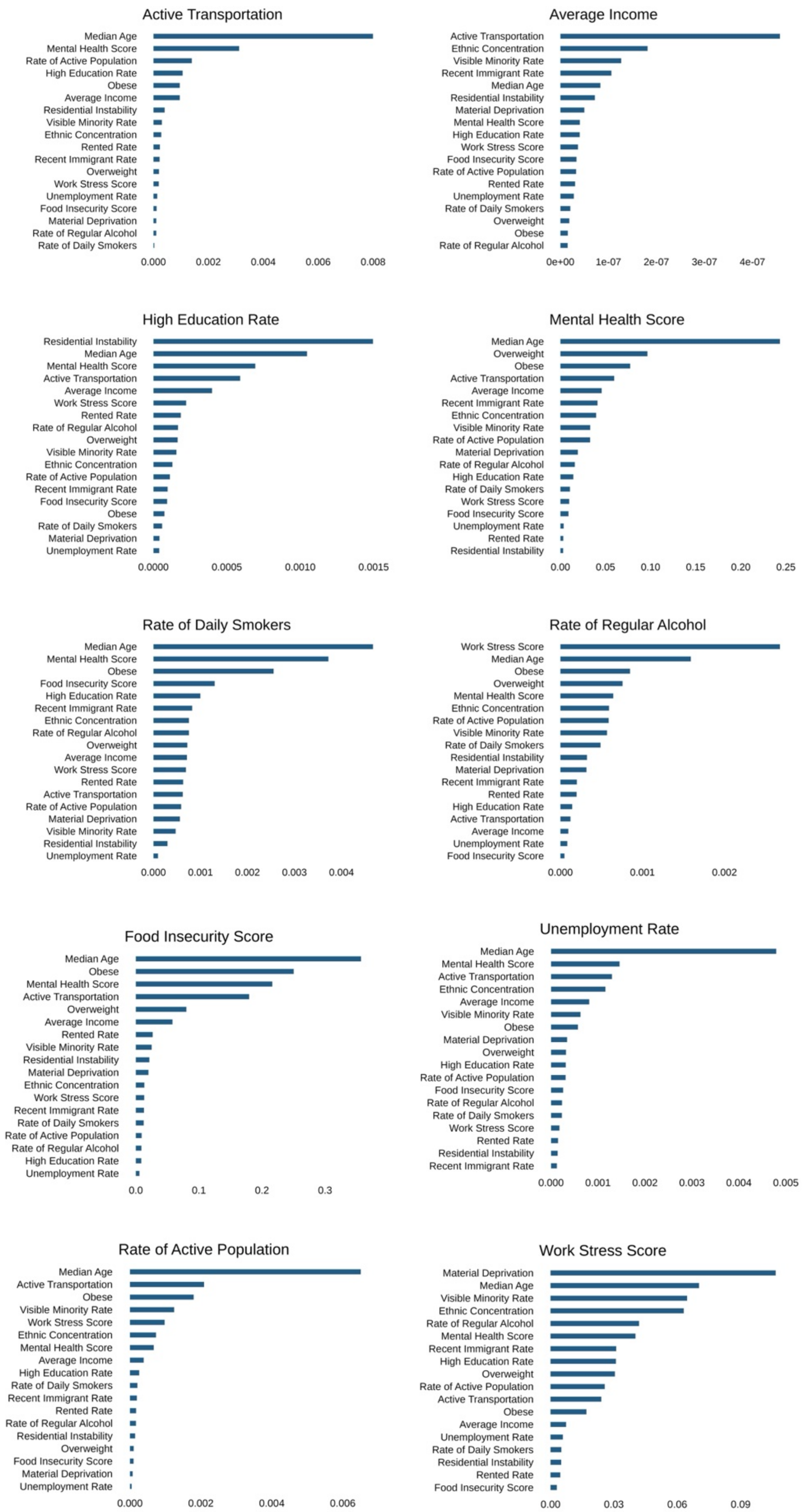


(b) The extended feature set with nine features.

Supplementary Figure 3. Iterative SHAP visualization: Analyzing feature impact by selecting one as treatment and others as covariates in a) the SVM feature set, and b) the extended feature set. Log median age, Obesity, and Overweight are excluded from the treatment sets.

2.2 Understanding Feature Interactions at Scale

Supplementary Figure 4 provides a detailed view of the full feature set for the SVM model, examining how each variable influences neighborhood-level diabetes risk when designated as the treatment while the rest act as covariates. The sub-figures highlight `Active Transportation`, `Average Income`, `High Education Rate`, `Mental Health Score`, `Rate of Daily Smokers`, `Rate of Regular Alcohol`, `Food Insecurity Score`, `Unemployment Rate`, `Rate of Active Population`, and `Work Stress Score`. The expanded selection of variables indicates that psychosocial and socioeconomic factors may overlap in their effects, at times obscuring simpler associations observed in smaller feature sets. For instance, `Work Stress Score` can appear less influential in the presence of `Mental Health Score`, and `Average Income` may combine with `High Education Rate` in ways that affect local prevalence. The figure suggests that diabetes risk arises from a complex mix of personal habits, resource availability, and social conditions.



Supplementary Figure 4. Iterative SHAP visualization: Analyzing feature impact by selecting one as treatment and others as covariates in the full feature set.