Additional file 1

Incremental value of a multidomain epidemiological questionnaire for discriminating thyroid cancer case status: a case–control study

Fig. S1. Numbers of cases and controls in each stratified outer test fold

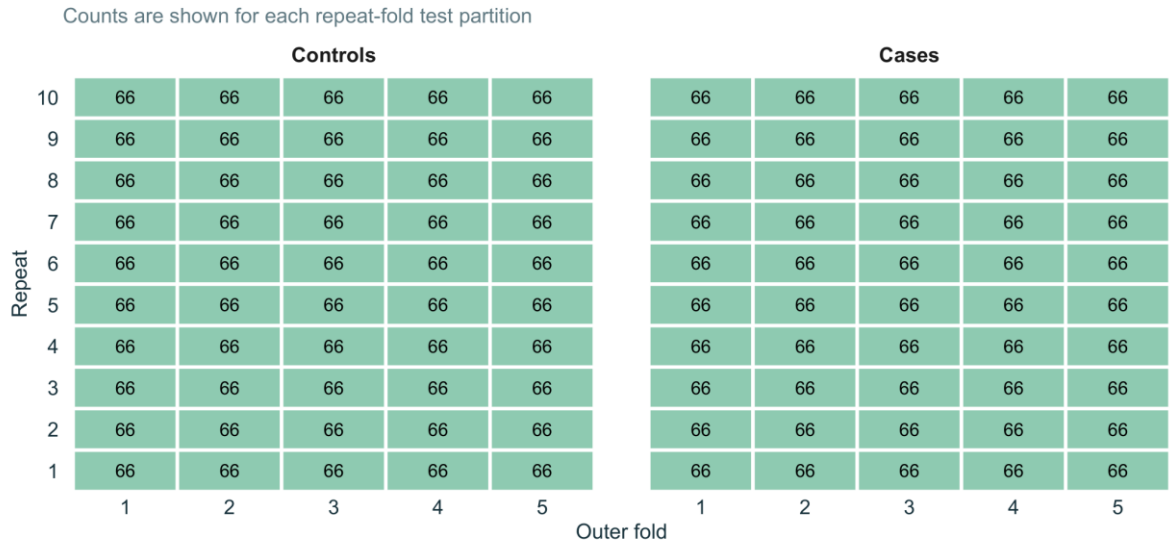
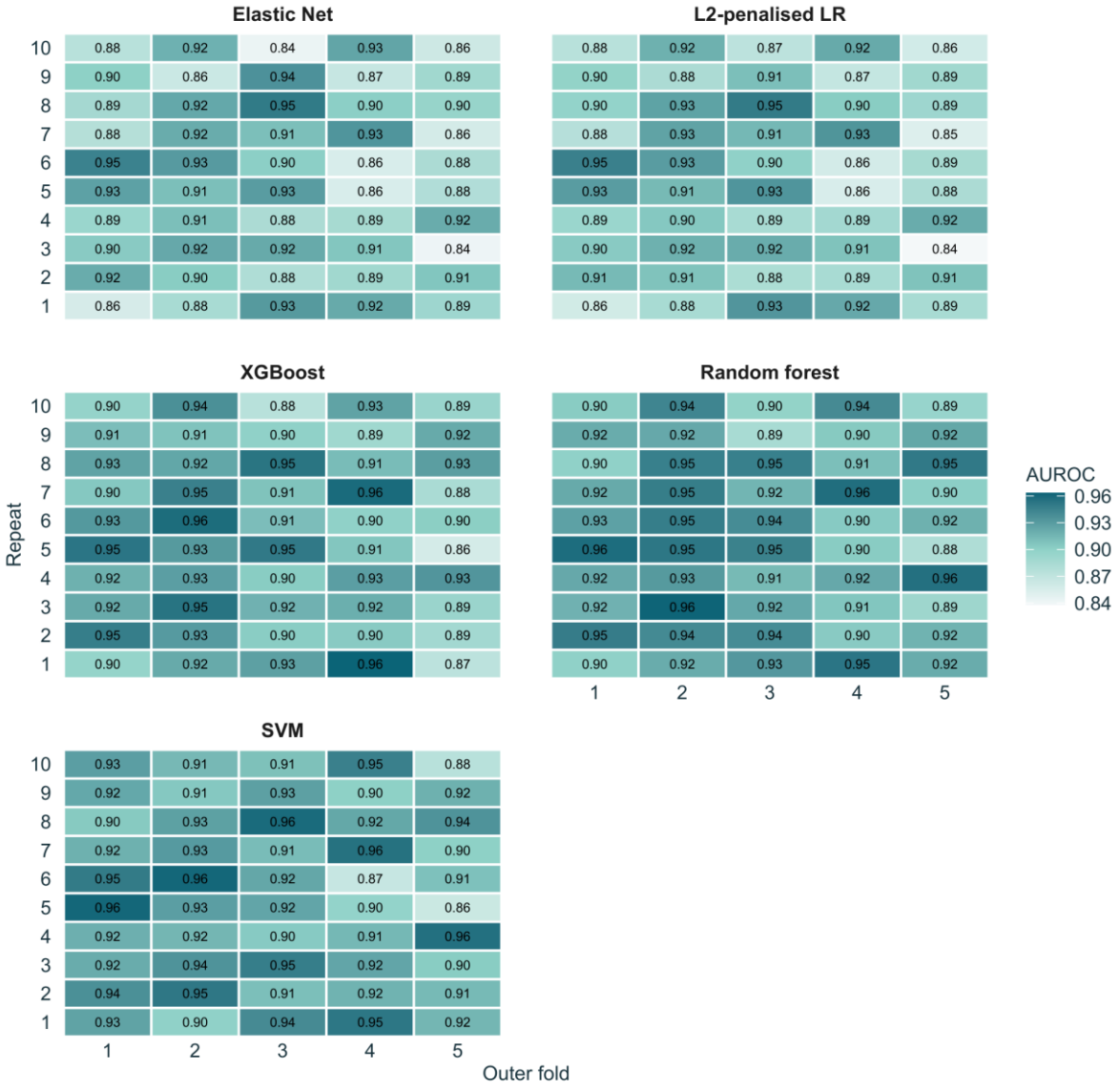


Fig. S2. AUROC heatmap for five algorithms across 50 outer test folds



AUROC, area under the receiver operating characteristic curve.

Fig. S3. Distributions of participant-level mean outer-test scores for the five full models

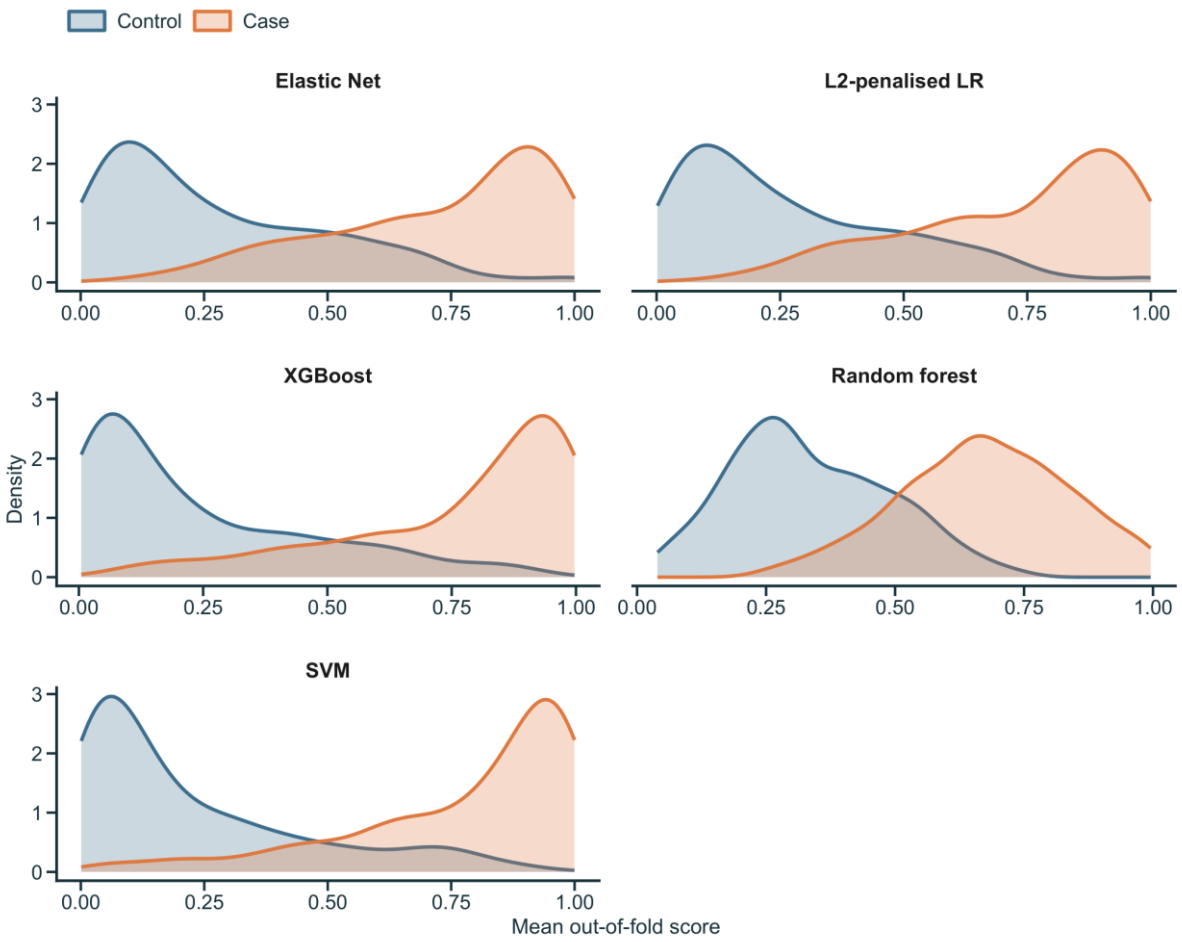


Fig. S4. Sensitivity analysis of progressive elastic net models

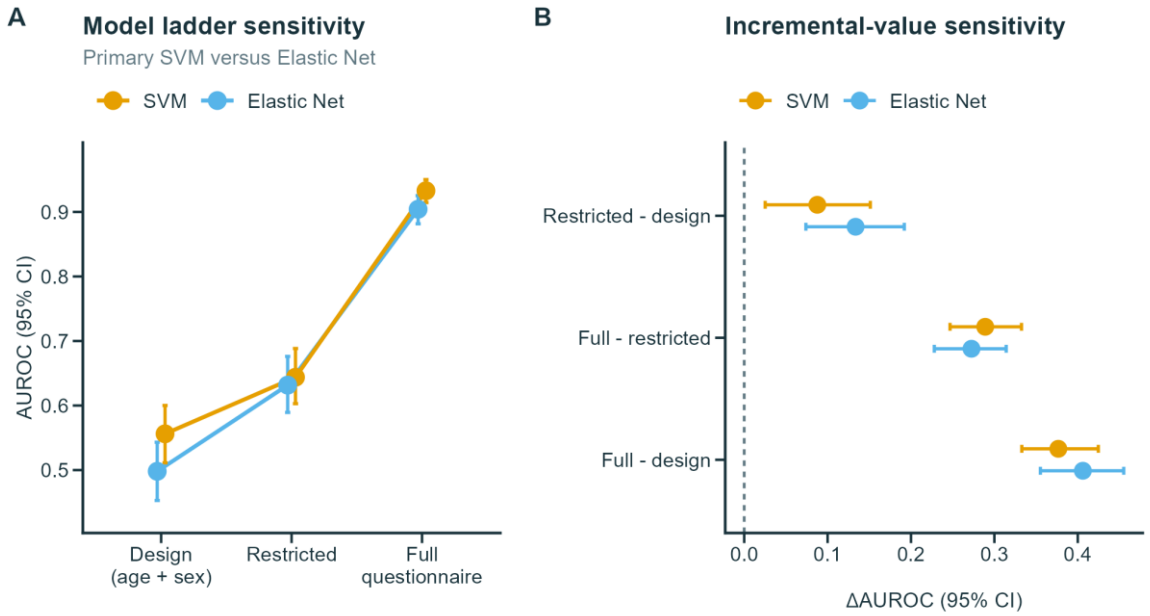


Fig. S5. Sensitivity analysis of domain addition and ablation using elastic net

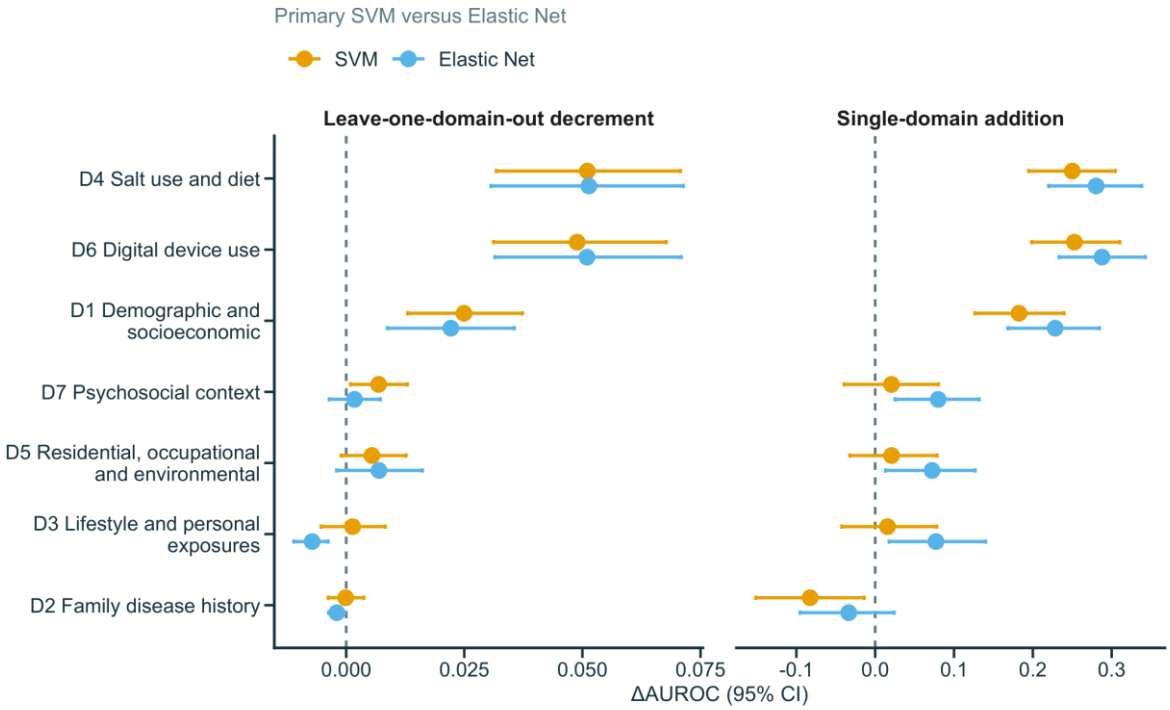
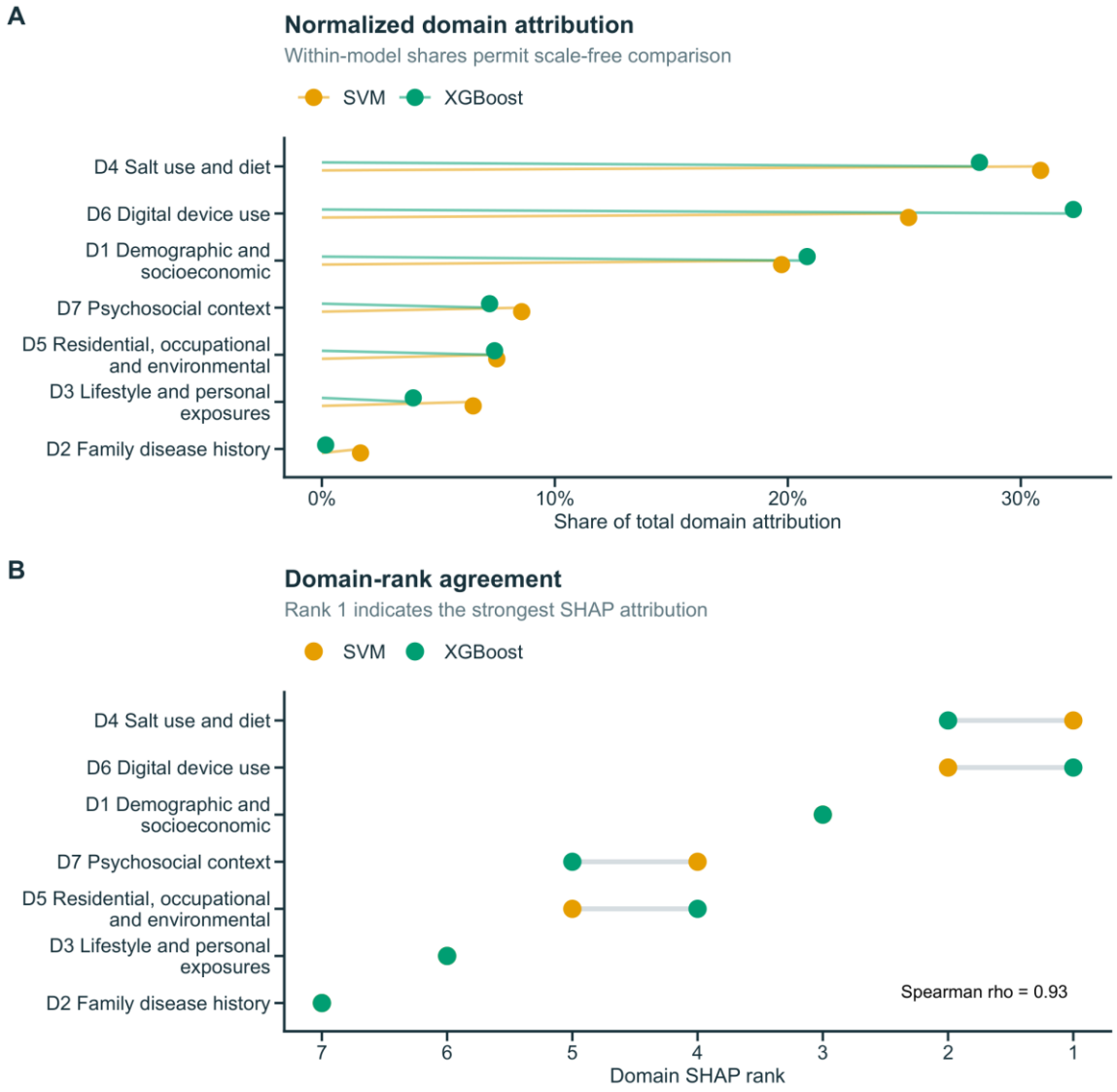
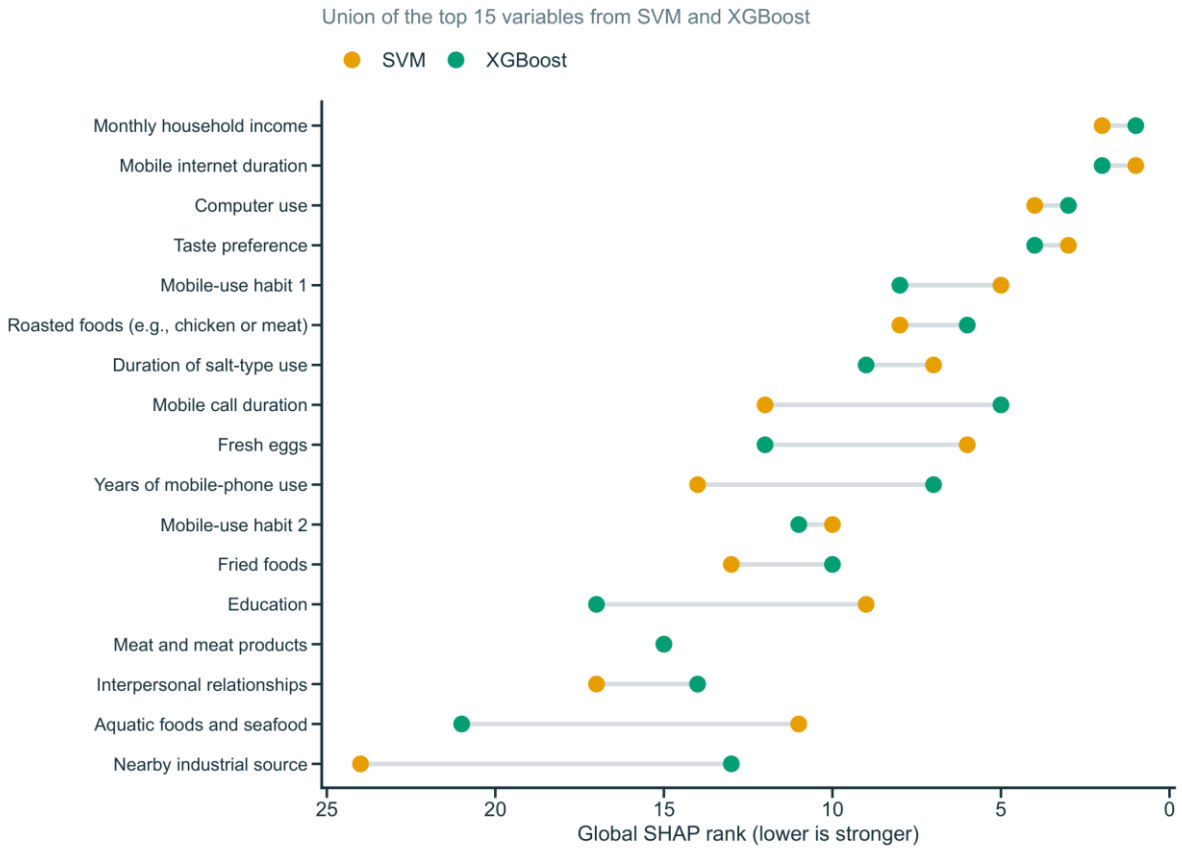


Fig. S6. Sensitivity analysis of domain-level SHAP attributions using XGBoost



SHAP, SHapley Additive exPlanations.

Fig. S7. Agreement between variable-level SHAP rankings from the support vector machine and XGBoost



SHAP, SHapley Additive exPlanations.

Fig. S8. Sensitivity analysis after adding the medical radiation domain

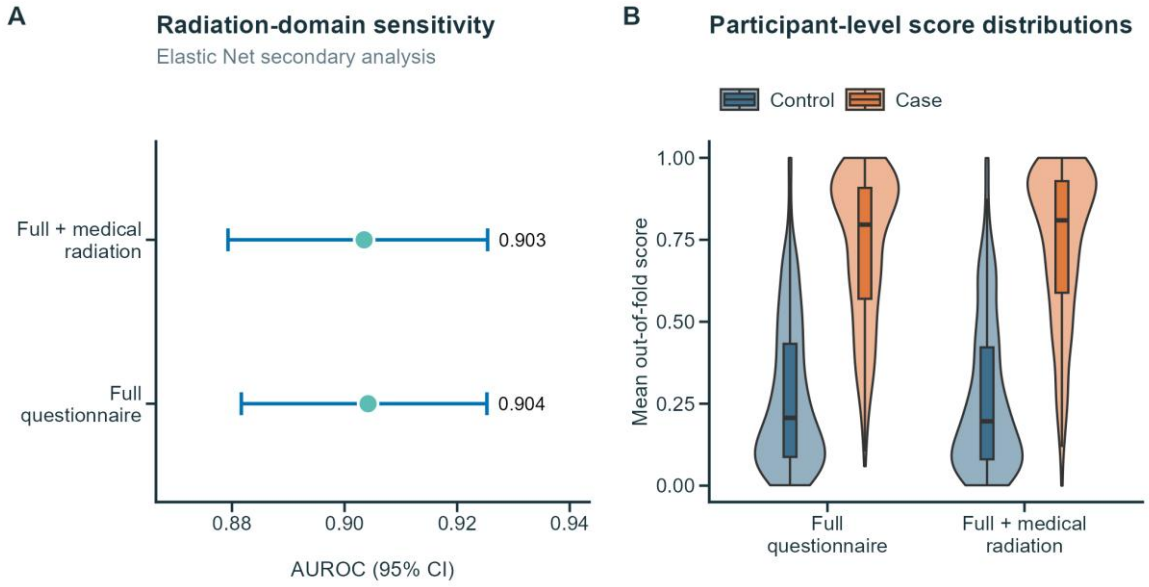


Fig. S9. Progressive model analysis in the female subgroup

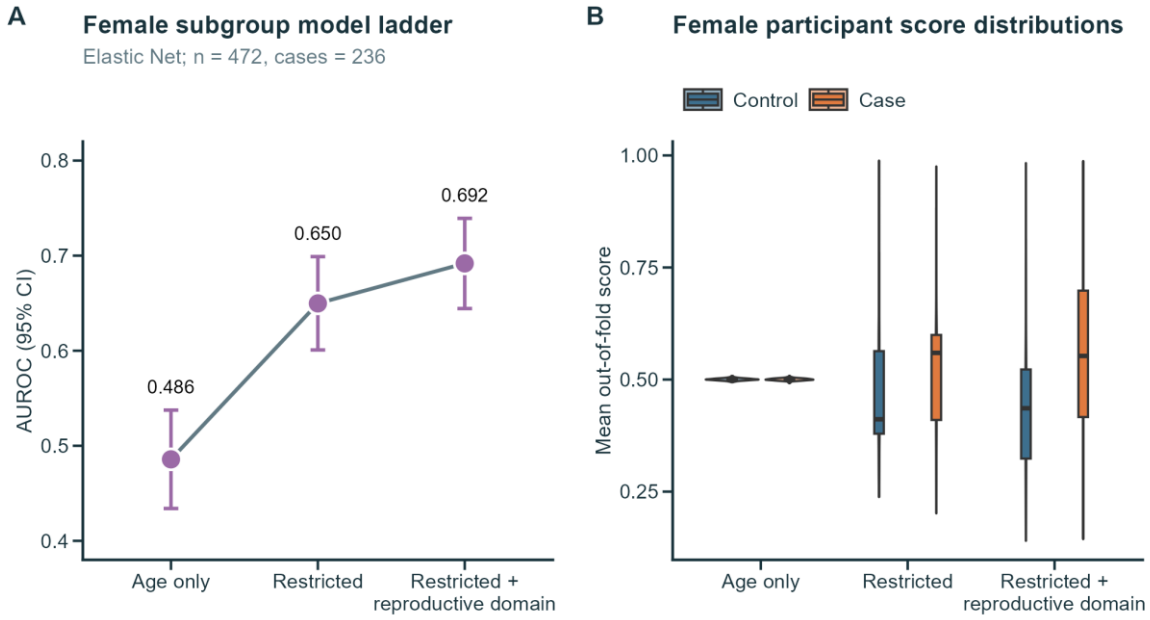


Fig. S10. Sensitivity analysis with algorithm selection within each outer training set

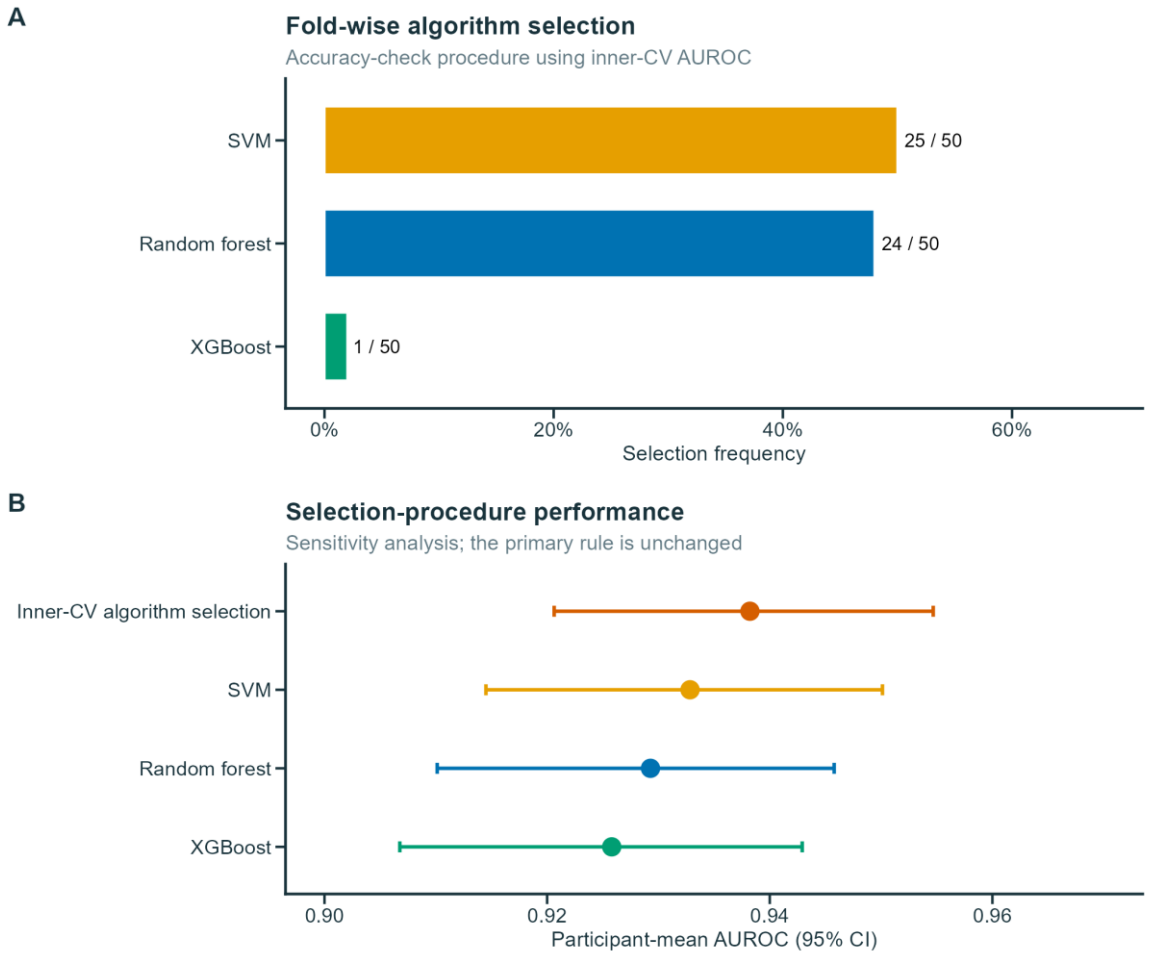
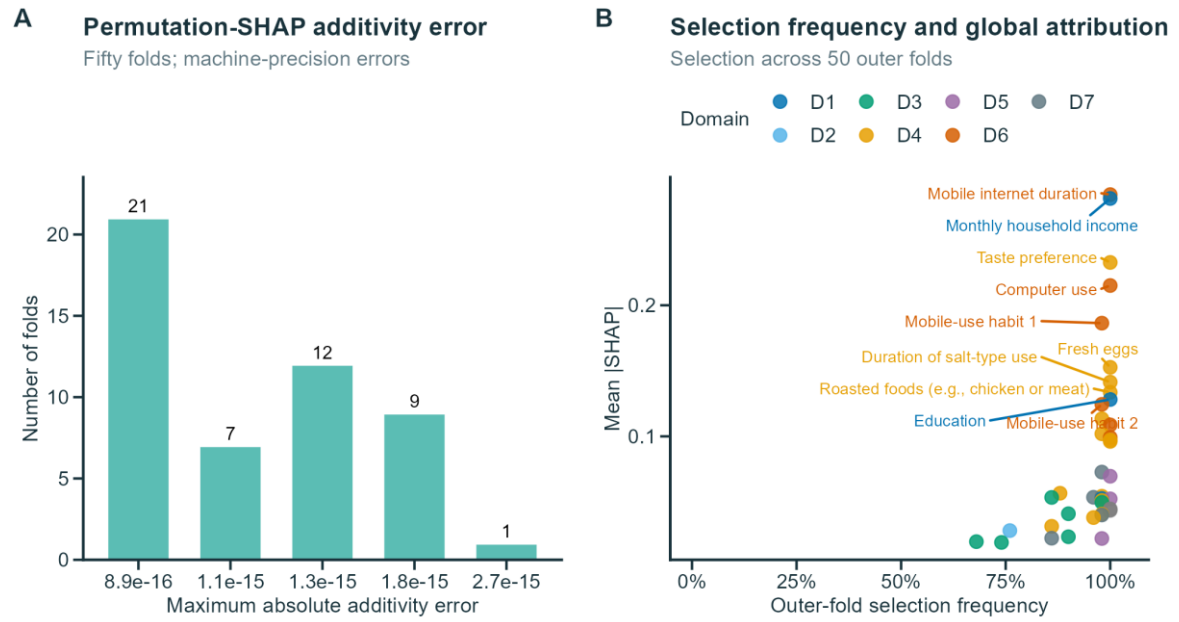
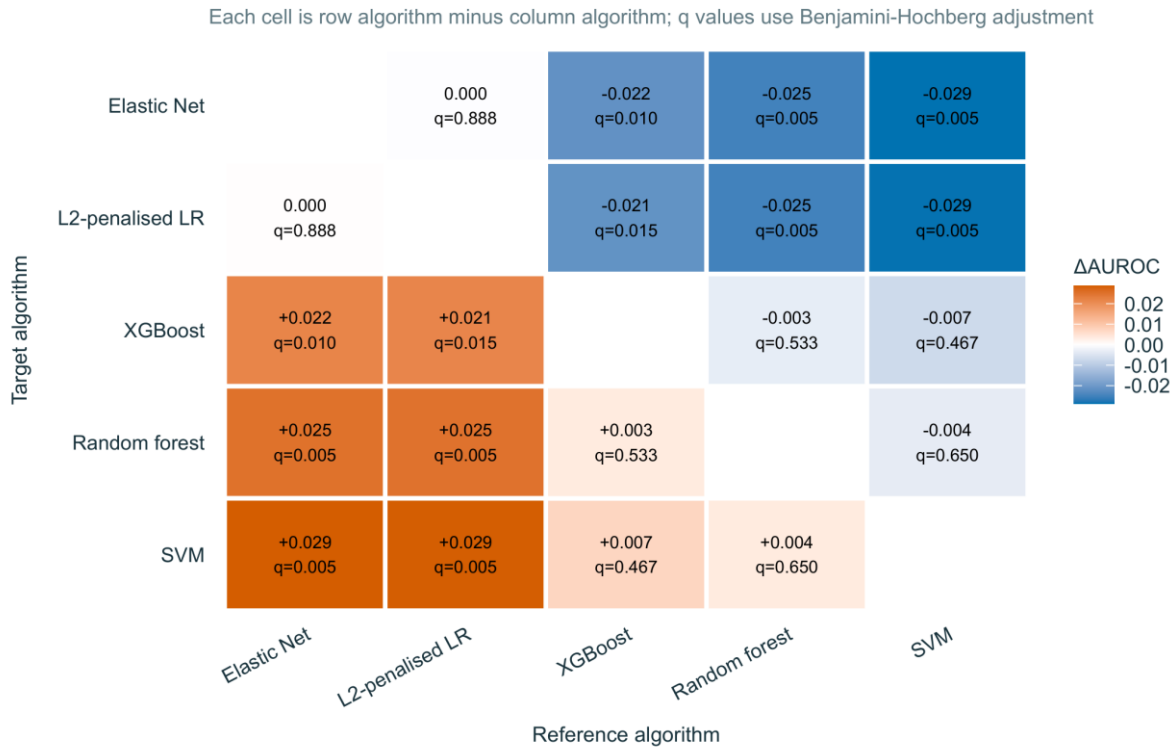


Fig. S11. Numerical quality of SHAP explanations and stability of variable selection



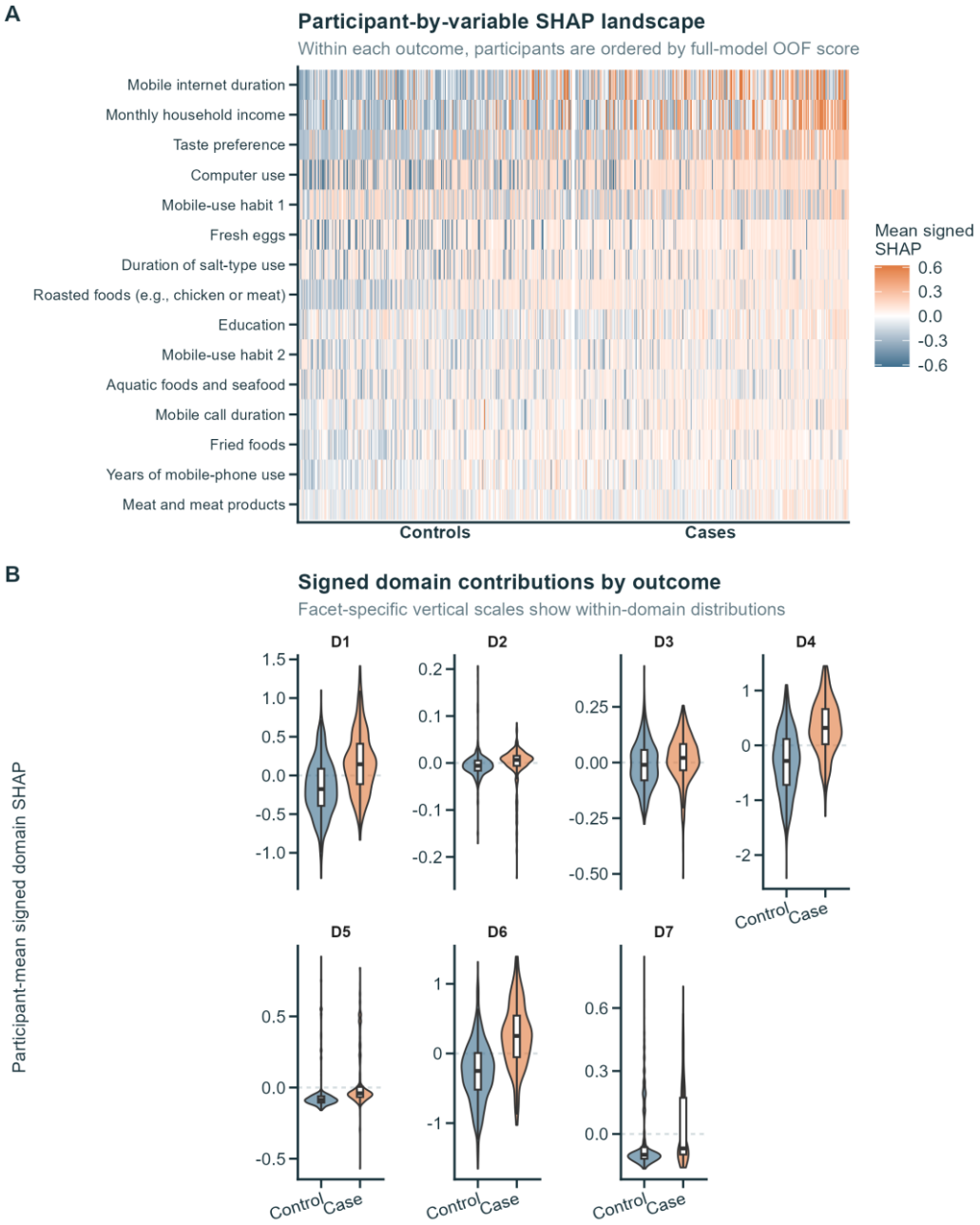
SHAP, SHapley Additive exPlanations.

Fig. S12. Pairwise AUROC differences among five candidate algorithms using predictions for the same participants



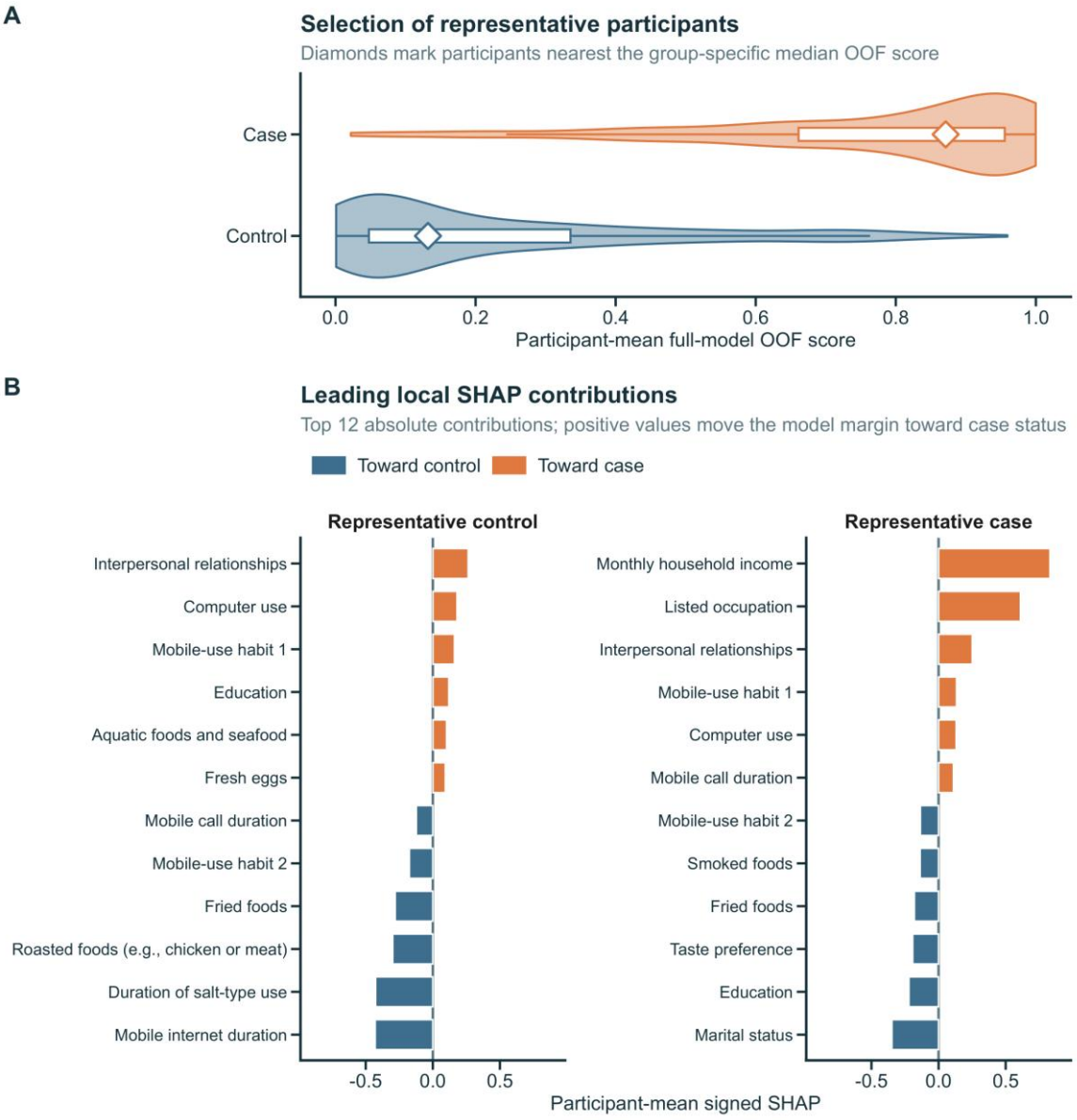
AUROC, area under the receiver operating characteristic curve.

Fig. S13. Overall participant-level pattern of SHAP attributions



SHAP, SHapley Additive exPlanations.

Fig. S14. Local SHAP attributions for representative case and control participants



SHAP, SHapley Additive exPlanations.