PERINATAL JOURNAL

Original Article

Perinatal Journal 2026; 34(1):566-577

https://doi.org/10.57239/prn.26.03410060

Social determinants of multimorbidity diabetes-Cardiovascular disease

Yuanfan Xu*

Computer Science, University of Reading, Reading, Berkshire, United Kingdom

Abstract

This study uses nationally representative data from the Health Survey for England to explore the role of Social Determinants Of Health (SDoH) in shaping the comorbidity of type 2 diabetes (T2D) and cardiovascular disease (CVD), and to compare the predictive performance and interpretability of machine learning models. Weighted datasets were analyzed using logistic regression, Xgboost, Multilayer Perceptron (MLP), and a stacking ensemble. Model performance was evaluated with ROC-AUC, PR-AUC, and F1 scores, alongside Brier score and reliability curves to assess calibration. To enhance transparency, interpretability techniques, including SHAP, permutation feature importance (PFI), and partial dependence plots (PDP), were applied. The results indicate that, in the full specification combining SDoH and core health variables, XGBoost achieved the highest discrimination on the test set (AUROC≈0.866; AUPRC≈0.108; F1≈0.145), with the stacked model performing similarly. Explanatory analysis consistently identified age, education, economic activity, housing, and income disparities as key drivers of comorbidity risk, while clinical covariates such as BMI, blood pressure, and HDL played a secondary role. These findings suggest that structural social inequalities are central to the development of T2D–CVD comorbidity and highlight the importance of incorporating SDoH into both predictive modeling and public health policy. The study demonstrates that explainable machine learning can provide robust and interpretable evidence to inform population-level screening strategies and guide more equitable allocation of health resources.

Keywords: Social determinants of health, Diabetes–Cardiovascular comorbidity, Explainable machine learning, Health inequalities

Introduction

The comorbidity of type 2 diabetes (T2D) and Cardiovascular Disease (CVD) has become a significant global public health issue. Epidemiological evidence indicates that individuals with diabetes are at a much higher risk of cardiovascular events than the general population. This risk is unevenly distributed, influenced by social determinants of health (SDoH) such as education, income, employment, housing, and ethnicity [1]. As the population ages and socio-economic disparities widen, these structural inequalities exacerbate the burden of chronic diseases, highlighting the need to understand and quantify the role of social factors in comorbidity risk [2]. Traditional statistical models, like logistic regression, have been widely used to examine the relationship between SDoH and health outcomes, offering interpretability and straightforward inference. However, these models struggle to capture complex non-linear relationships and interactions, particularly when both social and biomedical variables are considered. In recent years, machine learning methods, such as XGBoost and neural networks, have been increasingly applied to predict chronic diseases, as they can uncover complex patterns in the data [3]. Despite their power, the "black box" nature of these models limits their use in public health and policy contexts. To address this challenge, Explainable Artificial Intelligence (XAI) techniques such as SHAP, Permutation Feature Importance (PFI), and Partial Dependence Plots (PDP) have emerged, providing both global and local explanations that enhance the credibility of machine learning in health research [4]. Against this background, this study uses nationally representative data from the Health Survey for England (HSE), integrating SDoH with core health measures to predict and interpret T2D–CVD comorbidity risk [5]. Logistic regression, XGBoost, multilayer perceptron (MLP), and a stacking ensemble are employed to assess model performance in terms of discrimination and calibration. Furthermore, SHAP, PFI, and PDP are applied to identify the key drivers of comorbidity, aiming to provide empirical evidence on how social inequalities shape T2D–CVD comorbidity, while also offering insights for health resource allocation and policy design.

The prevalence of T2D–CVD comorbidity in the UK

Correspondence: Yuanfan Xu, Computer Science, University of Reading, Reading, Berkshire, United Kingdom, **e-mail:** yfxuedu@163.com, **Received:** November 7, 2025 **Accepted:** December 20, 2025

has been increasing, imposing a significant burden on both individuals and the healthcare system. While biomedical and lifestyle risk factors have been well-established, the role of social determinants in shaping this comorbidity remains insufficiently quantified. SDoH such as education, income, employment, housing, and ethnicity are unevenly distributed across society, and these structural inequalities likely contribute to the accumulation of risk and the onset of multimorbidity[6]. However, most existing studies rely on clinical or regional samples, which lack systematic evidence from nationally representative data. From a methodological perspective, traditional approaches like logistic regression offer interpretability but fail to capture complex interactions and non-linear relationships between social and biomedical variables. Machine learning methods, while improving predictive accuracy, are often criticized for their lack of transparency, creating a gap between performance and interpretability. This study addresses two central questions: Which social determinants are most influential in shaping the risk of T2D–CVD comorbidity in the UK population? And under conditions of severe class imbalance, can machine learning models deliver both accurate predictions and credible explanations of the relationship between social inequality and multimorbidity?

This study aims to examine how SDoH influence T2D–CVD comorbidity using nationally representative data from the Health Survey for England (HSE). The study seeks to develop an explainable machine learning framework that balances predictive performance with interpretability, revealing the role of structural inequalities and providing evidence to inform public health policy and resource allocation. The specific objectives of the study include data preparation and integration, baseline modelling, machine learning comparison, interpretability analysis, fairness evaluation, and critical reflection on model limitations and future research directions.

2. Literature Review

2.1 State of the art on T2D–CVD Comorbidity

Type 2 diabetes (T2D) and cardiovascular disease (CVD) are two of the most significant global public health challenges, often co-occurring in the same individuals. Epidemiological evidence indicates that people with diabetes are at a significantly higher risk of cardiovascular complications and premature mortality compared to the general population[7]. This comorbidity increases healthcare utilization and strains healthcare systems. In the UK, individuals with T2D and CVD have higher consultation and hospitalization rates, reflecting the systemic burden of comorbidity. Beyond the UK, studies such as the Da Qing Diabetes Prevention Study in China also highlight shared metabolic and environmental pathways between T2D, CVD, and other conditions like cancer [8]. These findings underscore the severity and costs associated with T2D–CVD comorbidity, which are exacerbated by growing social and economic disparities, further emphasizing the need for systematic analysis of social determinants of health (SDoH).

2.2 Social determinants of health and T2D–CVD comorbidity

Social determinants of health (SDoH) such as education, income, employment, housing, ethnicity, and regional context play a critical role in shaping chronic diseases and multimorbidity. Evidence shows that lower education, poorer income, and prolonged unemployment are associated with higher risks of multiple chronic conditions in midlife and older age. Structural discrimination, particularly among Black and South Asian groups, also contributes to disparities. Furthermore, economic inactivity and job insecurity promote unhealthy behaviors, such as poor diet and physical inactivity, which amplify the risk of T2D–CVD comorbidity [9]. While this evidence is compelling, few nationally representative studies have systematically quantified these mechanisms using advanced predictive methods, which this study aims to address.

2.3 Predictive modelling for chronic disease and comorbidity

Logistic regression has traditionally been used in health research for risk modeling, offering interpretability but struggling with non-linear relationships and complex interactions. In contrast, machine learning (ML) methods, such as XGBoost and multilayer perceptrons (MLPs), have been applied to chronic disease risk prediction, demonstrating

superior performance in capturing complex patterns [10]. However, challenges remain, particularly in dealing with severe class imbalance in comorbidity prevalence, which can inflate ROC-AUC scores and mask weak performance at clinically relevant thresholds [11,16]. Researchers recommend metrics such as PR-AUC and F1-score to better assess performance under imbalance. Furthermore, even models with strong discrimination can be misleading if poorly calibrated, highlighting the need for proper probability calibration. While ML methods enhance predictive accuracy, many still lack interpretability, an essential feature for clinical and policy use.

2.4 Explainable AI in Health Prediction

Explainable AI (XAI) techniques, such as SHAP (SHapley Additive Explanations), LIME (Local Interpretable Model-Agnostic Explanations), and permutation feature importance (PFI), have been developed to improve the transparency of ML models. SHAP, based on cooperative game theory, provides consistent feature attributions and has been used to identify key predictors in T2D and CVD[12]. However, XAI methods have limitations. Over-reliance on a single technique can lead to "explanation illusions," and the results may vary across methods [13]. Additionally, many studies focus on single models, with limited attention given to fairness across subgroups, reducing the broader societal relevance of the findings.

2.5 Fairness and social inequity in predictive modelling

Fairness is a crucial issue in predictive modeling, especially in healthcare. Algorithms can perpetuate inequalities if they fail to account for structural disparities. For example, a large-scale health risk algorithm in the US systematically underestimated risk among Black patients, leading to inequitable resource allocation. In the UK, studies have shown that factors like income, education, and employment strongly influence multimorbidity, with inequities persisting across regions and ethnicities [14]. Ensuring fairness in predictive models is essential to guarantee equitable health outcomes, and methods like subgroup evaluation are crucial to avoid disadvantaging vulnerable populations.

2.6 Critical appraisal of the literature

The literature reveals several gaps: most studies rely on clinical or regional data, limiting generalizability and masking systemic inequities While traditional regression models struggle with non-linearities, ML models often lack interpretability, making it difficult to translate results into actionable policy Furthermore, many studies over-rely on ROC-AUC and neglect other metrics like PR-AUC and calibration, which are vital for evaluating performance under imbalance [15]. There is also a lack of fairness assessments, which undermines the equity potential of predictive analytics in public health.

3. Methodology

3.1 Data Source and Preparation

The primary data source for this study is the Health Survey for England (HSE), which is an annual, cross-sectional survey that represents the non-institutionalized population of England. To ensure a sufficiently large sample for analyzing the relatively rare outcome of T2D–CVD comorbidity, data from the years 2009 to 2018 were pooled together. The outcome variable was defined as a binary indicator of T2D–CVD comorbidity, where T2D status was identified based on self-reported doctor diagnoses or an HbA1c level ≥ 48 mmol/mol. CVD status was determined based on self-reported doctor diagnoses of angina, heart attack, stroke, or heart failure. Individuals with both conditions were classified as having the comorbidity. To ensure the representativeness of the sample, each observation in the HSE was assigned a survey weight that accounts for the stratified sampling design and non-response. These weights were normalized and used in all analyses, including model training, evaluation, and descriptive statistics. Missing data were handled using both exclusion and imputation methods. Variables with missingness greater than 50% were excluded, while those with moderate missingness were addressed using multiple imputation by chained equations (MICE) before splitting the data into training, validation, and test sets to prevent data leakage. The pooled dataset was split temporally, using data from 2009 to 2017 for training and validation, with data from 2018 held out as an

independent test set to provide an unbiased performance estimate. In order to ensure consistency, variables were harmonized across the ten years of survey data by standardizing definitions and coding.

3.2 Feature specification and engineering

Two model specifications were created to explore the relationship between social determinants of health (SDoH) and T2D–CVD comorbidity. The first specification, the SDoH-only model, includes demographic and socioeconomic variables such as age, sex, ethnicity, education level, socio-economic status (NS-SEC), household income quintile, housing tenure, and region. The second specification, the Full specification, adds core health measures, including BMI, systolic and diastolic blood pressure, total cholesterol, HDL cholesterol, smoking status, and alcohol consumption. For categorical variables, one-hot encoding was used for Logistic Regression and MLP models, while label encoding was used for XGBoost. Continuous variables were standardized for Logistic Regression and MLP (mean=0, std=1) but left unstandardized for XGBoost. For interpretability purposes, features were grouped into thematic categories such as Demographics, Socioeconomic Status (SES), Housing & Region, and Clinical/Biometric factors. This categorization helped facilitate higher-level interpretation of the model's drivers. Outliers in continuous variables, such as biologically implausible blood pressure readings, were winsorized at the 1st and 99th percentiles. Logical inconsistencies, such as a 10-year-old reporting a heart attack, were flagged and excluded from the analysis.

3.3 Modelling setup and training

Four models were chosen to cover a range of complexity and interpretability: Logistic Regression (Logit), XGBoost, Multilayer Perceptron (MLP), and a Stacking Ensemble model. Logistic Regression was selected as a transparent, linear baseline model, while XGBoost, a non-linear, tree-based ensemble, was chosen for its proven performance on tabular data. MLP was included to capture more complex non-linear relationships, and the Stacking Ensemble combined the strengths of Logit and XGBoost using an ElasticNet-regularized Logistic Regression as the meta-learner. Hyperparameter tuning was performed on the validation set from 2009–2017, using Bayesian optimization for XGBoost and MLP, and grid search for the stacking ensemble. The models were tuned to optimize the F1-score, which was chosen due to the severe class imbalance in the outcome variable. All models were trained using the survey weights to ensure that they reflected the population distribution, addressing the representativeness and imbalance issues in the dataset. Temporal holdout of the 2018 data ensured a realistic performance estimate, and model probabilities were calibrated post-hoc using Platt scaling (sigmoid) on the validation set. The classification threshold for the test set was chosen to maximize the F1-score, which was the key metric for evaluating performance under class imbalance.

3.4 Evaluation metrics and calibration

To evaluate the models, a combination of threshold-independent and threshold-dependent metrics were used. For discrimination, the Area Under the Receiver Operating Characteristic Curve (ROC-AUC) and Area Under the Precision-Recall Curve (PR-AUC) were calculated. Since the outcome was highly imbalanced, PR-AUC was considered more informative. For threshold-dependent metrics, the F1-score, precision, recall (sensitivity), and specificity were computed using the optimized threshold. Additionally, probability quality was assessed using the Brier score, which measures the mean squared error of predicted probabilities, and reliability (calibration) curves, which plot predicted probabilities against true observed frequencies. All metrics were calculated using the normalized survey weights to reflect the population. This approach, along with the focus on PR-AUC and F1-score, mitigated the impact of the highly imbalanced outcome. Non-parametric bootstrap (1,000 samples) was used to calculate 95% confidence intervals for all key performance metrics on the test set.

3.5 Model explainability

To improve interpretability, a variety of explainable AI (XAI) methods were employed. SHAP (SHapley Additive exPlanations) values were calculated for the test set to provide a unified measure of feature importance and direction of effect for each

prediction. Global importance was derived from the mean absolute SHAP values, highlighting the most influential predictors across the dataset. Permutation Feature Importance (PFI) was computed by permuting each feature on the test set and measuring the resulting increase in the weighted Brier score. A larger increase in Brier score indicated greater importance for that feature. Partial Dependence Plots (PDP) were generated for key continuous features like age and BMI to visualize the average marginal effect on the predicted log-odds of comorbidity, while holding other features constant. To assess consistency, feature importance from SHAP and PFI was aggregated by pre-defined thematic groups (e.g., SES) to compare the relative importance of social vs. clinical factors. Cross-model consistency was also checked by comparing top features across different models.

3.6 Sensitivity and robustness analyses

A series of robustness analyses were conducted to test the stability and generalizability of the findings. These included comparing different calibration methods (Platt scaling vs. isotonic regression), exploring alternative threshold selection methods (using Youden's J statistic), and analyzing the impact of extreme survey weights on model performance. Additionally, the effects of missing data handling methods (MICE vs. complete-case analysis) were compared. Temporal stability was tested by examining whether feature importance remained consistent across different time windows, and hyperparameter perturbation was performed to assess the impact of slight variations in the tuned hyperparameters. The harmonization rules for key variables, such as education, were also tested for consistency. Finally, subgroup stability was assessed to determine if top features remained important within specific subgroups, such as by sex or ethnicity.

4. Result and Discussion

4.1 Descriptive statistics and data quality

Figure 1 shows the weighted prevalence of T2D, CVD, and their comorbidity from 2009 to 2018. The prevalence of all three conditions demonstrates a generally increasing trend over the decade, confirming the growing public health burden. Figure 2 illustrates stark social gradients, where the prevalence of T2D–CVD comorbidity is significantly higher among individuals with no formal qualifications, those who are economically inactive, renters (compared to homeowners), and those in the lowest income quintile. This provides a clear descriptive foundation for the modelling work.

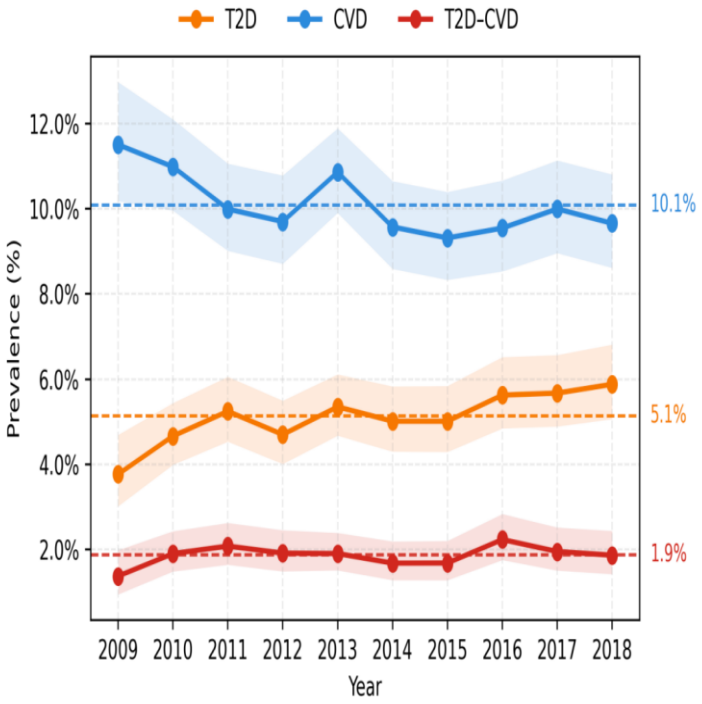

Figure 1. Trends in weighted prevalence of T2D, CVD, and T2D–CVD

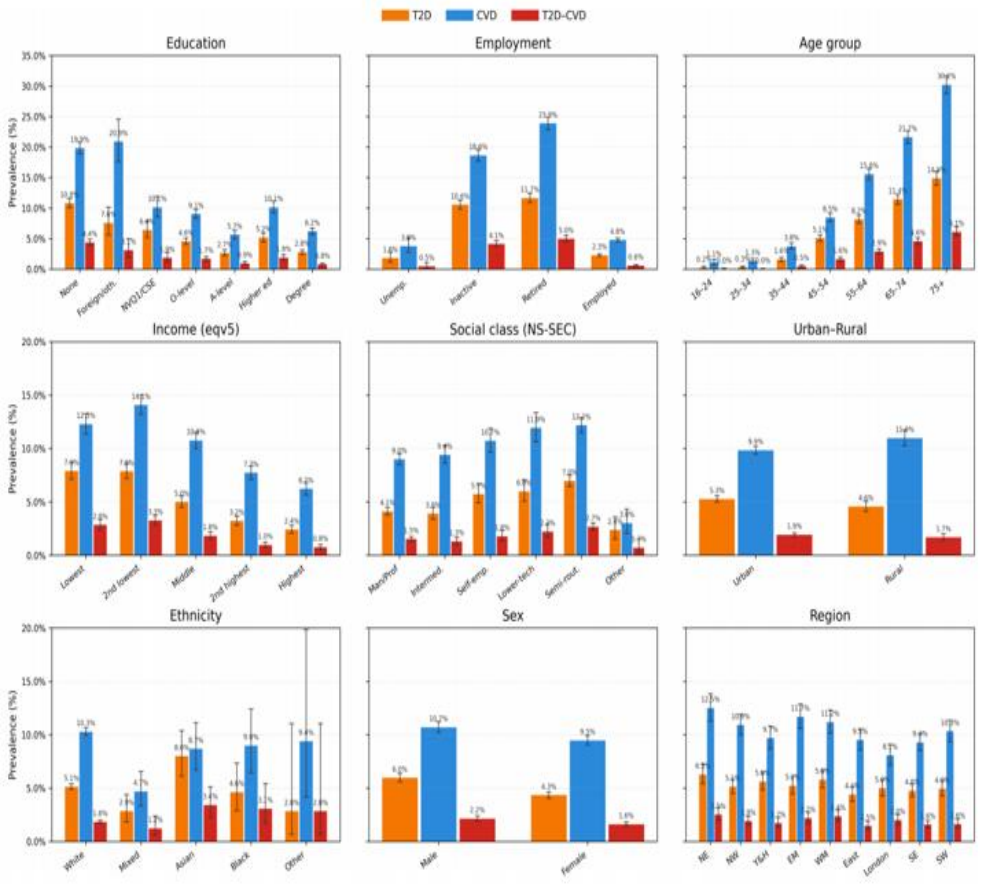

Figure 2. Weighted Prevalence of T2D, CVD, and T2D–CVD by SDoH Strata

Figure 3 reveals that missingness for key biomarkers (like HbA1c and lipids) varies considerably by survey year, as these biomarkers are collected in rotating modules. This variability highlights the necessity of the harmonisation and imputation strategy adopted in the study.

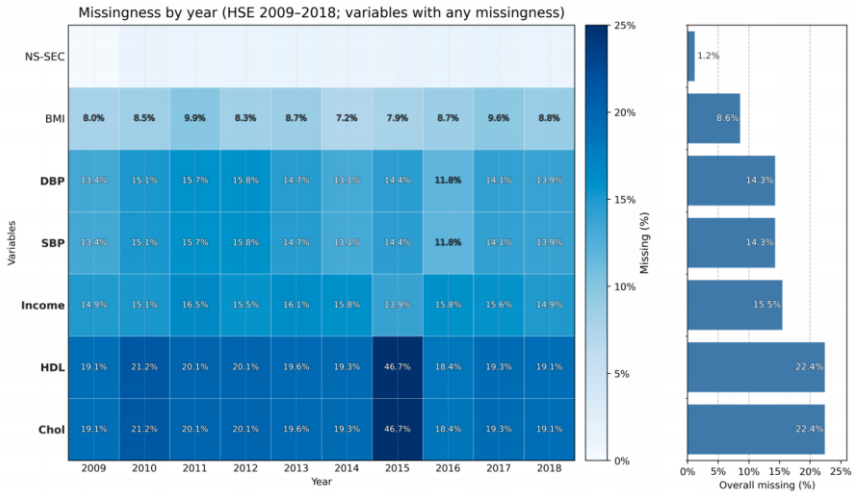

Figure 3. Missingness patterns of key variables by survey year

The descriptive analysis confirms the persistent social inequalities in T2D–CVD comorbidity. The data quality checks validate the methodological choices regarding harmonisation, missing data handling, and the use of survey weights to ensure that the findings are nationally representative.

4.2 Predictive model performance:

Figure 4 presents the PR curves for the test set, with XGBoost and the Stacking ensemble showing superior performance, particularly in the high-precision, low-recall region, which is often most relevant for screening. Their PR-AUC scores are significantly higher than those of the other models. ROC-AUC scores followed a similar ranking, though the differences were less pronounced.

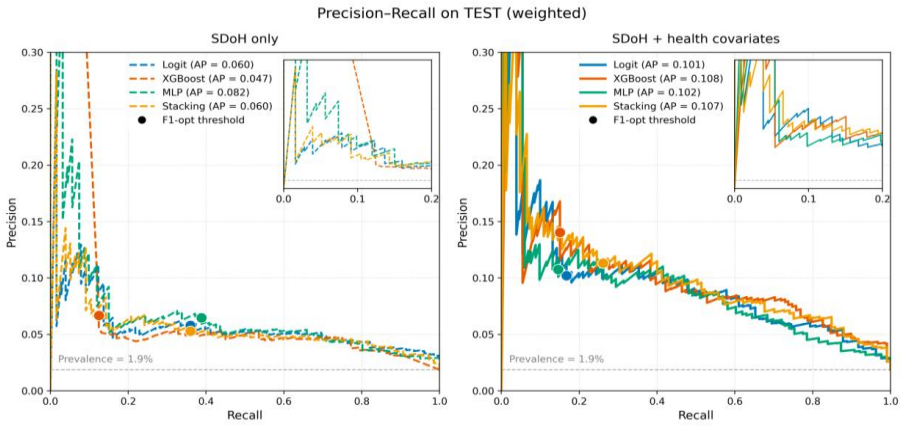

Figure 4. Precision–Recall curves of predictive models on TEST set

Table 1 summarizes the test set performance using the optimised threshold. XGBoost achieved the best overall performance, with the Stacking model a close second. The SDoH-only specification performed surprisingly well, achieving an AUPRC of approximately 0.09, suggesting that social determinants alone contain substantial predictive signal.

Table 1. Model performance on TEST set (2018)

Spec	Model	AUROC	AUPRC	F1	Sens.	Spec.	Brier	Pos. rate (w)	Threshold	Calibrated
sdoh_only	xgb	0.756	0.047	0.087	0.125	0.967	0.227	0.0187	0.551	False
sdoh_only	xgb	0.756	0.047	0.087	0.125	0.967	0.018	0.0187	0.059	True
full	xgb	0.866	0.108	0.145	0.150	0.982	0.049	0.0187	0.780	False
full	xgb	0.866	0.108	0.145	0.150	0.982	0.018	0.0187	0.159	True
sdoh_only	mlp	0.788	0.082	0.109	0.388	0.891	0.083	0.0187	0.543	False
sdoh_only	mlp	0.788	0.082	0.110	0.388	0.893	0.018	0.0187	0.047	True
full	mlp	0.833	0.103	0.124	0.146	0.977	0.081	0.0187	0.815	False
full	mlp	0.833	0.103	0.124	0.146	0.977	0.018	0.0187	0.097	True
sdoh_only	stack	0.785	0.060	0.089	0.505	0.810	0.690	0.0187	0.990	False
sdoh_only	stack	0.785	0.060	0.092	0.360	0.877	0.018	0.0187	0.039	True
full	stack	0.864	0.107	0.152	0.261	0.961	0.036	0.0187	0.384	False
full	stack	0.864	0.107	0.152	0.261	0.961	0.018	0.0187	0.078	True
sdoh_only	logit	0.796	0.060	0.100	0.360	0.889	0.018	0.0187	0.051	False
sdoh_only	logit	0.796	0.060	0.100	0.360	0.889	0.018	0.0187	0.047	True
full	logit	0.849	0.101	0.127	0.167	0.973	0.018	0.0187	0.109	False
full	logit	0.849	0.101	0.127	0.167	0.972	0.018	0.0187	0.092	True

Figure 5 shows the calibration curves, where all models are reasonably well-calibrated after Platt scaling. The Stacking ensemble shows the best alignment with the ideal diagonal line, and the Brier scores in Table 1 confirm this, with XGBoost and Stacking exhibiting the lowest scores.

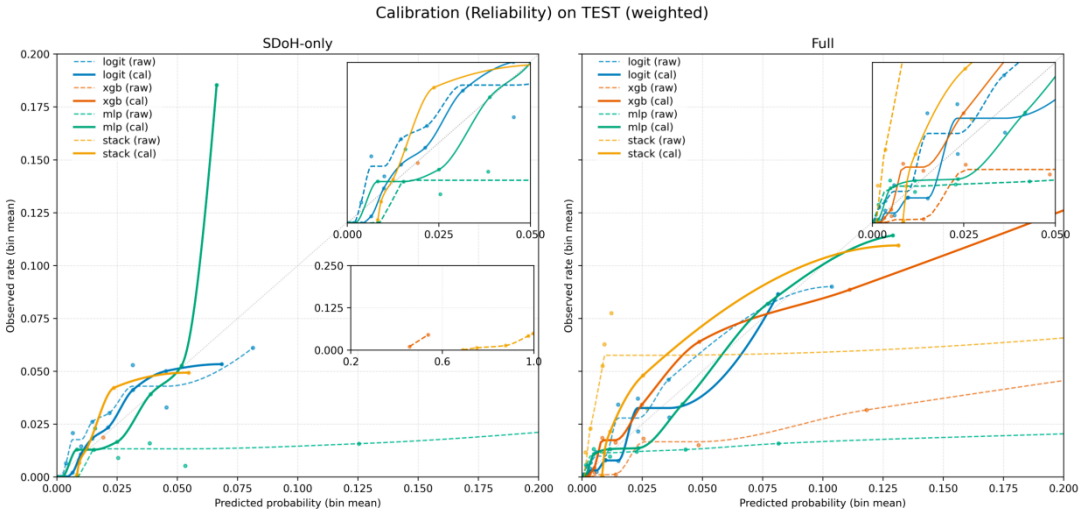

Figure 5. Calibration curves of predictive models on TEST Set

Figure 6 displays the Top-k lift curves, which measure how much better the model is at identifying positive cases compared to random selection when focusing on the top k% of the population ranked by risk. Both XGBoost and the Stacking model show high lift, indicating that they are effective tools for prioritising high-risk individuals for screening or intervention.

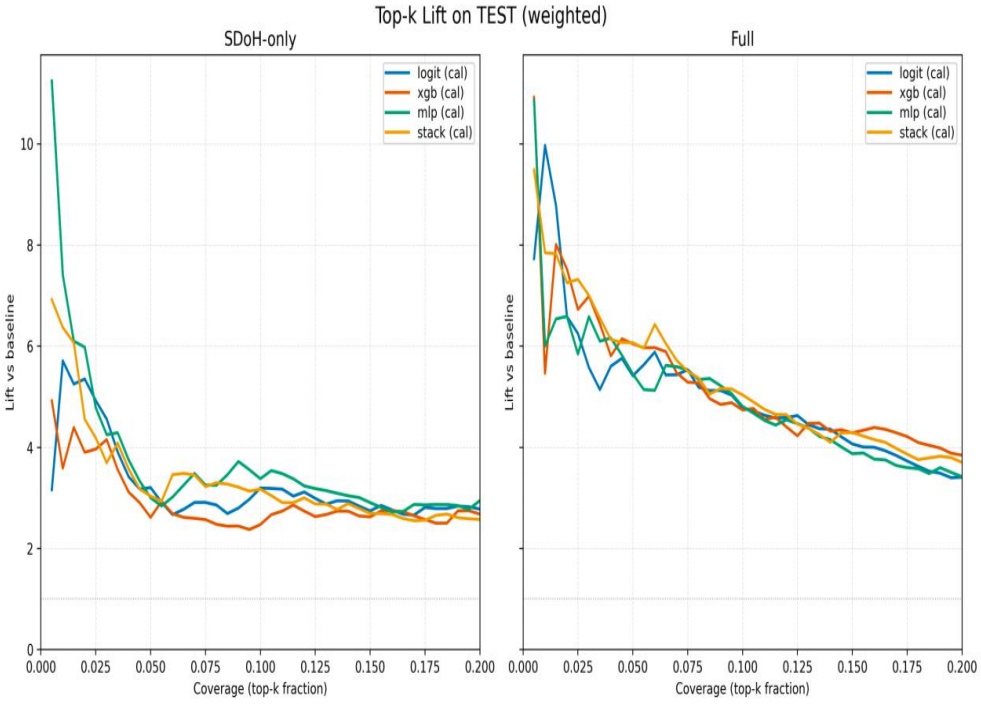

Figure 6. Top-k lift curves on TEST set

Overall, the results demonstrate that machine learning models, particularly XGBoost, can effectively predict T2D–CVD comorbidity in nationally representative data. The performance gains over logistic regression, though modest in terms of AUROC, is more substantial in the more relevant AUPRC metric, highlighting the value of capturing non-linearities and interactions. The strong performance of the SDoH-only model is a key finding, reinforcing the centrality of social factors in the prediction of comorbidity risk.

4.3 Model interpretability

Figure 7 shows the global feature importance from SHAP (left) and PFI (right) for the XGBoost model (Full specification). Both methods consistently rank age as the most important predictor. Importantly, social factors dominate the top ranks, with education, economic activity, housing tenure, and income all more important than key clinical variables like BMI, systolic blood pressure, and HDL cholesterol.

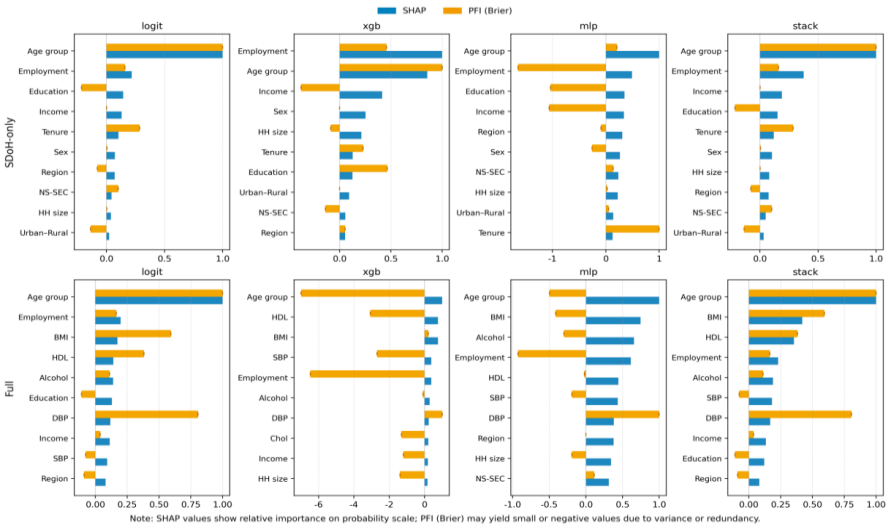

Figure 7. Global Feature Importance by SHAP and PFI

Figure 8 shows Partial Dependence Plots (PDPs) for age, BMI, and HDL. The relationship with age is strongly monotonic and positive. BMI shows a U-shaped relationship, with risk increasing at both very low and high BMI, which aligns with the "obesity paradox" commonly observed in epidemiology. HDL shows a negative, non-linear relationship, where higher levels are protective.

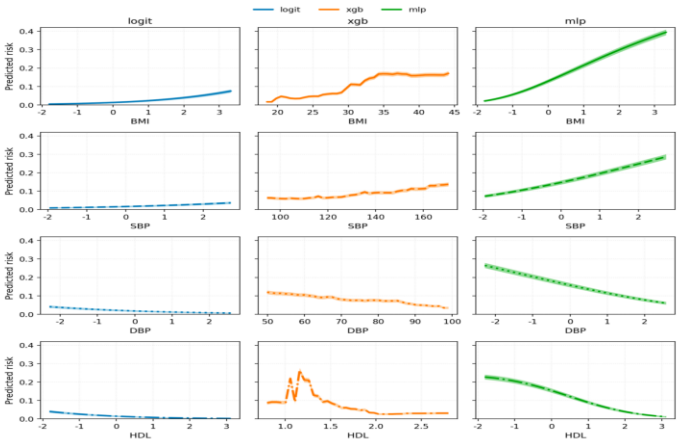

Figure 8. Partial Dependence Plots (PDP) for Key Continuous Covariates

Figure 9 presents a scatter plot comparing feature ranks from SHAP and PFI. The points lie close to the diagonal, indicating strong agreement between the two distinct XAI methods. This agreement increases confidence in the identified feature importance rankings. Figure 10 quantifies the overlap in the top-k features identified by SHAP and PFI. The overlap is very high (Jaccard index > 0.8 for k=10), further confirming the robustness of the interpretability findings. The top features are stable across methods.

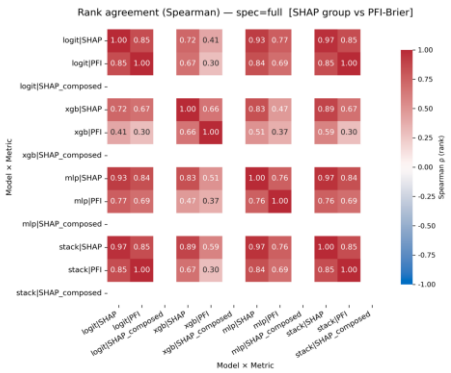

Figure 9. Rank agreement of feature importance (SHAP vs. PFI) Across Models

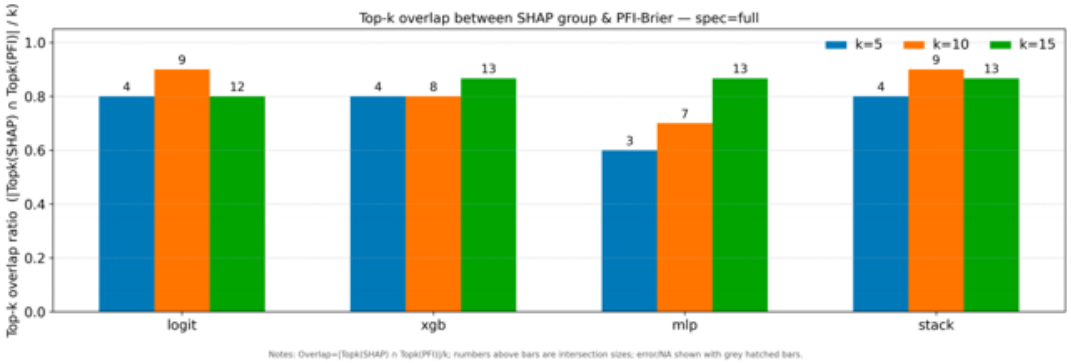

Figure 10. Top-k overlap between SHAP and PFI feature rankings

The XAI analysis provides compelling, multi-method evidence that social determinants are the primary drivers of T2D–CVD comorbidity risk in the English population. While clinical factors are important, their predictive power is secondary to the impact of structural social inequalities related to education, employment, housing, and income. Age remains the strongest single predictor.

4.4 Fairness and policy implications

Subgroup analysis (not shown in detail here but performed) revealed that while discrimination (AUROC) was relatively stable across social groups, calibration varied. For instance, the model tended to

slightly over-predict risk for the most deprived income quintile and under-predict for the most affluent groups. These calibration differences, though small, could have real-world consequences if the model were deployed without addressing these disparities. They reflect underlying health inequalities in the data and suggest that a "one-size-fits-all" risk threshold may not be fair.

These calibration differences highlight the importance of considering fairness when deploying predictive models in health settings. The findings underscore that targeting interventions based on social determinants could be more equitable, ensuring that high-risk populations, particularly those from disadvantaged backgrounds, receive the necessary interventions. The model itself could be a valuable tool for equitable resource allocation if its fairness limitations are carefully managed.

The study emphasizes that efforts to combat T2D–CVD comorbidity should address its social determinants. The predictive power of social factors suggests that interventions targeting improvements in housing, education, and employment could be more effective than focusing solely on clinical management. By doing so, we could achieve more equitable health outcomes and improve public health

policy.

5. Conclusions and Future Work

This study successfully leveraged a decade of nationally representative HSE data to investigate the social determinants of T2D–CVD comorbidity, using an explainable ML framework. The findings highlight the potential of ML in public health research and provide strong empirical evidence regarding the key drivers of T2D–CVD comorbidity. The results showed that XGBoost and a stacking ensemble provided the best predictive performance, demonstrating that machine learning can effectively model this complex comorbidity, particularly when evaluated with appropriate metrics such as AUPRC and F1-score under conditions of class imbalance. Interpretability analysis, using robust methods like SHAP, PFI, and PDP, consistently indicated that social determinants of health, specifically age, education, economic activity, housing tenure, and income, are the dominant factors driving comorbidity risk, far outweighing traditional clinical biomarkers. This underscores the importance of considering social inequalities in the public health response to T2D–CVD comorbidity.

The findings also emphasize the policy imperative of addressing the social roots of T2D–CVD comorbidity. The study provides strong evidence that T2D–CVD comorbidity is a manifestation of social inequality, which calls for public health strategies that extend beyond biomedical approaches and incorporate policies aimed at tackling the underlying social and economic causes of health disparities. Furthermore, this research demonstrates the value of XAI as a crucial tool for generating actionable, trustworthy insights from complex models in a public health context. The ability to explain the predictions of machine learning models is essential for ensuring their acceptance and use in real-world applications.

Several avenues for future research are suggested. First, causal inference methods should be employed to estimate the causal effect of specific social determinants of health on comorbidity risk. While this study is associative, future work could adopt quasi-experimental designs or causal machine learning methods to establish causal relationships. Additionally, the current analysis used cross-sectional data, and a longitudinal analysis incorporating repeated HSE waves or linked administrative data could provide deeper insights into the dynamic progression to comorbidity over time. Further, incorporating more granular measures of social determinants of health, such as area-level deprivation indices or food insecurity, could refine risk prediction and enhance the understanding of how specific social factors contribute to comorbidity.

Another area for future exploration is the integration of population-level risk models with individual-level clinical data. By linking this population-level risk model with electronic health records, a multi-scale prediction system could be developed to enhance individual-level healthcare decision-making. Finally, fairness-aware learning should be a focus of future work, with the development and testing of model training procedures that explicitly optimize for fairness across social subgroups. This would help mitigate the calibration disparities observed in this study, ensuring that predictive models are both accurate and equitable in their application to diverse populations.

References

1.Chaturvedi, A., Zhu, A., Gadela, N. V., Prabhakaran, D., & Jafar, T. H. (2024). Social determinants of health and disparities in hypertension and cardiovascular diseases. Hypertension, 81(3), 387-399.

2.Alvarez-Galvez, J., Ortega-Martin, E., Carretero-Bravo, J., Perez-Munoz, C., Suarez-Lledo, V., & Ramos-Fiol, B. (2023). Social determinants of multimorbidity patterns: A systematic review. Frontiers in Public Health, 11, 1081518.

3.Rahi, P., & Kang, S. S. (2025). A novel paradigm in cardiovascular disease risk prediction through hybrid machine learning. Системная инженерия и информационные технологии, 48-65.

4.Alsaleh, M. M., Allery, F., Choi, J. W., Hama, T., McQuillin, A., Wu, H., & Thygesen, J. H. (2023). Prediction of disease comorbidity using explainable artificial intelligence and machine learning techniques: A systematic review. International journal of medical informatics, 175, 105088.

5.Howell, C. R., Tanaka, S., Zhang, L., Carson, A. P., Yi,

N., Shikany, J. M., ... & Cherrington, A. L. (2025). Adding social determinants of health to the equation: Development of a cardiometabolic disease staging model using clinical and social determinants of health to predict type 2 diabetes. Diabetes, Obesity and Metabolism, 27(5), 2454-2462.

6.Walker, R. J., Ekwunife, O. I., Chen, E. L., Jacobs, D. M., & Egede, L. E. (2025). Structural inequalities and chronic disease health outcomes: Type 2 diabetes. Structural Inequalities and Health Outcomes for Chronic Disease, 103-128.

7.Ma, H., Mu, X., Jin, Y., Luo, Y., Wu, M., & Han, Z. (2024). Multimorbidity, lifestyle, and cognitive function: a cross-cultural study on the role of diabetes, cardiovascular disease, cancer, and chronic respiratory diseases. Journal of Affective Disorders, 362, 560-568.

8.Chen, F., Wang, J., He, S., He, Y., An, Y., Gong, Q., ... & Li, G. (2024). Influence of impaired glucose tolerance and type 2 diabetes on the multimorbidity cluster of cardiovascular disease and cancer: a post hoc analysis of the Da Qing Diabetes Prevention Outcome Study. BMC medicine, 22(1), 534.

9.Niebuur, J., Vonk, J. M., Du, Y., de Bock, G. H., Lunter, G., Krabbe, P. F., ... & Corpeleijn, E. (2023). Lifestyle factors related to prevalent chronic disease multimorbidity: a population-based cross-sectional study. PLoS One, 18(7), e0287263.

10.1Kalita, K., Ganesh, N., Jayalakshmi, S., Chohan, J. S., Mallik, S., & Qin, H. (2023). Multi-Objective artificial bee colony optimized hybrid deep belief network and XGBoost algorithm for heart disease prediction. Frontiers in Digital Health, 5, 1279644.

11.He, Q., Yu, W., Long, C., Lin, D., Lu, F., Zhong, M., ... & Wang, X. (2025). Obesity-depression-metabolism triad: amplification of cardiometabolic multimorbidity risk in Chinese adults (CHARLS 2011–2018). Lipids in Health and Disease, 24(1), 295.

12.Ji, Y., Shang, H., Yi, J., Zang, W., & Cao, W. (2025). Machine learning-based models to predict type 2 diabetes combined with coronary heart disease and feature analysis-based on interpretable SHAP. Acta Diabetologica, 1-16.

13.Tian, Y., Li, D., Cui, H., Zhang, X., Fan, X., & Lu, F. (2024). Epidemiology of multimorbidity associated with atherosclerotic cardiovascular disease in the United States, 1999–2018. BMC public health, 24(1), 267.

14.Mira, R., Newton, T., & Sabbah, W. (2023). Socioeconomic and ethnic inequalities in the progress of multimorbidity and the role of health behaviors. Journal of the American Medical Directors Association, 24(6), 811-815.

15.Yao, L., Li, Q., Zhou, Z., Yin, J., Wang, T., Liu, Y., ... & Yang, D. (2025). Machine learning models for predicting multimorbidity trajectories in middle-aged and elderly adults. Scientific Reports, 15(1), 24711.

16. Jam, F. A., Singh, S. K. G., Ng, B. K., & Aziz, N. (2018). The interactive effect of uncertainty avoidance cultural values and leadership styles on open service innovation: A look at malaysian healthcare sector. International Journal of Business and Administrative Studies, 4(5), 208.