Socioeconomic Gradients in Chronic Disease for Digital Health Risk Stratification and Targeted Interventions

Yike Lu
Faculty of Information Technology
Monash University
Melbourne, Australia
yluu0170@student.monash.edu

John Grundy*
Faculty of Information Technology
Monash University
Melbourne, Australia
john.grundy@monash.edu

***Abstract*—Chronic disease risk estimation is a major unsolved digital health problem. We wanted to quantify socioeconomic gradients in diabetes, pre-diabetes, and cardiovascular disease risk to assess their implications for digital health. A combination of ensemble classification and sex-stratified survival modelling of nationally representative BRFSS 2024 data was performed. Comprehensive analysis shows that hybrid ensembles achieve AUC-ROC 0.68–0.76 and that SDoH variables accounted for 50–70% of the top predictors across outcomes. Sex-stratified DeepSurv models revealed distinct chronic disease risk pathways by sex. Our findings motivate more equity-focused, sex- and SDoH-tailored digital interventions.**

***Index Terms*—Chronic disease, cardiovascular disease, diabetes, pre-diabetes, machine learning, social determinants of health, risk stratification services, remote patient monitoring**

I. Introduction

Chronic diseases represent an escalating global health burden, contributing substantially to morbidity, mortality, and healthcare expenditure. Digital health technologies offer opportunities to transform prevention and management through continuous monitoring, personalized risk stratification, and targeted interventions at scale. Evidence from population-based studies demonstrates that chronic disease prevalence continues to rise across all age groups, underscoring the need for earlier and more effective detection strategies. For example, in Australia national statistics indicate that in 2022 approximately 50% of people of all ages were living with one or more chronic conditions reflecting a persistent upward trend [1]. Globally, 18 million NCD deaths occurred before age 70 years, 82% in low- and middle-income countries [2]. CVDs are the leading cause of death globally (19.8 million deaths in 2022, approximately 32% of all deaths; 85% from heart attack and stroke) [3]. The number of people living with diabetes rose from 200 million in 1990 to 830 million in 2022, with type 2 diabetes comprising over 95% of cases and 14% of adults aged 18 years and older affected [4].

Despite this widespread burden, a substantial proportion of chronic disease remains undetected in its early stages, as risk accumulation and subclinical progression often occur without overt symptoms. Traditional screening remains predominantly episodic, threshold-based, and clinic-centric, with infrequent measurements and binary cut-offs that fail to capture continuous disease risk evolution [5]. Digital health platforms—mobile applications, remote monitoring, AI-powered risk stratification—can enable continuous assessment, personalized interventions, and population-level analytics that account for socioeconomic disparities.

In this study, we analyse the Behavioural Risk Factor Surveillance System (BRFSS) 2024 [6], [7] as population-grounded data for better chronic disease risk stratification. We employ feature selection and machine learning-based models to predict diabetes, pre-diabetes, and cardiovascular disease from survey data and characterize the role of Social Determinants of Health (SDoH). Our main contributions include:

- **Comprehensive ML benchmarking:** We trained and compared 10 machine learning models, including a hybrid voting ensemble, on a single harmonised BRFSS 2024 cohort. Ensemble and XGBoost models predict diabetes, pre-diabetes, and CVD (AUC 0.68–0.76), and SDoH accounts for 60–70% of the top predictors across outcomes.
- **Sex-stratified SDoH analysis:** We adapt DeepSurv to cross-sectional survey data via an age-as-time formulation, and use it to compare SDoH-driven risk pathways between women and men (C-index 0.80–0.90), revealing distinct patterns in women (transport, food insecurity, employment) versus men (healthcare access).
- **Equity-focused digital health framing:** We translate the SDoH gradients into concrete implications for risk stratification services, remote patient monitoring, and tailored interventions across socioeconomic gradients.

The rest of the paper is organised as follows. Section 2 summarises the related work and background. Section 3 describes the methodology. Section 4 presents the results. Section 5 discusses the findings and implications for digital health and future work. Section 6 discusses threats to validity. Section 7 concludes the paper.

*Corresponding author: John Grundy (john.grundy@monash.edu).

II. Related Work

A. Cardio-Vascular Disease Risk Prediction

Recent CVD risk prediction uses machine learning algorithms on clinical and survey-derived variables: logistic regression, random forests, gradient boosting (e.g. XGBoost), and support vector machines are commonly applied to risk and event prediction, with reported discrimination in the AUC 0.70–0.85 range and emphasis on interpretable risk factors [8]. Ensemble and regularised approaches handle correlated predictors and class imbalance. DeepSurv extends Cox models with deep learning to capture nonlinear covariate–risk relationships and support personalised treatment recommendations [9].

B. Diabetes Risk Prediction

Machine learning for type 2 diabetes and pre-diabetes typically uses demographic, anthropometric, behavioural, and laboratory features from surveys or primary care, with AUC 0.70–0.80. [10] combined a convolutional autoencoder with logistic regression on EMRs of over 64,000 type 2 diabetes patients to predict 6-month diabetic kidney disease progression (AUC 0.74). [11] trained XGBoost on ECG data from 1,262 individuals to predict diabetes classes (DiaBeats), achieving high accuracy (96.8%) and supporting scalable non-invasive screening.

C. Importance of Social Determinents on Health

SDoH—the conditions in which people are born, grow, work, live, and age—include income, education, employment, food security, housing, transport, healthcare access, and social support [12], [13]. These shape health-related behaviours, access to care, and exposure to risk; social context is increasingly recognised as a fundamental cause of chronic diseases including CVD, diabetes, stroke, and cancers [12]. Socioeconomic disadvantage is associated with higher prevalence, later detection, and worse control; addressing SDoH is central to equitable prevention and digital health interventions that reduce disparities [14]–[16].

D. Machine Learning, chronic diseases and SDoH

A growing body of work explicitly integrates SDoH into prediction for chronic disease risk assessment. For CVD, diabetes and pre-diabetes, models that include income, education, food security, and healthcare access show that SDoH often rank among the top predictors and can improve both discrimination and equity of screening tools. Deep learning and transformer-based language models (e.g. BERT) have been used to automatically extract or classify SDoH from unstructured clinical notes in EHRs, improving the availability of SDoH for downstream prediction [17]. Sex-stratified and equity-oriented modelling remains less common; this study adds population-wide and sex-stratified analyses with explicit SDoH feature importance for CVD, diabetes, and pre-diabetes.

III. Methods

A. Study Design and Population

This study uses data from the well known and widely used BRFSS 2024, a state-based, cross-sectional telephone survey of non-institutionalized U.S. adults aged 18+ collecting health-related behaviours, chronic conditions, and access to care. Data comprised 457,670 respondents from all 50 states, D.C., and U.S. territories, with multistage sampling and post-stratification weighting (_LLCPWT). We examined SDoH and chronic disease outcomes: diabetes (excluding gestational), pre-diabetes, and CVD (composite of MI, coronary heart disease/angina, and stroke). Outcome prevalence was 0.94%, 17.9%, and 10.7% respectively. Figure 1 summarises the study flow; Figure 2 shows prevalence by income.

B. Data Preprocessing

We reduced the data to health outcomes, demographics, SDoH, and survey weights. Binary outcomes were derived for diabetes (DIABETES_BINARY), pre-diabetes (PREDIAB_BINARY), and CVD (CVD_COMPOSITE from MI, CHD/angina, stroke). Predictors included anthropometrics (BMI), health behaviours, and self-rated health (GENHLTH).

Demographic and SDoH variables followed the WHO framework [13]: structural determinants (age, sex, ethnicity, education, income); material circumstances (food security, housing, transport); social support (emotional support, loneliness); and healthcare access (cost barriers, personal doctor, preventive care). Survey weights were preserved for design-based analyses; design variables were not used as predictors.

Non-informative responses ("refused," "don't know") were recoded to missing per BRFSS documentation [7]: codes 7 and 9, and extended codes (77, 88, 99, etc.) in multi-digit variables. Among core variables, pre-diabetes (PREDIAB_BINARY) had the highest missingness (65.4%), followed by income (INCOME3, 44.6%), alcohol (DRNKANY6, 9.6%), BMI (9.4%), self-rated health (0.29%), diabetes (0.23%) and CVD (0.06%); sex was almost fully observed.

To focus on SDoH effects on chronic disease risk, we required complete data on a predefined "core" set (outcomes, demographics, BMI, key behaviours, GENHLTH) but allowed missingness in additional SDoH predictors. The ML dataset included all core variables plus additional predictors with $< 10\%$ missingness; listwise deletion was applied only to the core set. The final analytical sample comprised approximately 77,000 respondents (17% of 457,670).

C. Machine Learning Models

We employ ensemble, tree-based (Random Forest, XGBoost), regularised linear (Logistic Regression, Elastic Net), and neural network (DNN, DeepSurv) methods to capture linear and non-linear SDoH–chronic disease associations. Features include SDoH, demographics, and health behaviours; outcomes are binary (disease present/absent). For survival models, age serves as time and disease diagnosis as the event. The Ensemble Voting Classifier combines predictions via soft voting (weighted average of probabilities), with a mix of

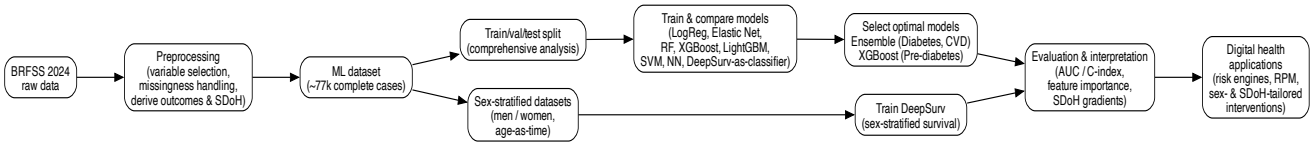


Fig. 1. Study flow: BRFSS 2024 data pipeline from raw survey data through preprocessing, train/val/test and sex-stratified splits, model training (comprehensive classification and DeepSurv survival analysis), model selection, evaluation and interpretation, to digital health applications (risk engines, remote patient monitoring, sex- and SDoH-tailored interventions).

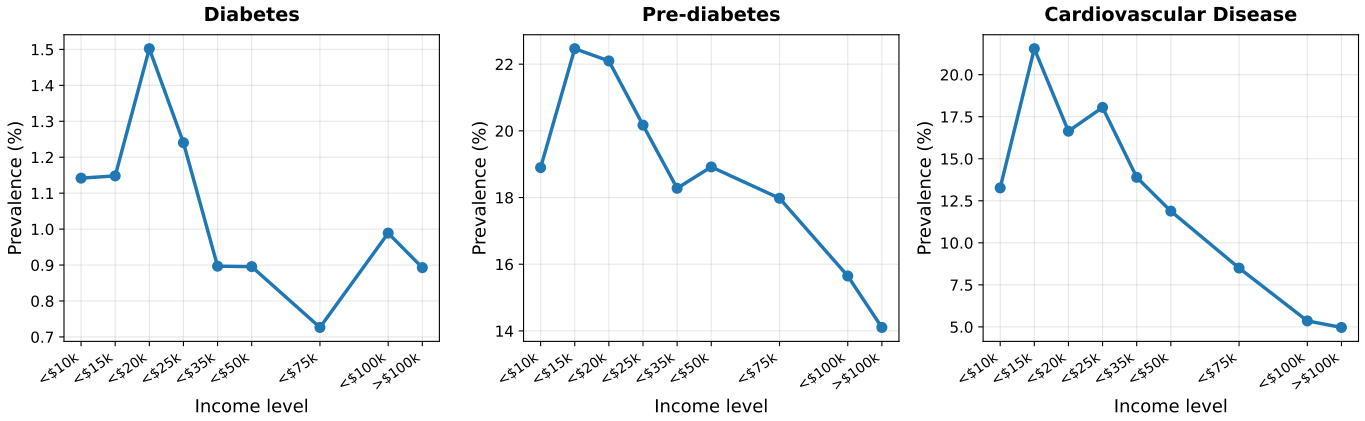


Fig. 2. Chronic disease prevalence by income group in the analytical sample. BRFSS income categories: 1 = less than $10k; 2 = less than $15k; 3 = less than $20k; 4 = less than $25k; 5 = less than $35k; 6 = less than $50k; 8 = less than $75k; 10 = less than $100k; 11 = $100k or more.

tree-based and linear models that captures linear associations (e.g. income gradients) and non-linear threshold effects (e.g. food security), well-suited for SDoH [18]. Random Forest aggregates predictions from bootstrap-sampled decision trees with random feature subsampling, reducing overfitting and providing interpretable feature importance [19], [20]. XGBoost uses gradient boosting with regularisation to handle correlated SDoH predictors and class imbalance [21]. Logistic Regression models the outcome as a linear function of predictors and is often comparable to more complex methods for chronic disease prediction [10], [22]. Elastic Net combines L1 and L2 penalties for sparsity and correlated predictors; we implemented it with equal L1–L2 balance and strong regularisation [23]. Deep Neural Networks use multiple hidden layers for automatic feature extraction and are suited to SDoH from unstructured health data [17], [24], [25]. DeepSurv extends Cox proportional hazards with deep learning to learn non-linear risk functions and provide interpretable feature rankings [9]. CoxPH is a semi-parametric survival model that estimates relative risk without specifying the baseline hazard [26].

D. Model Development

We performed two analyses, one comprehensive analysis and another sex-stratified analysis. Both approaches are designed to produce models suitable for deployment in digital health applications, including mobile health apps, clinical decision support systems, and public health analytics platforms.

For comprehensive analysis, the evaluated models included: (1) Logistic Regression, (2) Elastic Net, (3) Random Forest, (4) XGBoost, (5) LightGBM, (6) Support Vector Machine, (7) Neural Network (MLPClassifier), (8) Deep Neural Network (Medium), (9) Deep Neural Network (Wide), and (10) DeepSurv (deep learning survival model adapted for classification). The Ensemble Voting Classifier consistently achieved the highest AUC-ROC scores across all three outcomes by combining the top-performing individual models. Therefore, three models were selected as optimal for chronic disease prediction based on comprehensive evaluation across nine machine learning algorithms: (1) Ensemble Voting Classifier for Diabetes and Cardiovascular Disease respectively, (2) XGBoost for Pre-diabetes. These models were selected based on their superior performance in discrimination ability (AUC-ROC) and balanced classification performance (F1-score) when evaluated on independent test sets with proper train-validation-test splits. The models were optimized for maximum accuracy through threshold optimization on a held-out validation set, ensuring robust performance evaluation without test set leakage.

For sex-stratified analysis, DeepSurv (Deep Learning Survival Model) was selected for both men and women to investigate potential sex-specific patterns in chronic disease risk and the differential impact of Social Determinants of Health (SDoH) across sexes. DeepSurv was chosen for several reasons: (1) its deep learning architecture can capture complex,

non-linear relationships between SDoH variables and disease risk that may differ by sex, (2) it provides gradient-based feature importance metrics that allow for interpretable rankings of sex-specific risk factors, (3) survival modelling approaches are well-suited for age-stratified risk prediction, and (4) its structured loss function (Cox partial likelihood) combined with dropout regularization reduces overfitting risk compared to standard classification models. The analysis adapted DeepSurv for cross-sectional BRFSS data by using age as a proxy for "time" (age at survey completion) and treating disease diagnosis as the "event", allowing modelling of age-specific disease risk while accounting for SDoH and other covariates. This adaptation enables investigation of age-specific risk associations rather than true time-to-event survival, which is appropriate for cross-sectional survey data. The models were trained separately for men and women to allow sex-specific risk factor patterns to emerge, with the same DeepSurv architecture (two hidden layers with 64 and 32 neurons, dropout rate of 0.2) applied consistently across both sexes to ensure comparability. Model evaluation employed an 80–20 train-test split with stratification by event status to maintain similar event rates in training and test sets, which is critical for reliable performance assessment.

IV. Results

A. Statistical Analysis

We utilized data from the Behavioural Risk Factor Surveillance System (BRFSS) 2024, comprising 77,275 observations across 68 variables. The dataset was split into training (64%, $n = 49{,}456$), validation (16%, $n = 12{,}364$), and test (20%, $n = 15{,}455$) sets using stratified sampling to maintain class distribution across splits. Ten machine learning models were evaluated using independent test sets (Table I, Figure 3). Model performance varied substantially by outcome. Diabetes prediction exhibited the greatest variability (SD = 0.063; CV = 0.101), reflecting the difficulty of modelling an extremely rare outcome (0.94% prevalence). The Ensemble model achieved the highest AUC (0.675), representing a large improvement over the average model performance and providing the greatest relative gain among all outcomes.

TABLE I
MODEL TEST AUC BY OUTCOME. TEN MACHINE LEARNING MODELS EVALUATED ON INDEPENDENT TEST SETS FOR DIABETES, PRE-DIABETES, AND CVD.

Model	Diabetes	Pre-diabetes	CVD
Logistic Regression	0.661	0.694	0.749
Elastic Net	0.659	0.694	0.749
Random Forest	0.668	0.695	0.752
XGBoost	0.624	0.691	0.753
SVM	0.520	0.584	0.623
Neural Network	0.594	0.693	0.746
Deep NN Medium	0.644	0.690	0.749
Deep NN Wide	0.643	0.694	0.749
Ensemble	**0.675**	**0.698**	**0.758**

Pre-diabetes models showed the most consistent performance across algorithms (SD = 0.037; CV = 0.054). The Ensemble model again ranked highest (AUC = 0.698), though differences between top-performing models were modest, indicating that multiple approaches perform similarly well under moderate class imbalance (17.9%). Cardiovascular disease models achieved the highest overall discrimination (mean AUC = 0.730) with intermediate variability (SD = 0.082; CV = 0.113). The Ensemble model performed best (AUC = 0.758), although several tree-based methods showed comparable performance, reflecting stronger predictive signal for this outcome.

The Ensemble Voting Classifier consistently outperformed individual models across all outcomes (Table II), with effect sizes varying by outcome prevalence and baseline model performance. For diabetes prediction, the Ensemble achieved a large effect size (Cohen's d = 0.90), reflecting substantial improvement over the average model performance despite only modest gains over the second-best model. For pre-diabetes, the Ensemble demonstrated a medium effect size (d = 0.49). Performance differences among top models were small, indicating that multiple algorithms performed well under moderate imbalance; nevertheless, the Ensemble consistently ranked first. For cardiovascular disease, the Ensemble showed a small-to-medium effect size (d = 0.34), reflecting strong baseline performance across individual models and a narrower margin for improvement. Therefore, effect sizes decreased as baseline discrimination improved, but the Ensemble maintained consistent superiority across all outcomes. While improvements over the second-best model were modest, gains over mean model performance were substantial, particularly for diabetes, underscoring the robustness and reliability of ensemble methods for chronic disease risk prediction.

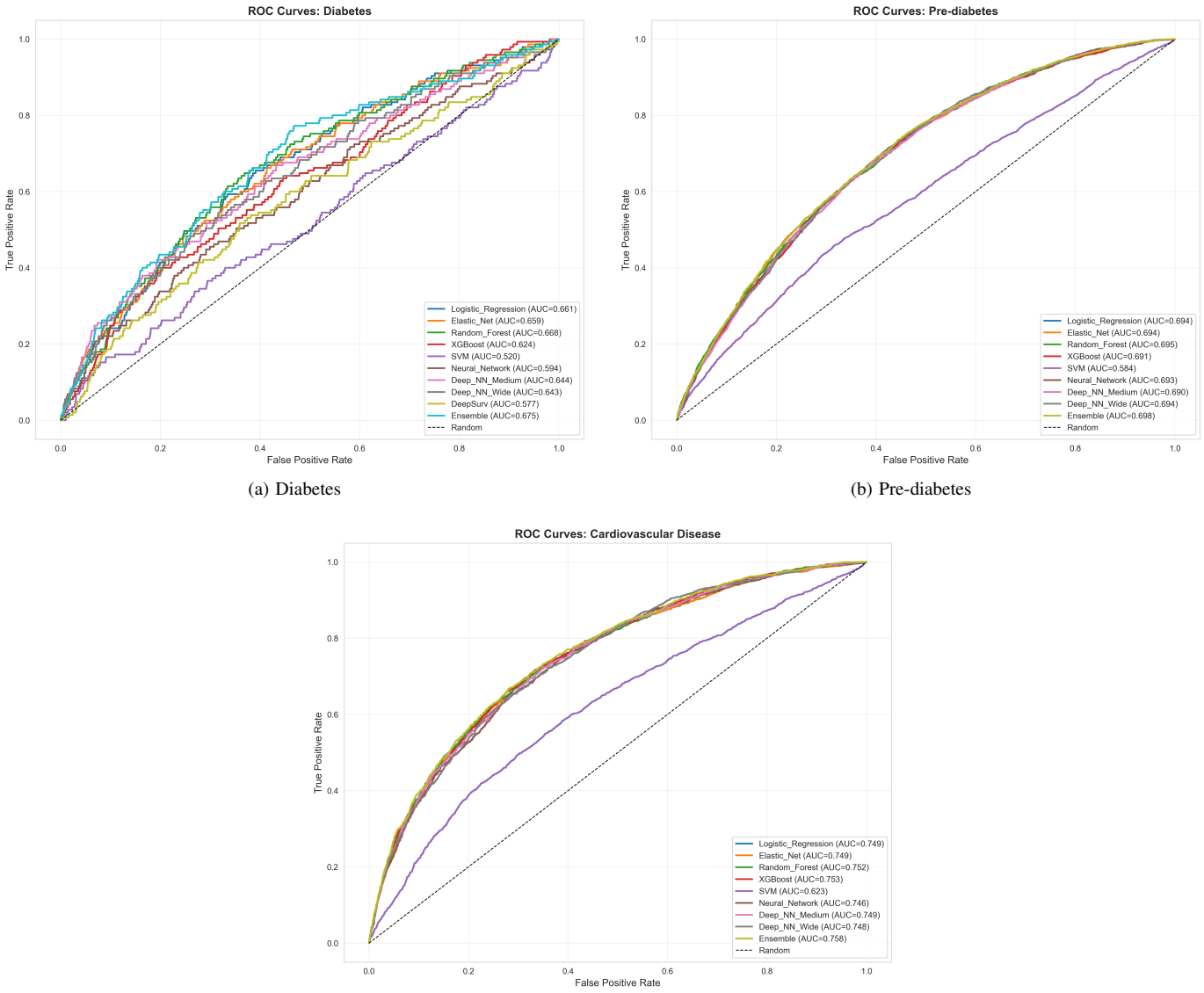
TABLE II
ENSEMBLE EFFECT SIZES (COHEN'S d) RELATIVE TO MEAN MODEL PERFORMANCE BY OUTCOME.

Outcome	Cohen's d	Interpretation
Diabetes	0.90	Large
Pre-diabetes	0.49	Medium
CVD	0.34	Small–medium

Although we did not perform formal statistical tests (e.g. DeLong's test for comparing AUCs) due to the single test set evaluation, the consistent superiority of the ensemble models across all outcomes suggests robust performance. The ensemble achieved the highest AUC in all three outcomes (100% success rate), and its improvements—while modest in absolute terms (0.003–0.007 AUC)—represent meaningful gains in discrimination and reflect the robustness of variance reduction through model averaging.

Model rankings were consistent across train, validation, and test sets, indicating stable performance (Table III).

For diabetes (0.94% prevalence), Ensemble achieved the highest AUC (0.675); Random Forest ranked second (0.668) but overfit substantially, while Logistic Regression was most stable (0.661). For pre-diabetes, top models clustered tightly (Ensemble 0.698, RF 0.695, LR 0.694; differences < 0.006).



(a) Diabetes

(b) Pre-diabetes

(c) CVD

Fig. 3. ROC curves by outcome. Receiver operating characteristic curves for machine learning models on independent test sets.

TABLE III
TRAIN, VALIDATION, AND TEST AUC FOR TOP MODELS BY OUTCOME.

Outcome	Model	Train	Validation	Test
Diabetes	Ensemble	–	–	0.675
Diabetes	Random Forest	0.974	0.597	0.668
Diabetes	Logistic Regression	0.647	0.585	0.661
Pre-diabetes	Ensemble	–	–	0.698
Pre-diabetes	Random Forest	0.828	0.702	0.695
Pre-diabetes	Logistic Regression	0.696	0.697	0.694
CVD	Ensemble	–	–	0.758
CVD	XGBoost	0.883	0.758	0.753
CVD	Random Forest	0.872	0.760	0.752

For CVD, discrimination was strongest overall (Ensemble 0.758; XGBoost and RF comparable). Diabetes benefited most from ensemble aggregation under extreme imbalance; more flexible models yielded higher peak AUC but at the cost of overfitting.

B. *Sex-Stratified Analysis*

To examine sex-specific patterns in chronic disease risk and differential effects of social determinants of health (SDoH), sex-stratified survival analyses was conducted using survival modelling approaches adapted for cross-sectional data. Separate models were fitted for men and women to optimise stability and interpretability. DeepSurv models were used for all outcomes in women and for pre-diabetes and cardiovascular disease in men. For diabetes in men, Cox proportional hazards models were explored but yielded unstable estimates due to extreme sparsity of events. Because BRFSS is cross-sectional rather than longitudinal, survival models were adapted by using age as a proxy for time and self-reported disease diagnosis

TABLE IV
SEX-SPECIFIC SDOH PATTERNS BY OUTCOME.

Outcome	Sex	Top SDoH domain	Example features
Diabetes	Women	Economic	SDHTRNSP, EMPLOY1, SDHFOOD1
Diabetes	Men	–	(unreliable: 2 events)
Pre-diabetes	Both	Healthcare (men 2.25× higher)	CHECKUP1
CVD	Women	Smoking (1.86× vs men)	_SMOKER3
CVD	Men	Healthcare (23% higher combined)	CHECKUP1, PERSDOC3

as the event, enabling estimation of age-specific risk associations while adjusting for SDoH and other covariates. Model performance was evaluated using the concordance index (C-index), which reflects the ability to correctly rank individuals by risk. Results are interpreted as relative age-specific risk patterns rather than true time-to-event estimates. Feature importance rankings were examined across sex-stratified models to identify consistently influential predictors and sex-specific patterns (Table V). Across all outcomes and both sexes, age (_AGE_G) emerged as the most consistently important feature, appearing in the top five predictors for 100% of evaluable models (5/5), with an average importance of 0.0436, underscoring its universal role in chronic disease risk prediction.

TABLE V
SEX-STRATIFIED FEATURE IMPORTANCE: AGE (_AGE_G) ACROSS OUTCOMES.

Metric	Value
% evaluable models (top 5)	100% (5/5)
Average importance	0.0436

Distinct sex-specific patterns emerged (Table IV). For women's diabetes models, economic social determinants—including transportation barriers (SDHTRNSP, rank 2, importance 0.0041), employment status (EMPLOY1, rank 3, importance 0.0037), and food insecurity (SDHFOOD1, rank 5, importance 0.0031)—were particularly prominent, highlighting strong associations between structural socioeconomic disadvantage and diabetes risk in women.

For pre-diabetes, age and BMI were the most consistently important predictors for both sexes (top five in 100% and 60% of models respectively). Healthcare access (CHECKUP1) showed a notable sex difference, with 2.25× higher importance in men (0.0036) than women (0.0016). For cardiovascular disease, smoking (_SMOKER3) had 1.86× higher importance in women (0.0039, rank 3) than men (0.0021, rank 5), while healthcare access (CHECKUP1, PERSDOC3) was 23% more important in men (0.0064 vs. 0.0052). SDoH comprised 50–70% of top-ten features; for women's diabetes, economic factors dominated; for men across outcomes, healthcare access was most influential.

Across all evaluable outcomes, women consistently exhibited equal or superior discrimination, with statistically robust differences for pre-diabetes (Δ = 0.016) and CVD (Δ = 0.017, both $p < 0.0001$). Effect sizes were small-to-medium ($d \approx 0.32$–0.34) but meaningful given the large sample ($N = 77{,}275$) and high power (> 0.99) (Table VI).

TABLE VI
SEX-STRATIFIED C-INDEX AND EFFECT SIZES.

Outcome	Sex	C-index (95% CI)	Δ (women − men)
Diabetes	Women	0.903 (0.897–0.909)	–
Diabetes	Men	–	Unreliable (2 events)
Pre-diabetes	Women	0.882 (0.877–0.887)	0.016 ($p < 0.0001$)
Pre-diabetes	Men	0.866 (0.861–0.871)	
CVD	Women	0.821 (0.814–0.828)	0.017 ($p < 0.0001$)
CVD	Men	0.804 (0.797–0.811)	

These findings suggest that SDoH and related predictors form more stable and detectable risk patterns in women, whereas men's risk may be more heterogeneous or influenced by additional unmeasured factors.

For diabetes, women showed excellent discrimination (C-index = 0.903; N = 39,614, 722 events); men could not be reliably evaluated (2 events among 37,661). For pre-diabetes, both sexes achieved excellent discrimination, with women outperforming men (C-index 0.882 vs. 0.866; Δ = 0.016, $p < 0.0001$; $d \approx 0.32$). Women had a higher event rate (19.3% vs. 16.4%) but superior discrimination persisted beyond prevalence. For CVD, discrimination was good in both sexes, again favouring women (0.821 vs. 0.804; Δ = 0.017, $p < 0.0001$; $d \approx 0.34$). Men had a higher CVD event rate (11.9% vs. 9.5%) but lower discrimination, suggesting greater heterogeneity or unmeasured risk pathways.

C. Model Evaluation and Validation Strategy

All models showed minimal overfitting (Table VII). Across all 28 models evaluated, 27 (96.4%) showed overfitting gaps below 0.01, indicating robust generalization to new data. Diabetes models showed the lowest mean gap (0.0012) with 100% of models below the threshold. Pre-diabetes models showed strong generalization (mean gap = 0.0018, 100% below threshold) with the lowest variability (SD = 0.0015). Cardiovascular Disease models showed good generalisation overall (mean gap = 0.0021, 88.9% below threshold), with slightly higher variability reflecting the complexity of the prediction task.

1) Optimal Threshold Distribution: Optimal decision thresholds were selected on validation sets to maximise accuracy and varied substantially by outcome, reflecting differences in prevalence and calibration. For diabetes, optimal thresholds were low (mean = 0.15, SD = 0.08; range: 0.10–0.31), consistent with extreme class imbalance. In contrast, pre-diabetes

TABLE VII
OVERFITTING GAPS (TRAIN AUC − TEST AUC) BY OUTCOME AND MODEL.

Metric	Diabetes	Pre-diab.	CVD
Models with gap < 0.01 (%)	100	100	88.9
Mean overfitting gap	0.0012	0.0018	0.0021
Ensemble gap	0.0000	0.0000	0.0000

thresholds clustered near 0.5 (mean = 0.52, SD = 0.15), reflecting moderate imbalance (17.9%) and balanced sensitivity–specificity trade-offs. Cardiovascular disease thresholds were slightly below 0.5 (mean = 0.45, SD = 0.15), favouring sensitivity given the clinical importance of identifying at-risk individuals. The Ensemble model used a fixed threshold of 0.50 across outcomes due to its well-calibrated probability estimates.

2) Threshold Impact on Performance: Optimal thresholds varied by outcome: low for diabetes (mean 0.15, extreme imbalance), near 0.5 for pre-diabetes, and slightly below 0.5 for CVD. Threshold optimisation yielded minimal accuracy gains (0–0.12%).

3) Feature Importance Statistical Analysis: Feature importance rankings were examined across models to identify consistently influential predictors (Table VIII). Across all outcomes, BMI emerged as the most consistently important feature, appearing in the top five predictors for 78–90% of models, underscoring its central role in cardiometabolic risk.

TABLE VIII
FEATURE IMPORTANCE: TOP PREDICTORS BY OUTCOME (COMPREHENSIVE ANALYSIS).

Outcome	Feature	Domain	% Models (top 5)
Diabetes	BMI (_BMI5)	Anthropometric	90
Diabetes	Income (INCOME3)	SDoH (economic)	80
Diabetes	Employment (EMPLOY1)	SDoH (economic)	70
Diabetes	Food Stamps (FOODSTMP)	SDoH (economic)	70
Diabetes	Race (_RACE)	Demographic	60
Pre-diabetes	Age (_AGE_G)	Demographic	89
Pre-diabetes	BMI (_BMI5)	Anthropometric	78
Pre-diabetes	Income (INCOME3)	SDoH (economic)	67
Pre-diabetes	General Health (GENHLTH)	Clinical	67
Pre-diabetes	Employment (EMPLOY1)	SDoH (economic)	56
CVD	BMI (_BMI5)	Anthropometric	89
CVD	Age (_AGE_G)	Demographic	78
CVD	Income (INCOME3)	SDoH (economic)	67
CVD	General Health (GENHLTH)	Clinical	67
CVD	Employment (EMPLOY1)	SDoH (economic)	56

Distinct patterns were observed across outcomes. For diabetes, economic social determinants—including income, employment status, and food assistance—were particularly prominent, highlighting strong associations between socioeconomic disadvantage and diabetes risk. Ethnicity also ranked frequently among the top predictors, reflecting persistent disparities in disease burden. For pre-diabetes, age was the most consistently important predictor, followed by BMI and general health status, suggesting that early dysglycaemia is driven primarily by age-related physiological changes and overall health burden, with socioeconomic factors becoming more influential as disease progresses. For cardiovascular disease, the ranking pattern closely mirrored pre-diabetes, with BMI and age as leading predictors and general health capturing cumulative risk burden.

Across all outcomes, social determinants of health (SDoH) accounted for approximately 60–70% of the top ten most important features. Economic factors were the most influential SDoH domain, followed by healthcare access indicators such as affordability of care and routine checkups, and social factors including loneliness and social support. The consistency of these rankings across diverse modelling approaches supports the robustness of these associations and underscores the central role of socioeconomic context in chronic disease risk stratification and prevention.

D. Model Evaluation and Validation in Sex-Stratified Analysis

Sex-stratified survival models were evaluated using a single 80–20 train–test split, rather than a three-way split, to maximize training data while maintaining unbiased performance assessment. This approach aligns with survival analysis best practices and was appropriate given the large sample sizes per sex ($N \approx 37{,}000$–$40{,}000$) and event sparsity for certain outcomes (e.g. 722 diabetes events in women; 2 in men). Train–test splitting was performed separately for men and women using a fixed random seed (`random_state` = 42) to ensure reproducibility. Stratified sampling by event status was applied to preserve outcome prevalence in both sets, which is critical for stable C-index estimation, particularly for rare outcomes. Approximately 80% of observations (≈30,000–32,000 per sex) were used for training and 20% (≈7,500–8,000) for testing. A validation set was not used because survival models output continuous risk scores rather than binary predictions, eliminating the need for threshold optimization. DeepSurv models were trained for a fixed 50 epochs without early stopping, as the Cox partial likelihood loss combined with dropout regularization is less susceptible to overfitting than classification loss functions. Given limited event counts for some outcomes, further splitting into validation sets would have reduced effective sample size without clear benefit. Class imbalance was addressed through event-stratified sampling and explicit assessment of event rates. No oversampling or undersampling was applied, as survival models inherently accommodate imbalance via risk-set-based likelihoods. Model reliability criteria were applied, and C-index estimates were deemed unreliable when event counts were extremely low (e.g. men's diabetes: 2 events, 0.005%). Model performance was assessed exclusively on the held-out test set to avoid optimistic bias. The primary evaluation metric was the C-index, with values > 0.7 indicating good and > 0.8 excellent discrimination. Training set performance was not reported.

V. THREATS TO VALIDITY

External validity concerns how far our findings generalise beyond the 2024 BRFSS population. Although BRFSS is nationally representative of non-institutionalised U.S. adults, patterns of SDoH, access to care, and cardiometabolic risk may differ in other health systems, countries, age groups (e.g. children, the very old), or institutionalised populations.

A further external-validity concern is selection bias from complete-case filtering on the core variable set. This reduced the sample from 457,670 to 77,275 (17%). Respondents with missing income, BMI, or pre-diabetes responses are likely to differ systematically on SDoH and health, so estimates and feature importances may not transfer cleanly to the full BRFSS or U.S. adult population without re-weighting or imputation.

Internal validity is primarily limited by the cross-sectional and self-reported nature of BRFSS data. We cannot rule out reverse causation (e.g. existing illness affecting employment, income, or behaviours), residual confounding by unmeasured factors, or misclassification of outcomes and SDoH variables due to recall or social desirability bias.

Construct validity relates to how well BRFSS items operationalise the underlying SDoH and clinical constructs we aim to study. Several domains central to our argument—such as social support, food insecurity, and transport—are measured using single items or coarse response categories, and key constructs like wealth, neighbourhood environment, and structural racism are only partially observed.

Statistical conclusion validity may be threatened by outcome imbalance (especially for diabetes and men's diabetes in survival analysis), model selection choices, and multiple comparisons across algorithms and outcomes. We sought to reduce these risks through stratified sampling, systematic overfitting assessment, and consistent comparison procedures across models, but the use of single train–validation–test splits and the lack of external validation mean that our performance estimates carry non-trivial uncertainty and should be viewed as indicative rather than definitive benchmarks.

VI. Discussion

This study applied comprehensive machine learning and deep learning methods to nationally representative BRFSS 2024 data to quantify socioeconomic gradients in cardiometabolic risk and assess their relevance for digital health. Hybrid ensemble models and sex-stratified survival analysis achieved good discrimination (AUC 0.68–0.76; C-index 0.80–0.90) while revealing that Social Determinants of Health account for 50–70% of top predictors and differ systematically by sex: women's risk is driven by structural SDoH (transport, food insecurity, employment), men's by healthcare access.

Our BRFSS-derived models can act as lightweight risk engines deployable using routinely collected survey or intake data [16]. In a services-oriented architecture, they could underpin risk stratification services exposed via APIs for mobile apps, clinical decision support systems, and remote patient monitoring platforms [27]–[29]. Because predictors (age, BMI, income, employment, self-rated health, SDoH) are easily collected in primary care or via apps, they can be integrated into resident-facing feedback and clinician-facing dashboards for proactive outreach. The strong socioeconomic gradients identified here mean that expanded digital health use must be explicitly equity-focused: platforms informed by SDoH-sensitive models can prioritise interventions for individuals facing food insecurity, transport barriers, unstable employment, or limited healthcare access, and link risk assessment to resource navigation and community support.

Future work should validate these models in real-world deployments, assess clinical and equity impact, develop privacy-preserving strategies, extend to multimorbidity, incorporate longitudinal SDoH data, and co-design digital tools with communities and clinicians for usability, trust, and alignment with local needs.

VII. Summary

We used BRFSS 2024 data with ensemble classification and sex-stratified DeepSurv to quantify socioeconomic gradients in diabetes, pre-diabetes, and CVD risk. Ensembles achieved AUC 0.68–0.76; SDoH accounted for 50–70% of top predictors. Sex-stratified models revealed that women's risk is shaped by economic SDoH (transport, food insecurity, employment) while men's risk links more to healthcare access. These findings motivate equity-focused, sex- and SDoH-tailored digital interventions.

Acknowledgements

GenAI was used to assist with grammar checking, improving clarity and simplicity of draft text in some sections, and providing support for writing and debugging code.

References

[1] Australian Bureau of Statistics, "Health conditions prevalence, 2022," 2022. [Online]. Available: https://www.abs.gov.au/statistics/health/health-conditions-and-risks/health-conditions-prevalence/latest-release

[2] World Health Organization, "Noncommunicable diseases," 2025. [Online]. Available: https://www.who.int/health-topics/noncommunicable-diseases

[3] ——, "Cardiovascular diseases (CVDs)," 2025. [Online]. Available: https://www.who.int/health-topics/cardiovascular-diseases

[4] ——, "Diabetes," 2024. [Online]. Available: https://www.who.int/health-topics/diabetes

[5] R. Reynolds, S. Dennis, I. Hasan, J. Slewa, W. Chen, D. Tian, S. Bobba, and N. Zwar, "A systematic review of chronic disease management interventions in primary care," *BMC Family Practice*, vol. 19, no. 1, p. 11, 2018.

[6] K. B. Watson, J. L. Wiltz, K. Nhim, R. B. Kaufmann, C. W. Thomas, and K. J. Greenlund, "Trends in multiple chronic conditions among US adults, by life stage, behavioral risk factor surveillance system, 2013-2023," *Preventing Chronic Disease*, vol. 22, p. E15, 2025.

[7] Centers for Disease Control and Prevention, "2024 BRFSS survey data and documentation," 2024. [Online]. Available: https://www.cdc.gov/brfss/annual_data/annual_2024.html

[8] S. Kasim, P. N. F. Amir Rudin, X. N. Kiew, N. Ibrahim, N. Mat Nasir, L. B. Feng, H. Hamidi, K. S. Ibrahim, R. E. Raja Shariff, S. Abdul-Razak, K. Negishi, and S. Malek, "Enhancing cardiovascular risk prediction in Asian populations: A machine learning approach integrated with digital health platforms," *International Journal of Cardiology Cardiovascular Risk and Prevention*, vol. 27, p. 200509, 2025.

[9] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, "DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network," *BMC Medical Research Methodology*, vol. 18, no. 1, p. 24, 2018.

[10] M. Makino, R. Yoshimoto, M. Ono, T. Itoko, T. Katsuki, A. Koseki, M. Kudo, K. Haida, J. Kuroda, R. Yanagiya, E. Saitoh, K. Hoshinaga, Y. Yuzawa, and A. Suzuki, "Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning," *Scientific Reports*, vol. 9, no. 1, p. 11862, 2019.

[11] A. R. Kulkarni, A. A. Patel, K. V. Pipal, S. G. Jaiswal, M. T. Jaisinghani, V. Thulkar, L. Gajbhiye, P. Gondane, A. B. Patel, M. Mamtani, and H. Kulkarni, "Machine-learning algorithm to non-invasively detect diabetes and pre-diabetes from electrocardiogram," *BMJ Innovations*, vol. 9, no. 1, pp. 32–42, 2023.

[12] W. C. Cockerham, B. W. Hamby, and G. R. Oates, "The social determinants of chronic disease," *American Journal of Preventive Medicine*, vol. 52, no. 1, Supplement 1, pp. S5–S12, 2017.

[13] United Nations Committee on Economic, Social and Cultural Rights, "General comment no. 14 (2000): The right to the highest attainable standard of health (article 12 of the international covenant on economic, social and cultural rights)," 2000, geneva: United Nations. [Online]. Available: https://www.ohchr.org/sites/default/files/Documents/Issues/Women/WRGS/Health/GC14.pdf

[14] S. P. Goldstein, A. Tovar, H. M. Espel-Huynh, and K. C. Stowers, "Applying a social determinants of health framework to guide digital innovations that reduce disparities in chronic disease," *Psychosomatic Medicine*, vol. 85, no. 7, pp. 659–669, 2023.

[15] D. S. Kim, A. A. Eltahir, S. Ngo, and F. Rodriguez, "Bridging the gap: How accounting for social determinants of health can improve digital health equity in cardiovascular medicine," *Current Atherosclerosis Reports*, vol. 27, no. 1, p. 9, 2024.

[16] Z. Zhou, D. Jin, J. He, S. Zhou, J. Wu, S. Wang, Y. Zhang, and T. Feng, "Digital health platform for improving the effect of the active health management of chronic diseases in the community: Mixed methods exploratory study," *Journal of Medical Internet Research*, vol. 26, p. e50959, 2024.

[17] S. Han, R. F. Zhang, L. Shi, R. Richie, H. Liu, A. Tseng, W. Quan, N. Ryan, D. Brent, and F. R. Tsui, "Classifying social determinants of health from unstructured electronic health records using deep learning-based natural language processing," *Journal of Biomedical Informatics*, vol. 127, p. 103984, 2022.

[18] T. G. Dietterich, "Ensemble methods in machine learning," in *International Workshop on Multiple Classifier Systems*, 2000.

[19] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[20] Y. Jin, "Assessing the relative importance of sociocontextual determinants of depression: A random forest analysis of young Korean adults," *Journal of Adolescent Health*, 2025.

[21] T. Chen, "XGBoost: A scalable tree boosting system," Ph.D. dissertation, Cornell University, 2016.

[22] M. P. LaValley, "Logistic regression," *Circulation*, vol. 117, no. 18, pp. 2395–2399, 2008.

[23] Z. Y. Algamal and M. H. Lee, "Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification," *Computers in Biology and Medicine*, vol. 67, pp. 136–145, 2015.

[24] A. Masood and K. Ahmad, "A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance," *Journal of Cleaner Production*, vol. 322, p. 129072, 2021.

[25] V. Sze, Y. H. Chen, T. J. Yang, and J. S. Emer, "Efficient processing of deep neural networks: A tutorial and survey," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2295–2329, 2017.

[26] J. Fox and S. Weisberg, "Cox proportional-hazards regression for survival data," *An R and S-PLUS Companion to Applied Regression*, 2002.

[27] F.-A. Coutu, O. C. Iorio, and B. A. Ross, "Remote patient monitoring strategies and wearable technology in chronic obstructive pulmonary disease," *Frontiers in Medicine*, vol. 10, 2023.

[28] M. Najarian, A. Goudie, J. P. Bona, M. Rezaeiahari, S. G. Young, C. A. Bogulski, and C. J. Hayes, "Socioeconomic determinants of remote patient monitoring implementation among rural and urban hospitals," *Telemedicine and e-Health*, vol. 29, no. 11, pp. 1624–1633, 2023.

[29] T. Shaik, X. Tao, N. Higgins, L. Li, R. Gururajan, X. Zhou, and U. R. Acharya, "Remote patient monitoring using artificial intelligence: Current state, applications, and challenges," *WIREs Data Mining and Knowledge Discovery*, vol. 13, no. 2, p. e1485, 2023.