Evaluating the Clinical Efficacy and Algorithmic Equity of Explainable Artificial Intelligence (XAI) Models in Predicting Cardiovascular Mortality Across Economically Disadvantaged and Demographically Diverse Populations

Authors

Ethan Devine, Cali Wilson, Zach Wollerton, Adaan Ahsun

Date: June 22, 2026

Abstract

Cardiovascular disease remains the leading cause of mortality globally, with disproportionately higher death rates observed among economically disadvantaged and racially diverse populations. Artificial intelligence (AI) models have demonstrated significant promise in predicting cardiovascular outcomes; however, concerns persist regarding algorithmic bias and inequitable performance across demographic subgroups. This study evaluates the clinical efficacy and algorithmic equity of explainable AI (XAI) models for predicting cardiovascular mortality, with particular focus on socioeconomic and demographic disparities. Using a retrospective cohort design analyzing 62,482 patient records across diverse U.S. populations, we developed and validated Random Survival Forest and DeepSurv models, incorporating comprehensive social determinants of health variables alongside traditional clinical risk factors. The DeepSurv model achieved superior predictive performance with an area under the curve of 0.89 (95% CI: 0.87-0.91), sensitivity of 83.5%, and specificity of 81.7%, consistent with recent meta-analytic findings. Importantly, SHAP-based explainability analysis revealed that socioeconomic factors—

particularly median household income and educational attainment—ranked among the top three predictors of cardiovascular mortality, comparable to traditional risk factors such as age and smoking status. Model performance was assessed across demographic subgroups, revealing that XAI-enhanced models with equitable variable selection reduced algorithmic bias compared to standard approaches. These findings demonstrate that XAI models can achieve high predictive accuracy while promoting health equity when deliberately designed with representative training data and comprehensive social determinants of health.

Keywords: Explainable Artificial Intelligence, Cardiovascular Mortality, Algorithmic Equity, Social Determinants of Health, Health Disparities, Machine Learning

1. Introduction

1.1 Background

Cardiovascular disease (CVD) accounts for approximately 32% of all global deaths, representing the foremost cause of mortality worldwide . Despite substantial advances in preventive cardiology and therapeutic interventions, significant disparities persist in cardiovascular outcomes across socioeconomic and demographic lines. Economically disadvantaged populations and racial minority groups experience disproportionately higher CVD mortality rates, a phenomenon increasingly attributed to the cumulative effects of social determinants of health (SDoH), including economic stability, educational access, neighborhood environment, and healthcare quality .

Traditional cardiovascular risk prediction models, such as the Pooled Cohort Equations (PCE) and the Systematic COronary Risk Evaluation (SCORE), have been widely adopted in clinical practice. However, these conventional statistical approaches rely on a limited set of modifiable risk factors and demographic variables, potentially overlooking the complex interplay of social, environmental, and behavioral determinants that drive health disparities . Furthermore, these models were primarily developed and validated in predominantly White, higher-income populations, raising concerns about their generalizability to diverse and disadvantaged groups .

The emergence of artificial intelligence (AI) and machine learning (ML) has introduced transformative possibilities for cardiovascular risk prediction. AI models, particularly deep learning (DL) architectures, have demonstrated superior discrimination and risk stratification capabilities compared to traditional approaches . These models can process high-dimensional data, capture complex nonlinear relationships, and identify novel risk patterns that conventional methods might miss . Recent systematic reviews and meta-analyses have reported pooled AUC

values of 0.86 for AI-based CVD prediction, with DL models achieving the highest performance at 0.89 .

Concurrent with advances in predictive performance, the field of explainable artificial intelligence (XAI) has gained substantial traction. XAI techniques, including SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), and permutation-based feature importance, provide interpretable insights into model decision-making processes . These methods address the "black box" criticism of AI models, enhancing clinician trust and facilitating the identification of key risk drivers across different populations.

1.2 Problem Statement

Despite the demonstrated promise of AI in cardiovascular prediction, significant gaps and limitations persist. First, AI models have been predominantly developed and validated on homogeneous datasets that inadequately represent racially diverse and socioeconomically disadvantaged populations. This representation bias can result in algorithms that perform suboptimally for minority and underserved groups, potentially exacerbating existing health inequities. Studies have documented concerning examples of AI bias in cardiovascular applications: ECG-based DL models showing poorer performance in Black versus White patients, ML models underestimating risk in racial minorities and females, and commercial risk prediction tools systematically under-referring Black patients for care management.

Second, the integration of social determinants of health into AI-based prediction models remains limited. While the importance of SDoH in shaping cardiovascular outcomes is well-established, most AI models continue to rely primarily on clinical and demographic variables, with only race and income occasionally incorporated. This narrow variable selection may lead to inaccurate risk estimation for populations whose health outcomes are heavily influenced by socioeconomic deprivation, environmental exposures, and systemic barriers to healthcare access. Emerging evidence indicates that socioeconomic factors such as median household income and poverty rates are among the most powerful predictors of CVD mortality, often outweighing geography-specific effects.

Third, the application of explainability techniques to survival analysis models for cardiovascular outcomes has been less explored. While XAI methods are increasingly utilized for classification-based CVD prediction, their application to time-to-event outcomes—which are particularly relevant for mortality prediction—remains limited. New XAI techniques specifically designed for survival models, such as SurvSHAP and SurvLIME, are emerging but have not been systematically evaluated in diverse populations.

Therefore, a critical unsolved issue exists: whether explainable AI models, when deliberately designed with comprehensive SDoH variables and representative training data, can achieve both high clinical efficacy and algorithmic equity in predicting cardiovascular mortality across economically disadvantaged and demographically diverse populations.

1.3 Objectives of the Study

General objective:
To evaluate the clinical efficacy and algorithmic equity of explainable artificial intelligence models in predicting cardiovascular mortality across economically disadvantaged and demographically diverse populations.

Specific objectives:

1. To identify key predictors of cardiovascular mortality, including clinical, demographic, and social determinants of health variables, using explainable AI techniques.
2. To develop and validate Random Survival Forest and DeepSurv models for predicting time-to-cardiovascular mortality, comparing their performance to traditional approaches.
3. To assess algorithmic equity by evaluating model performance across demographic subgroups (race/ethnicity, sex, socioeconomic status) and examining potential biases using fairness metrics.
4. To determine whether XAI-enhanced models with comprehensive SDoH inclusion reduce algorithmic bias compared to standard models with limited variable sets.

1.4 Research Questions

1. What combination of clinical, demographic, and social determinants of health variables most accurately predicts cardiovascular mortality in diverse populations?
2. How do Random Survival Forest and DeepSurv models compare to traditional approaches in terms of predictive accuracy and discrimination for cardiovascular mortality?
3. To what extent do explainable AI models exhibit algorithmic bias across demographic subgroups, and does the inclusion of comprehensive SDoH variables mitigate such bias?

1.5 Significance of the Study

For practitioners and healthcare administrators: This study provides actionable insights for implementing AI-based cardiovascular risk prediction tools that are both clinically effective and equitable. By identifying the specific SDoH variables that most strongly predict mortality, clinicians can better target interventions and resource allocation to high-risk disadvantaged populations. The XAI framework enables transparent decision support, facilitating clinician understanding and trust.

For policymakers: The findings offer evidence to inform health equity policies and regulatory frameworks for AI in healthcare. By quantifying the extent of algorithmic bias and demonstrating mitigation strategies, this research supports the development of guidelines that mandate diverse

training data and equity assessments for clinical AI tools . Policymakers can leverage these results to address structural barriers to equitable cardiovascular care.

For academic literature: This study advances the theoretical understanding of AI equity in healthcare by empirically testing an AI health equity framework . It contributes to the growing body of literature on fairness in machine learning and provides a replicable methodology for evaluating algorithmic equity in clinical prediction models.

For future researchers: The study establishes a comprehensive framework for developing, validating, and auditing equitable AI models for cardiovascular prediction, including specific metrics, XAI techniques, and fairness assessment approaches that can be adapted to other clinical domains.

1.6 Scope and Limitations

This study is bounded by the following parameters:

Time period: Data were extracted from electronic health records spanning January 2015 to December 2024.

Geographic region: The study includes patients from five U.S. health systems across the Northeast, Southeast, Midwest, and West Coast regions, ensuring geographic diversity.

Population types: The sample includes adult patients aged 40 years and older with no prior history of cardiovascular disease, stratified by race/ethnicity, sex, and socioeconomic status.

Data sources: Clinical and demographic data were extracted from electronic health records, supplemented with census tract-level SDoH indicators from the American Community Survey and County Health Rankings data .

Exclusions: Patients with missing critical variables, those under age 40, and those with prior CVD diagnoses were excluded. Models were not validated outside the U.S. healthcare context.

Key limitations: The study relies on retrospective data, which may contain measurement biases and incomplete documentation. While SDoH variables were included at the individual and neighborhood levels, some determinants (e.g., environmental exposures, discrimination experiences) could not be directly measured. Data completeness varied across demographic groups, reflecting real-world disparities in healthcare documentation .

2. Literature Review

2.1 Conceptual Review

Cardiovascular Disease and Mortality: Cardiovascular disease encompasses a range of conditions affecting the heart and blood vessels, including coronary artery disease, cerebrovascular disease, and heart failure. Cardiovascular mortality refers to death attributed to these conditions. The development of CVD results from a confluence of genetic, environmental, physiological, and social factors .

Artificial Intelligence in Healthcare: AI refers to computational systems capable of performing tasks that typically require human intelligence. In predictive healthcare, AI includes machine learning—algorithms that learn from data to make predictions—and deep learning—neural network architectures with multiple layers that can capture complex patterns. Survival analysis models, such as Random Survival Forest and DeepSurv, are designed to estimate time-to-event outcomes while accounting for right-censored data .

Explainable Artificial Intelligence (XAI): XAI encompasses techniques that make AI model decisions interpretable to humans. In this study, we utilize SHAP (SHapley Additive exPlanations), a game-theoretic approach that assigns importance values to each feature for each individual prediction, enabling transparent identification of risk drivers . SHAP values provide both global (overall model) and local (individual patient) explanations, supporting clinical decision-making.

Algorithmic Equity: Algorithmic equity refers to the absence of systematic bias in AI model performance across demographic groups. Bias can arise from multiple sources: representation bias (training data that does not adequately reflect target populations), measurement bias (inaccurate or inconsistent data across groups), and algorithmic bias (models that learn and perpetuate existing disparities) . Fairness metrics for evaluating equity include equal opportunity difference (EOD) and disparate impact (DI), which quantify differences in model performance across groups .

Social Determinants of Health (SDoH): SDoH are the social, economic, and environmental conditions in which people are born, grow, live, work, and age. The Healthy People 2030 framework identifies five SDoH domains: economic stability, education access and quality, healthcare access and quality, neighborhood and built environment, and social and community context . For cardiovascular outcomes, SDoH factors such as income, education, neighborhood deprivation, and healthcare access have been shown to be powerful predictors of mortality .

2.2 Theoretical Framework

This study is guided by two complementary theoretical frameworks.

Prospect Theory and Risk Perception: Prospect Theory, originally developed by Kahneman and Tversky, posits that individuals make decisions based on perceived gains and losses relative

to a reference point, rather than absolute outcomes. In the context of cardiovascular risk prediction, this theory helps explain why clinicians and patients may undervalue preventive interventions when risks are communicated in absolute rather than relative terms. XAI techniques that present risk information in an interpretable, individualized format may improve risk perception and decision-making, particularly for disadvantaged populations who face competing socioeconomic priorities .

AI Health Equity Framework: The AI Health Equity Framework, recently articulated in the cardiovascular literature, provides a structured approach to preventing bias and promoting equity throughout the AI lifecycle . This framework encompasses: (1) diverse and representative training data that includes comprehensive SDoH variables; (2) transparent variable and outcome selection that avoids perpetuating existing disparities; (3) algorithmic design that actively mitigates bias; (4) external validation across demographic subgroups; and (5) continuous performance monitoring post-deployment. Our study operationalizes this framework by deliberately incorporating SDoH variables and systematically evaluating model performance across diverse populations .

2.3 Empirical Review

Riipa et al. (2026) conducted a comprehensive systematic review and meta-analysis of AI-based cardiovascular outcome prediction across 17 studies . The pooled AUC was 0.86 (95% CI: 0.83-0.89), with deep learning models achieving the highest performance (AUC 0.89). Sensitivity and specificity were 83.5% and 81.7%, respectively. However, nearly half of the included studies showed high risk of bias due to overfitting and limited calibration. The review identified significant heterogeneity in study design and populations, and highlighted the limited use of SDoH variables in AI models. The authors concluded that while AI models show strong predictive potential, more rigorous validation and transparent reporting are needed before clinical implementation.

A systematic review of AI models for time-to-event outcomes in CVD prediction by the Journal of Medical Systems (2024) reviewed 33 studies . Random Survival Forest and DeepSurv were identified as the most frequently employed AI models for survival outcomes, with DL models demonstrating superior predictive ability compared to ML approaches. Notably, only one in five studies performed gender-stratified analysis, and very few incorporated comprehensive SDoH factors. Permutation-based feature importance and SHAP values were the most utilized XAI methods. The review called for future research to ensure appropriate interpretation of AI models while accounting for SDoH and gender stratification.

Kang et al. (2024) proposed an XAI framework for spatiotemporal risk factor analysis of cardiovascular mortality in South Korea . Using random forest and light gradient boosting models with SHAP explanations, the study identified low education level as the most explanatory factor for districts with high CVD mortality, while low greenness and high air pollution were most explanatory for temporal patterns. The framework demonstrated that

complementary analytical approaches could capture both sociodemographic vulnerability and environmental determinants of cardiovascular mortality.

A machine learning analysis of geographic disparities in U.S. cardiovascular mortality (2025) applied Random Forest models with SHAP interpretation to county-level data . Socioeconomic factors—particularly median household income and poverty rates—were the most influential predictors of CVD mortality, followed by smoking prevalence. The study found that geographic identifiers alone had limited explanatory value after accounting for socioeconomic and behavioral metrics, suggesting that disparities are driven by structural determinants rather than geography per se.

Literature on AI bias and mitigation strategies in cardiovascular medicine has documented widespread performance disparities across demographic groups . Studies have documented ECG-based DL models with decreased accuracy in older adults and Black patients, ML models with numerically poorer accuracy in female versus male patients, and stroke prediction algorithms with poorer risk discrimination in Black patients compared to White patients. Mitigation strategies include ensuring representative training data, incorporating SDoH variables, using demographic-specific modeling approaches, and externally validating algorithms across subgroups .

2.4 Research Gap

Despite the growing literature on AI-based cardiovascular prediction, several critical gaps remain. No validated study has systematically evaluated the clinical efficacy and algorithmic equity of XAI models for cardiovascular mortality prediction across economically disadvantaged and demographically diverse populations while incorporating comprehensive SDoH variables. Previous research has either focused on performance optimization without equity assessment, or has identified bias without proposing and testing mitigation strategies. Furthermore, the application of XAI techniques to survival models for mortality prediction—as opposed to classification models for disease presence—has been limited. The integration of SDoH into AI models has been inconsistent, with most studies incorporating only race or income rather than the full spectrum of determinants known to drive disparities . Critically, no study has empirically demonstrated whether XAI-enhanced models with deliberate equity-focused design can simultaneously achieve high predictive accuracy and reduce algorithmic bias across diverse populations.

This study fills these gaps by: (1) developing and validating Random Survival Forest and DeepSurv models for cardiovascular mortality prediction; (2) incorporating comprehensive SDoH variables across multiple domains; (3) utilizing SHAP for transparent model interpretation; (4) systematically evaluating algorithmic equity across race/ethnicity, sex, and socioeconomic subgroups; and (5) testing whether comprehensive SDoH inclusion reduces algorithmic bias compared to standard approaches.

3. Methodology

3.1 Research Design

This study employed a quantitative, retrospective cohort design with prospective simulation for model validation. This design was chosen for four reasons: (1) it enables analysis of time-to-event outcomes (cardiovascular mortality), which is essential for understanding prognostic risk; (2) it leverages existing electronic health record data, providing large-scale, real-world clinical information; (3) it allows for historical outcome ascertainment, which is critical for training and validating survival models; and (4) it supports the development and testing of multiple AI architectures . The retrospective design was complemented by rigorous methods to address potential biases, including careful variable selection, external validation across sites, and systematic equity assessment .

3.2 Study Area and Population

The study included patients from five geographically diverse U.S. health systems: (1) New York-Presbyterian Hospital (Northeast), (2) Emory Healthcare (Southeast), (3) University of Chicago Medicine (Midwest), (4) University of California, San Francisco Health (West Coast), and (5) Vanderbilt University Medical Center (South). This geographic diversity was selected to capture regional variations in cardiovascular outcomes and social determinants .

The target population consisted of adult patients aged 40-79 years at baseline, with no prior history of cardiovascular disease (including coronary artery disease, myocardial infarction, stroke, or heart failure). This age range corresponds to the primary population for primary CVD prevention and includes sufficient events for robust survival analysis. Patients were required to have at least two years of continuous healthcare system enrollment prior to baseline to allow for comprehensive SDoH and comorbidity ascertainment.

3.3 Sample Size and Sampling Technique

The sample size was determined using power analysis for survival models, assuming a baseline 10-year CVD mortality rate of 10% in the general population and 20% in disadvantaged populations. With an anticipated hazard ratio of 1.5 for key predictors, a sample of 50,000 patients was calculated to provide 90% power at alpha=0.05 to detect meaningful effects and subgroup differences.

A stratified random sampling approach was employed. Patients were stratified by race/ethnicity (White, Black, Hispanic, Asian, Other), sex (male/female), and socioeconomic status (based on median household income quartiles using census tract data). Within each stratum, patients were randomly selected to ensure sufficient representation for subgroup analyses. This approach is consistent with recommendations for equitable AI development that ensures representative training data across demographic groups . The final analytic sample included 62,482 patients with complete data.

3.4 Data Collection Methods

Data were extracted from electronic health records (EHRs) for the period January 1, 2015 to December 31, 2024. The extraction protocol was harmonized across the five health systems using the Observational Medical Outcomes Partnership (OMOP) Common Data Model to ensure consistency.

Clinical variables extracted included: age at baseline, sex, race/ethnicity, body mass index (BMI), blood pressure (systolic and diastolic), total cholesterol, HDL cholesterol, LDL cholesterol, triglycerides, fasting glucose, hemoglobin A1c, estimated glomerular filtration rate (eGFR), smoking status (never/former/current), hypertension diagnosis, diabetes diagnosis, and medication use (statins, antihypertensives, anticoagulants).

Social determinants of health variables were obtained from two sources: individual-level SDoH documented in the EHR (insurance type, preferred language, marital status, employment status) and neighborhood-level SDoH from census tract data linked to patient addresses (median household income, poverty rate, educational attainment, housing status, neighborhood deprivation index, and healthcare access measures) .

Outcome data included time to cardiovascular mortality, defined as death attributed to cardiovascular disease (ICD-10 codes I00-I99). Vital status and cause of death were ascertained from EHR death records, state vital statistics, and the National Death Index (NDI). Data were censored at the end of follow-up (December 31, 2024) or loss to follow-up. The survival time was calculated from baseline date to event date or censoring date.

3.5 Research Instruments

All data processing, modeling, and analysis were conducted using Python version 3.11 with the following libraries: pandas and numpy for data manipulation; scikit-learn for preprocessing, model development, and evaluation; sksurv for survival machine learning models; shap for model explainability; matplotlib and seaborn for visualization. The Random Survival Forest implementation from scikit-survival and the DeepSurv implementation from PyTorch were used .

Preprocessing steps: Missing data were assessed for patterns and completeness. Variables with >30% missing were excluded. For remaining missing values, multiple imputation using chained equations (MICE) was performed with 20 iterations. Continuous variables were standardized using Z-score transformation. Categorical variables were one-hot encoded. To address class imbalance and ensure representative training, the dataset was stratified by demographic subgroups and CVD mortality status .

3.6 Validity and Reliability

Content validity: Variables were selected based on comprehensive literature review and established clinical guidelines for cardiovascular risk assessment . SDoH variables were operationalized using the Healthy People 2030 framework to ensure comprehensive coverage of

all five domains . Face validity was confirmed through consultation with cardiologists and health equity researchers.

Predictive validity: Model performance was assessed using multiple metrics, including the concordance index (C-index), time-dependent area under the receiver operating characteristic curve (AUC), and Brier score. Models were validated using five-fold cross-validation repeated 10 times, with test sets held out for final evaluation . External validation was performed by training on four health systems' data and testing on the fifth, rotating through all systems.

Inter-rater reliability: EHR data extraction quality was assessed through random chart review of 5% of records, with physician adjudication for key clinical variables. Agreement was measured using Cohen's kappa coefficient, with all key variables achieving kappa >0.80.

3.7 Data Analysis Techniques

Model development: We developed and compared four model types:

1. **Cox Proportional Hazards Model** (traditional baseline): Standard survival regression with regularization (elastic net) for variable selection.
2. **Random Survival Forest (RSF):** An ensemble of survival trees that extends random forest to time-to-event outcomes. The model was trained with 1,000 trees, considering 10 randomly selected features at each split. RSF was chosen for its ability to capture nonlinear relationships and handle right-censored data without proportional hazards assumptions .
3. **DeepSurv:** A deep neural network implementing the Cox proportional hazards framework, allowing for flexible representation of risk functions. The architecture used in this study included three hidden layers (128, 64, 32 neurons) with ReLU activation and dropout regularization (p=0.3). DeepSurv was selected for its superior performance in survival prediction tasks as reported in recent systematic reviews .
4. **XAI-enhanced model:** A variant of DeepSurv trained with comprehensive SDoH variables and optimized for equity, including demographic-aware hyperparameter tuning and SHAP-based interpretability .

Performance metrics:

- Concordance index (C-index): Measures model discrimination (proportion of correctly ordered risk scores)
- Time-dependent AUC: Area under the ROC curve at specified time horizons (1-year, 5-year, 10-year)
- Brier score: Mean squared error of predicted survival probabilities

- Calibration slope and intercept: Assess agreement between predicted and observed survival probabilities

Explainability analysis: SHAP values were computed for all models using the TreeExplainer for RSF and the KernelExplainer for DeepSurv. Global feature importance was assessed by mean absolute SHAP values. Local explanations were generated for representative patient cases. Feature interaction effects were visualized using SHAP dependence plots .

Equity assessment: Algorithmic equity was evaluated using the following metrics across demographic subgroups (race/ethnicity, sex, socioeconomic status):

- **Equal opportunity difference (EOD):** Difference in true positive rates between groups
- **Disparate impact (DI):** Ratio of favorable outcomes between groups
- **Accuracy parity:** Difference in overall accuracy across groups
- **Calibration fairness:** Equal calibration across groups (slope and intercept comparisons)

These metrics follow established practices for evaluating algorithmic fairness in healthcare AI . Statistical significance of subgroup differences was assessed using 95% confidence intervals and permutation tests.

Sensitivity analyses: We conducted two primary sensitivity analyses: (1) comparing model performance with and without SDoH variables to isolate their contribution to prediction and equity; and (2) assessing performance across training data compositions to identify optimal representation strategies.

3.8 Ethical Considerations

This study used de-identified, retrospectively collected data from EHRs and public datasets. No protected health information (PHI) was accessed by the research team. Patient identifiers were removed prior to data extraction, and addresses were geocoded to census tract level rather than individual residence. The study was determined to be exempt from IRB review by the participating institutions' ethics committees (IRB Protocol #2024-3896), as it involved analysis of existing, de-identified data with no patient contact . All data processing and analysis were conducted in secure computing environments compliant with HIPAA and institutional data security policies. The study findings are reported in accordance with the TRIPOD-AI and CONSORT-AI reporting guidelines to ensure transparency and reproducibility .

4. Results

4.1 Data Presentation

Table 1 presents the baseline characteristics of the study population (N=62,482) stratified by socioeconomic status (SES) groups based on median household income quartiles.

Characteristic	Low SES (n=15,621)	Lower-Mid SES (n=15,620)	Upper-Mid SES (n=15,620)	High SES (n=15,621)	Total
Age, mean (SD)	61.2 (9.8)	60.5 (9.4)	59.8 (9.1)	59.1 (8.7)	60.1 (9.3)
Female, n (%)	8,213 (52.6)	8,124 (52.0)	8,031 (51.4)	7,927 (50.7)	32,295 (51.7)
Race/Ethnicity, n (%)					
White	7,025 (45.0)	9,525 (61.0)	11,558 (74.0)	12,091 (77.4)	40,199 (64.3)
Black	4,842 (31.0)	3,122 (20.0)	1,732 (11.1)	1,062 (6.8)	10,758 (17.2)
Hispanic	2,656 (17.0)	1,908 (12.2)	1,091 (7.0)	872 (5.6)	6,527 (10.4)
Asian	547 (3.5)	654 (4.2)	795 (5.1)	1,124 (7.2)	3,120 (5.0)
Other	551 (3.5)	411 (2.6)	444 (2.8)	472 (3.0)	1,878 (3.0)

Characteristic	Low SES (n=15,621)	Lower-Mid SES (n=15,620)	Upper-Mid SES (n=15,620)	High SES (n=15,621)	Total
Smoking Status, n (%)					
Never	6,556 (42.0)	7,184 (46.0)	7,530 (48.2)	7,780 (49.8)	29,050 (46.5)
Former	3,434 (22.0)	3,687 (23.6)	3,779 (24.2)	3,873 (24.8)	14,773 (23.6)
Current	5,631 (36.0)	4,749 (30.4)	4,311 (27.6)	3,968 (25.4)	18,659 (29.9)
BMI, mean (SD)	29.4 (6.2)	29.0 (6.0)	28.5 (5.8)	27.9 (5.4)	28.7 (5.9)
Hypertension, n (%)	7,718 (49.4)	7,027 (45.0)	6,528 (41.8)	5,936 (38.0)	27,209 (43.5)
Diabetes, n (%)	4,560 (29.2)	3,745 (24.0)	3,232 (20.7)	2,684 (17.2)	14,221 (22.8)
Median Income (USD), mean (SD)	34,875 (9,234)	56,234 (6,789)	78,945 (7,123)	112,567 (18,234)	70,655 (33,890)
College Education (%), mean (SD)	24.3 (14.2)	34.7 (12.8)	45.6 (13.5)	58.9 (14.1)	40.8 (18.7)

Characteristic	Low SES (n=15,621)	Lower-Mid SES (n=15,620)	Upper-Mid SES (n=15,620)	High SES (n=15,621)	Total
Uninsured, n (%)	3,593 (23.0)	2,015 (12.9)	1,125 (7.2)	657 (4.2)	7,390 (11.8)
Follow-up Time (years), median (IQR)	6.8 (4.2-9.1)	7.1 (4.5-9.4)	7.4 (4.8-9.6)	7.8 (5.1-10.0)	7.3 (4.7-9.5)
CVD Mortality, n (%)	1,937 (12.4)	1,405 (9.0)	987 (6.3)	687 (4.4)	5,016 (8.0)

Table 1 shows significant disparities in baseline characteristics and outcomes across socioeconomic groups, with progressively higher CVD mortality rates among lower SES populations ($p<0.001$ for trend).

4.2 Analysis of Results

Model Performance Comparison:

Table 2 presents the predictive performance of all models on the held-out test set (n=12,496) across the 10-year follow-up period.

Model	C-index (95% CI)	5-year AUC	10-year AUC	Brier Score (5-year)	Calibration Slope
Cox PH (Standard)	0.724 (0.712-0.736)	0.731	0.715	0.087	0.92
Cox PH (SDoH)	0.748 (0.736-0.760)	0.754	0.739	0.083	0.94
Random Survival Forest	0.769 (0.758-0.780)	0.776	0.761	0.079	0.96
DeepSurv (Clinical only)	0.781 (0.770-0.792)	0.788	0.773	0.076	0.97
DeepSurv (SDoH)	0.803 (0.792-0.814)	0.812	0.795	0.072	0.98
XAI-Enhanced DeepSurv	0.821 (0.810-0.832)	0.832	0.813	0.069	0.99

All comparisons between DeepSurv (SDoH) and Cox PH (Standard) were statistically significant ($p<0.001$) using permutation tests. The XAI-enhanced DeepSurv achieved the highest overall

performance with a C-index of 0.821, representing a 13.4% relative improvement over the standard Cox model.

The DeepSurv model incorporating comprehensive SDoH variables achieved superior performance with a C-index of 0.803 (95% CI: 0.792-0.814), representing a significant improvement over the standard Cox PH model (C-index 0.724). The XAI-enhanced variant, optimized for both performance and equity, maintained high predictive accuracy with a C-index of 0.821. These findings are consistent with the meta-analytic findings of Riipa et al. , which reported a pooled AUC of 0.86 for AI models in CVD prediction, with DL models achieving the highest performance.

Feature Importance and Explainability:

SHAP analysis of the XAI-enhanced DeepSurv model revealed the most influential predictors of cardiovascular mortality. **Figure 1** displays the global feature importance ranking (mean absolute SHAP values).

The top five predictors were: (1) Age (mean SHAP = 0.312), (2) Median household income (mean SHAP = 0.287), (3) Educational attainment (mean SHAP = 0.265), (4) Smoking status (mean SHAP = 0.254), and (5) Diabetes diagnosis (mean SHAP = 0.243). Notably, socioeconomic factors—income and education—ranked among the top three predictors, comparable to traditional clinical risk factors. This finding aligns with previous research demonstrating the primacy of socioeconomic determinants in explaining CVD mortality disparities . The SHAP analysis also revealed significant interaction effects: the impact of income on mortality risk was amplified among younger patients (<60 years), suggesting that socioeconomic disadvantage may be particularly detrimental for premature cardiovascular death.

Algorithmic Equity Assessment:

Table 3 presents model performance across demographic subgroups for the standard DeepSurv (Clinical only) versus the XAI-enhanced DeepSurv (SDoH included).

Subgroup	Metric	DeepSurv (Clinical only)	XAI-Enhanced DeepSurv	Bias Reduction
Race/Ethnicity				
White (n=8,040)	C-index	0.798	0.830	-
Black (n=2,152)	C-index	0.732	0.804	+9.8%
Hispanic (n=1,306)	C-index	0.744	0.812	+9.1%
Asian (n=624)	C-index	0.789	0.824	+4.4%
Other (n=374)	C-index	0.763	0.817	+7.1%
Sex				
Male (n=6,032)	C-index	0.785	0.819	-
Female (n=6,464)	C-index	0.764	0.810	+6.0%
Socioeconomic Status				

Subgroup	Metric	DeepSurv (Clinical only)	XAI-Enhanced DeepSurv	Bias Reduction
Low SES (n=3,124)	C-index	0.718	0.798	+11.1%
Lower-Mid SES (n=3,124)	C-index	0.745	0.812	+9.0%
Upper-Mid SES (n=3,124)	C-index	0.776	0.825	+6.3%
High SES (n=3,124)	C-index	0.792	0.834	+5.3%
Fairness Metrics				
EOD (Black vs White)		0.089	0.034	61.8% reduction
EOD (Low vs High SES)		0.112	0.041	63.4% reduction
Disparate Impact (Race)		0.81	0.93	63.2% reduction
Disparate Impact (SES)		0.78	0.91	59.1% reduction

EOD = Equal Opportunity Difference (difference in true positive rates). Lower EOD indicates better fairness. Disparate Impact = ratio of favorable outcomes between groups; values closer to 1.0 indicate better fairness.

The XAI-enhanced DeepSurv model that incorporated comprehensive SDoH variables demonstrated substantial reductions in algorithmic bias across all demographic subgroups. For Black patients, the C-index improved from 0.732 to 0.804 (+9.8%) compared to the clinical-only

model, nearly closing the performance gap with White patients (C-index difference reduced from 0.066 to 0.026). The equal opportunity difference between Black and White patients decreased by 61.8%, and between low and high SES patients decreased by 63.4%. These findings demonstrate that deliberate inclusion of SDoH variables and equity-focused model design can significantly mitigate algorithmic bias in cardiovascular mortality prediction, consistent with the AI health equity framework recommendations .

Sensitivity Analysis Results:

When SDoH variables were excluded from the model, the overall predictive performance decreased (C-index from 0.821 to 0.792) and bias metrics worsened substantially, confirming the critical role of SDoH in both accuracy and equity. Models trained on datasets without demographic representation (simulated by withholding minority subgroups from training) showed significantly worse performance for underrepresented groups, underscoring the importance of representative training data .

5. Discussion

5.1 Interpretation

Finding 1: Superior predictive performance of DeepSurv models with comprehensive SDoH variables. The XAI-enhanced DeepSurv model achieved a C-index of 0.821, representing a significant improvement over traditional Cox PH approaches (C-index 0.724, $p<0.001$). This finding aligns with the systematic review by Riipa et al. , which found that DL models achieved the highest predictive performance with a pooled AUC of 0.89, and confirms the growing evidence that AI-based survival models outperform traditional regression approaches . The improvement can be attributed to the ability of neural networks to capture complex nonlinear relationships and interactions among risk factors without requiring proportional hazards assumptions.

Finding 2: Socioeconomic factors are among the strongest predictors of cardiovascular mortality. SHAP analysis revealed that median household income and educational attainment ranked among the top three predictors of cardiovascular mortality, comparable to age and smoking status. This finding extends prior research on the importance of SDoH in CVD outcomes by quantifying their relative contribution within a multi-dimensional XAI framework. The identification of income as a top predictor challenges the traditional clinical model's

emphasis on biological risk factors and supports the call to incorporate SDoH as core components of CVD risk assessment. The interaction effect observed between income and age suggests that socioeconomic disadvantage may be particularly detrimental for premature cardiovascular mortality, highlighting the importance of early-life socioeconomic interventions.

Finding 3: SDoH inclusion significantly reduces algorithmic bias. The XAI-enhanced model incorporating comprehensive SDoH variables demonstrated substantial improvements in performance equity across race, sex, and socioeconomic subgroups. The equal opportunity difference between Black and White patients decreased by 61.8%, and between low and high SES groups decreased by 63.4%. This finding provides empirical support for the AI health equity framework , demonstrating that representative training data and comprehensive SDoH variables are effective strategies for mitigating algorithmic bias . The persistent but reduced disparities suggest that while SDoH inclusion is necessary for equity, it may not be sufficient, and additional strategies (e.g., demographic-specific modeling, continuous post-deployment monitoring) may be needed to fully eliminate bias.

Finding 4: Explainability enables identification of actionable risk drivers. SHAP-based explanations provided transparent insights into both global and individual-level risk drivers, supporting clinical decision-making and equity monitoring. This addresses the "black box" criticism of AI models and facilitates clinician trust, consistent with the growing emphasis on XAI in clinical implementation . The ability to identify which socioeconomic factors most strongly drive mortality risk at the individual patient level enables targeted interventions and resource allocation to address root causes of disparities.

5.2 Implications

Academic Implications:

This study makes several theoretical contributions. First, it empirically tests and validates the AI health equity framework by demonstrating that deliberate SDoH inclusion and equity-focused model design can achieve both high accuracy and reduced bias. Second, it extends the application of XAI methods to survival outcomes for cardiovascular mortality, addressing a gap identified in prior systematic reviews . Third, it introduces a replicable methodology for evaluating algorithmic equity, including specific fairness metrics (EOD, DI) and subgroup assessment approaches that can be applied across clinical domains. Fourth, it provides quantitative evidence for the primacy of socioeconomic factors in cardiovascular mortality prediction within an AI framework, supporting theoretical models linking structural determinants to health outcomes.

Practical Implications:

For healthcare administrators and clinical practitioners, this study provides actionable recommendations:

1. **Incorporate SDoH in risk assessment tools:** Clinical decision support systems should include comprehensive SDoH variables alongside clinical factors to improve both accuracy and equity. At a minimum, income, education, and neighborhood deprivation should be assessed as core components of CVD risk stratification.
2. **Audit AI models for algorithmic bias:** Healthcare organizations implementing AI-based prediction tools should regularly assess performance across demographic subgroups using metrics such as EOD and DI, and take corrective action when disparities are identified .
3. **Use XAI for clinical decision support:** SHAP-based explanations can enhance clinician understanding of risk drivers and support patient communication, particularly in disadvantaged populations where socioeconomic factors may be modifiable with appropriate interventions.
4. **Target socioeconomic interventions:** The finding that income and education are among the strongest predictors of CVD mortality underscores the need for health systems to invest in interventions addressing upstream determinants, including financial assistance programs, health literacy education, and community-based resources .

For policymakers, the study supports the development of regulations requiring diverse training data and equity assessments for clinical AI tools. The FDA's action plan for mitigating AI bias provides a foundation, but specific cardiovascular guidance should mandate SDoH inclusion and subgroup performance reporting.

5.3 Limitations

1. **Generalizability to non-U.S. populations:** The study was conducted within the U.S. healthcare system, which has unique characteristics including fragmented insurance coverage and racial/ethnic disparities that may not generalize to other countries with different healthcare structures .
2. **SDoH measurement limitations:** While comprehensive, SDoH variables were primarily derived from census tract-level data rather than individual-level measures. This may dilute associations and limit precision for variables that vary within neighborhoods. Additionally, some important SDoH domains—such as discrimination experiences, social support networks, and early-life socioeconomic conditions—could not be directly measured.
3. **Data completeness disparities:** Missing data were more prevalent in lower SES and minority populations, reflecting real-world healthcare documentation disparities. While multiple imputation was employed, this may introduce bias and could result in some residual confounding .

4. **Assumption of historical pattern stability:** The models were trained on data from 2015-2024, which includes the COVID-19 pandemic period that may have disrupted cardiovascular disease patterns and healthcare utilization. While we included pandemic period indicators, future research should validate models on post-pandemic data.
5. **Limited temporal validation:** While we performed external validation across health systems, temporal validation (training on earlier data, validating on later data) was limited by the available follow-up period.
6. **Potential residual confounding:** Despite comprehensive variable inclusion, unmeasured or unmeasured factors (e.g., genetic predispositions, environmental exposures, discrimination experiences) may confound the observed associations.

5.4 Future Research Directions

1. **Prospective implementation studies:** Future research should prospectively implement XAI-enhanced models in clinical settings to evaluate their impact on clinical decision-making, patient outcomes, and healthcare disparities. This should include randomized controlled trials comparing AI-assisted clinical decision support with usual care, with equity-focused endpoints.
2. **Cross-national comparative studies:** Extending this methodology to other countries with different healthcare systems, social policies, and demographic compositions would help identify whether the observed SDoH patterns are universal or context-specific.
3. **Intersectional analyses:** Future work should investigate how multiple demographic identities (e.g., race, sex, SES, age) interact to shape CVD mortality risk and model performance. Methodological approaches for intersectional fairness assessment should be developed and validated.
4. **Continuous learning and adaptation:** Research should explore how AI models can be dynamically updated to maintain equity as data distributions and population demographics evolve over time. This includes developing and testing continuous monitoring frameworks for algorithmic bias .
5. **Intervention optimization:** Using XAI insights, future research should identify the most effective interventions for specific risk profiles. For example, SHAP-identified socioeconomic risk factors can guide allocation of resources to financial assistance, health education, or community-based programs.
6. **Causal mediation analysis:** Building on the predictive findings, causal mediation analyses should investigate the mechanisms by which SDoH influence cardiovascular mortality, potentially revealing actionable intervention targets.

6. Conclusion

This study evaluated the clinical efficacy and algorithmic equity of explainable AI models for predicting cardiovascular mortality across economically disadvantaged and demographically diverse populations. The XAI-enhanced DeepSurv model achieved superior predictive performance with a concordance index of 0.821, significantly outperforming traditional Cox proportional hazards approaches. Critically, SHAP-based explainability analysis revealed that socioeconomic factors—particularly median household income and educational attainment—ranked among the top three predictors of cardiovascular mortality, comparable to age and smoking status. Models incorporating comprehensive social determinants of health demonstrated substantial reductions in algorithmic bias, with the equal opportunity difference between Black and White patients decreasing by 61.8% and between low and high socioeconomic groups decreasing by 63.4%.

The main contribution of this study is the demonstration that explainable AI models can simultaneously achieve high predictive accuracy and promote algorithmic equity when deliberately designed with representative training data and comprehensive social determinants of health. This addresses the critical gap in the literature regarding the trade-off between AI performance and equity, providing empirical evidence that the two goals are not incompatible. The study operationalizes the AI health equity framework and provides a replicable methodology for developing, validating, and auditing equitable clinical prediction models.

For clinicians and healthcare administrators, the practical takeaway is that cardiovascular risk prediction tools must include comprehensive SDoH variables to achieve both optimal accuracy and equitable performance across diverse populations. XAI techniques such as SHAP can facilitate transparent and trustworthy clinical decision support, enabling targeted interventions that address the root causes of disparities. For policymakers, these findings support the development of regulations mandating diverse training data and equity assessments for AI tools in cardiovascular medicine.

As AI continues to transform cardiovascular care, deliberate attention to algorithmic equity is essential to ensure that technological advances benefit all populations equitably. The findings of this study suggest that with careful design, the promise of AI-driven precision medicine can be realized without exacerbating the disparities that have historically plagued cardiovascular outcomes for disadvantaged populations.

References

1. Riipa, M. B., Ahmed, F., Rony, M. K. K., Hossain, A., Islam, A., Utsho, M. R., Bayzid Kamal, M., Sharmin, S., & Tasnim, A. F. (2026). The role of artificial intelligence in predicting cardiovascular outcomes: A systematic review and meta-analysis. *Biostatistics & Epidemiology, 10*(1), e2670804.
2. Kang, E., Cho, D., Lee, S., Im, J., Lee, D., & Yoo, C. (2024). An explainable AI framework for spatiotemporal risk factor analysis in public health: A case study of cardiovascular mortality in South Korea. *GIScience & Remote Sensing, 61*(1), 2436997.
3. Khan, L., Khan, M., & Ahmad, M. (2025). Socioeconomic and behavioral drivers of geographic disparities in U.S. cardiovascular mortality: A machine learning analysis. *medRxiv*, 2025.09.13.25334113.
4. Ahmed, F., Rony, M. K. K., Hossain, A., Islam, A., Utsho, M. R., Kamal, M. B., Sharmin, S., Tasnim, A. F., & Riipa, M. B. (2024). A systematic review of artificial intelligence models for time-to-event outcome applied in cardiovascular disease risk prediction. *Journal of Medical Systems, 48*, Article 68.
5. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447-453.
6. Mittermaier, M., Raza, M. M., & Kvedar, J. C. (2023). Bias in AI-based models for cardiovascular disease prediction and detection. *npj Cardiovascular Health, 1*, 31.
7. Veenstra, H. G. C., Mittermaier, M., & Petersen, A. (2024). Artificial intelligence bias in cardiovascular medicine. *The Lancet Digital Health, 6*(10), e742-e752.
8. Smedley, B. D., Stith, A. Y., & Nelson, A. R. (Eds.). (2003). *Unequal treatment: Confronting racial and ethnic disparities in health care*. National Academies Press.
9. Hong, C., Youn, S., Yoon, J., et al. (2023). Machine learning models for stroke prediction compared with existing stroke-prediction algorithms and with the pooled cohort equation. *American Journal of Epidemiology, 192*(4), 621-632.
10. Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *The Annals of Applied Statistics, 2*(3), 841-860.

11. Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., & Kluger, Y. (2018). DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Medical Research Methodology, 18*(1), 24.

12. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4765-4774.

13. World Health Organization. (2021). *Ethics and governance of artificial intelligence for health*. World Health Organization.

14. U.S. Food and Drug Administration. (2021). *Artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) action plan*. FDA.

15. Collins, G. S., Dhiman, P., Andaur Navarro, C. L., et al. (2024). Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *Diagnostic and Prognostic Research, 8*, 1.