



(RESEARCH ARTICLE)



PREDICTING HEART DISEASE RISK FROM CLINICAL VARIABLES: A GENDER-SPECIFIC MACHINE LEARNING ANALYSIS AMONG HIGH-CHOLESTEROL PATIENTS

Taiwo Samson Adeyemo *

Department of Information Systems. Dakota State University.

* Corresponding Author

ORCID Details

Taiwo Samson Adeyemo: <https://orcid.org/0009-0009-6459-6249>

GSC Advanced Research and Reviews, 2026, 28(02), 151–159

Publication history: Received on 12 July 2026; revised on 22 August 2026; accepted on 24 August 2026

Article DOI: <https://doi.org/10.30574/gscarr.2026.28.2.0204>

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

Cardiovascular disease remains a major cause of mortality and economic burden in the United States. This study developed and evaluated supervised machine learning models to predict heart disease risk from routine clinical variables and tested whether male patients with high cholesterol have higher odds of heart disease than female patients. Using a publicly available clinical dataset of 918 patients, the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework guided exploratory analysis, data cleaning, median imputation of irregular cholesterol values, and feature selection using principal component analysis (PCA) and SelectKBest. Logistic Regression, Support Vector Machine (SVM), and a soft voting ensemble were trained and tuned using grid search with stratified 5-fold cross-validation. The ensemble achieved the highest predictive performance (accuracy = 0.940, F1 = 0.950, receiver operating characteristic area under the curve [ROC-AUC] = 0.958). Logistic Regression achieved comparable performance (accuracy = 0.929, F1 = 0.940, ROC-AUC = 0.958) and was selected for hypothesis testing because of its interpretability. Among patients with cholesterol ≥ 240 mg/dL, male sex was a statistically significant independent predictor of heart disease after controlling for other clinical variables. The findings support sex-specific screening and preventive strategies for high-cholesterol male patients and demonstrate the value of interpretable machine learning models for clinical decision support. Larger externally validated datasets are needed to assess generalizability.

Keywords: Heart Disease Prediction; Machine Learning; Logistic Regression; Clinical Decision Support; Cholesterol; Gender-Specific Risk

1 INTRODUCTION

Cardiovascular disease remains one of the leading causes of death in the United States, most often originating from heart disease. According to the Centers for Disease Control and Prevention, heart disease is the primary cause of death for men, women, and individuals across most racial and ethnic groups, with 919,032 deaths attributed to cardiovascular conditions in 2023 (1). The economic burden is comparably severe, with the cost of heart disease estimated at \$417.9 billion between 2020 and 2021 (1), and separately projected to rise from roughly \$400 billion in 2020 to as much as

\$1,344 billion by 2050, a sum approximately equivalent to 20% of the federal government's FY2024 budget (2). Globally, the American Heart Association reports that cardiovascular disease accounts for nearly 19.9 million deaths annually (3), and the CDC identifies high blood pressure, high cholesterol, and lifestyle behaviors such as smoking and poor diet as its primary drivers in the United States.

Although recent research indicates a moderate decline in overall heart disease prevalence, there is a growing burden of mortality linked to other related cardiac conditions, underscoring the continued need for targeted intervention (2). Cardiovascular risk factors do not affect all patients equally: prior research has found that men show higher rates of all-cause and cardiovascular mortality compared to women (4). Despite this, many predictive models for heart disease treat risk classification as a general problem, without systematically examining how specific clinical risk factors, such as cholesterol, interact with demographic variables, such as sex, to produce differential risk.

This study addresses that gap. Using a dataset of routine clinical variables collected from 918 patients, the study develops a predictive model to classify heart disease risk and specifically tests whether male patients with high cholesterol are more likely to develop heart disease than female patients with high cholesterol. The goal is twofold: first, to build an accurate and interpretable predictive model that can support early clinical screening, and second, to generate evidence on a specific, clinically actionable risk pattern, sex-specific risk among high-cholesterol patients, that can inform targeted prevention and resource allocation.

1.1 Research gap and objectives

Existing predictive models for heart disease have largely focused on maximizing classification accuracy across a population, often treating demographic and clinical variables as interchangeable predictors rather than examining specific, clinically meaningful interactions between them. While numerous predictive models have been developed to classify or detect heart disease, relatively few have gone further to provide actionable insights into the specific clinical variables that drive disease progression. This general-purpose approach can obscure differential risk patterns that matter for clinical decision-making, such as whether a specific combination of risk factors, in this case elevated cholesterol and male sex, is associated with disproportionately higher risk.

This study responds to that gap by pursuing two connected objectives: (1) developing a predictive model for heart disease using routine clinical variables, and (2) testing a gender-specific classification hypothesis within the subgroup of patients with high cholesterol. This matters because sex-specific evidence of this kind can directly inform clinical prioritization, guide resource allocation toward higher-risk subgroups, and support more targeted treatment planning, rather than relying on undifferentiated population-level risk estimates.

1.2 Research question and hypotheses

The study is guided by the following research question:

RQ: Does sex play a crucial role in heart disease among patients with high cholesterol? Specifically, are male patients with high cholesterol more likely to have heart disease than female patients with high cholesterol?

This translates into the following statistical hypotheses, tested within the subgroup of patients with high cholesterol (≥ 240 mg/dL):

- H_0 : Male patients are no more likely than female patients to develop heart disease when cholesterol is high.
- H_1 : Male patients are more likely than female patients to develop heart disease when cholesterol is high.

2 MATERIALS

2.1 Data source, description, and provenance

The study uses a publicly available clinical dataset retrieved from Kaggle.com, where it was first uploaded approximately two years prior to this analysis. The dataset is drawn from real clinical records of patients who received cardiovascular care and includes 918 patients with 11 predictor variables and one binary target variable, HeartDisease, indicating the presence (1) or absence (0) of heart disease. The dataset contains a mix of numerical and categorical diagnostic factors commonly used by clinicians, including age, blood pressure, cholesterol level, resting electrocardiogram (ECG) results, and exercise-related responses, all of which are clinically meaningful for this type of research. Because the dataset reflects only the group of patients from which it was originally collected, some degree of sampling bias may be present, and observed patterns may not perfectly generalize to other populations.

2.2 Data dictionary

The dataset comprises the following variables: Age (numeric), Sex (categorical), ChestPainType (categorical: ATA, NAP, ASY, TA), RestingBP (numeric, resting blood pressure), Cholesterol (numeric, mg/dL), FastingBS (binary, fasting blood sugar indicator), RestingECG (categorical: Normal, ST, LVH), MaxHR (numeric, maximum heart rate achieved during exercise testing), ExerciseAngina (binary, exercise-induced angina), Oldpeak (numeric, ST-segment depression after exercise relative to rest), ST_Slope (categorical: Up, Flat, Down), and HeartDisease (binary target variable).

To support the study's hypothesis, Sex was re-coded into a binary dummy variable (Sex_Male: Male = 1, Female = 0), and the original categorical Sex field was dropped to avoid redundancy, since the hypothesis required isolating the effect of cholesterol separately for each sex.

3 METHODOLOGY AND FRAMEWORK

The study follows the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework, encompassing exploratory data analysis (EDA), data cleaning, correlation analysis, dimensionality reduction via Principal Component Analysis (PCA), feature selection, classification model testing, hyperparameter tuning, and final model selection. The intended use of the resulting model is to provide a risk score and risk label that can support clinicians in early screening and decision-making, functioning as a decision-support tool rather than a diagnostic replacement for clinical judgment.



Figure 1: Methodology workflow

3.1 Exploratory data analysis

Initial inspection confirmed no missing values and no duplicate records across the 918 observations. Descriptive statistics showed a mean patient age of 53.5 years (range 28–77), a mean resting blood pressure of 132.4 mmHg, a mean (uncorrected) cholesterol level of 198.8 mg/dL, and a positive class prevalence of 55.3% (HeartDisease = 1).

Categorical distributions showed that asymptomatic chest pain (ASY) was by far the most frequent chest pain classification in the sample, followed by non-anginal pain (NAP), atypical angina (ATA), and typical angina (TA), which was least frequent. The majority of patients had a Normal resting ECG, with smaller proportions showing ST-segment abnormality or left ventricular hypertrophy (LVH). A Flat ST-segment slope was the most frequent ST_Slope classification, followed by an Up slope, with a Down slope observed least often. Roughly three in five patients did not report exercise-induced angina.

Examining chest pain and cholesterol by sex, male patients showed a markedly higher frequency of asymptomatic chest pain (ASY) than female patients, and the distribution of cholesterol values was both higher and more dispersed among male patients than female patients. This pattern directly supported the study's focus on the interaction between sex and cholesterol in the research hypothesis.

The high prevalence of asymptomatic chest pain (ASY) in the sample is clinically significant, as this classification represents a particularly dangerous presentation in which patients may exhibit no overt symptoms despite underlying

heart disease (5). The mean age of 53.5 years also aligns with the age range in which heart disease typically becomes clinically manifest; heart disease death rates among adults aged 45–60 increased by 1.5% between 2011 and 2016, and more patients in their 40s and 50s are now experiencing heart attacks

3.2 Data cleaning

Boxplot inspection identified outliers in RestingBP, Cholesterol, MaxHR, and Oldpeak. The outliers observed in RestingBP, MaxHR, and Oldpeak were judged moderate and plausibly reflective of real clinical variation. The Cholesterol variable, however, contained a substantial number of biologically implausible zero values, likely reflecting missing data at the time of entry, alongside extreme values exceeding 300 mg/dL that surpass typical clinical ranges.

Because the volume of affected records made deletion impractical without risking the loss of otherwise valid data, these irregular cholesterol values (equal to zero or above 300 mg/dL) were corrected using median imputation, calculated from the subset of biologically plausible cholesterol values (greater than 0 and less than or equal to 300 mg/dL). This approach is consistent with recommended practice for handling implausible or outlying clinical values, which can be addressed through deletion or imputation using the mean, median, or mode depending on the analytic context (6,7). Median imputation was selected to provide a more balanced and representative treatment of the affected values than mean imputation, which is more sensitive to the extreme values already identified as problematic.

3.3 Correlation analysis

Pairwise correlation analysis among numeric predictors found that Age was positively correlated with Oldpeak ($r = 0.26$) and HeartDisease ($r = 0.28$), and negatively correlated with MaxHR ($r = -0.38$). MaxHR showed the strongest negative correlation with HeartDisease ($r = -0.40$), while Oldpeak showed the strongest positive correlation with HeartDisease ($r = 0.40$). Cholesterol was negatively correlated with HeartDisease ($r = -0.23$), and FastingBS was positively correlated with HeartDisease ($r = 0.27$). Additional heatmap analysis found that biological sex was positively correlated with HeartDisease ($r = 0.31$), and chest pain type was correlated with both HeartDisease ($r = -0.39$) and, weaklier, Age ($r = -0.08$). Further inspection identified moderate associations between Oldpeak and ST_Slope ($r = -0.50$), ExerciseAngina and ST_Slope ($r = -0.43$), and Oldpeak and ExerciseAngina ($r = 0.41$). All pairwise correlations remained below the conventional 0.70 threshold associated with problematic multicollinearity, supporting the joint inclusion of these variables in subsequent modeling, subject to confirmation through PCA and feature-selection procedures.

3.4 Feature selection and dimensionality reduction

Feature selection used a triangulated approach combining two complementary methods. Categorical variables were first one-hot encoded, and all predictors were standardized prior to analysis.

Principal Component Analysis (PCA) was applied to identify which variables contributed most to the components explaining the greatest share of variance in the dataset. The first eight principal components collectively explained approximately 74% of total variance. Reviewing the top-loading variables across these eight components, ST_Slope, ExerciseAngina, and MaxHR loaded most heavily on the first component; Cholesterol, RestingECG, and Oldpeak loaded most heavily on the second; and RestingBP, Age, and ChestPainType loaded most heavily on the third, with Sex emerging as the dominant loading on the fifth component and Cholesterol re-emerging as the dominant loading on the eighth.

In parallel, univariate feature selection (SelectKBest, using the ANOVA F-statistic) ranked predictors by their individual statistical association with the HeartDisease outcome, from which the top eight were selected: Age, MaxHR, Oldpeak, Sex_Male, ChestPainType (ATA), ExerciseAngina, and the ST_Slope categories (Flat and Up).

Comparing the variables identified through PCA against those identified through univariate feature selection, the overlapping and complementary variables produced a final predictor set of eight conceptual variables: Age, RestingBP, Cholesterol, FastingBS, MaxHR, ChestPainType, Sex, and ST_Slope. Notably, Sex and Cholesterol, the two variables central to the study's hypothesis, were supported by both selection approaches, increasing confidence in their inclusion and relevance to the final model. This triangulated approach strengthens the reliability of variable selection by integrating complementary analytical perspectives (8)

3.5 Feature transformation and encoding

Categorical variables were numerically encoded for modeling: Sex (Female = 0, Male = 1); ChestPainType (ASY = 0, ATA = 1, NAP = 2, TA = 3); and ST_Slope (Down = 0, Up = 1, Flat = 2), with ChestPainType and ST_Slope also represented via one-hot dummy variables in the final feature matrix. Variables not retained in the final feature set, including RestingECG and, in the final model, Oldpeak, were excluded from subsequent modeling. To directly test the study's hypothesis, an interaction variable combining sex and cholesterol was constructed to evaluate the joint predictive direction of these two variables relative to their individual effects.

3.6 Modeling approach

The prediction task was framed as supervised binary classification, with HeartDisease (0/1) as the outcome variable. The data were split into training and validation sets (80/20) using stratified sampling. Three classification approaches were compared:

Logistic Regression was used as an interpretable baseline classifier. Logistic Regression assumes a binary outcome, independent observations, an approximately linear relationship between predictors and the log-odds of the outcome, and the absence of severe multicollinearity (9). It is well suited to this dataset because the outcome is binary, the predictor set mixes numeric and categorical clinical variables, and it produces directly interpretable coefficients and odds ratios, which are essential for testing whether sex and cholesterol are significantly associated with heart disease (10)

Support Vector Machine (SVM), implemented with a radial basis function (RBF) kernel, was used to capture potentially non-linear relationships among predictors. SVM assumes that classes can be separated by a decision boundary or hyperplane and that the chosen kernel and hyperparameters (C, gamma) are appropriately specified, but it does not require normality or strict linearity (11). This flexibility makes it well suited to structured clinical data with potentially non-linear boundaries between patients with and without heart disease.

Soft Voting Ensemble. A voting ensemble combining the Logistic Regression and SVM base learners was constructed, averaging their predicted class probabilities to produce a final prediction. The ensemble assumes that its base learners are trained on the same data, produce probabilistic outputs, and make partially independent errors, such that aggregating their predictions can outperform either model individually. This approach is well suited to the dataset because Logistic Regression captures clear, approximately linear risk patterns while SVM can capture more complex, non-linear boundaries; combining their predicted probabilities allows the ensemble to draw on both perspectives.

Model evaluation used stratified 5-fold cross-validation combined with grid search (GridSearchCV, optimized for ROC-AUC) for hyperparameter tuning. Performance was assessed using accuracy, precision, recall, F1 score, ROC-AUC, and the confusion matrix on the held-out validation set.

4 RESULTS AND DISCUSSION

4.1 Model performance comparison

All three models achieved strong discriminative performance on the validation set.

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	0.929	0.953	0.927	0.940	0.958
SVM (RBF/linear)	0.886	0.901	0.909	0.905	0.953
Soft Voting Ensemble (Logistic + SVM)	0.940	0.946	0.955	0.950	0.958

The ensemble model achieved the strongest pure predictive performance, correctly classifying the highest proportion of patients while missing the fewest true positive heart disease cases (confusion matrix: 68 true negatives, 6 false positives, 5 false negatives, 105 true positives). Logistic Regression performed very closely behind on nearly every metric, with balanced precision and recall for both classes (confusion matrix: 69 true negatives, 5 false positives, 8 false negatives, 102 true positives). The SVM model was the weakest of the three across all metrics, though it still achieved a strong ROC-AUC of 0.953.

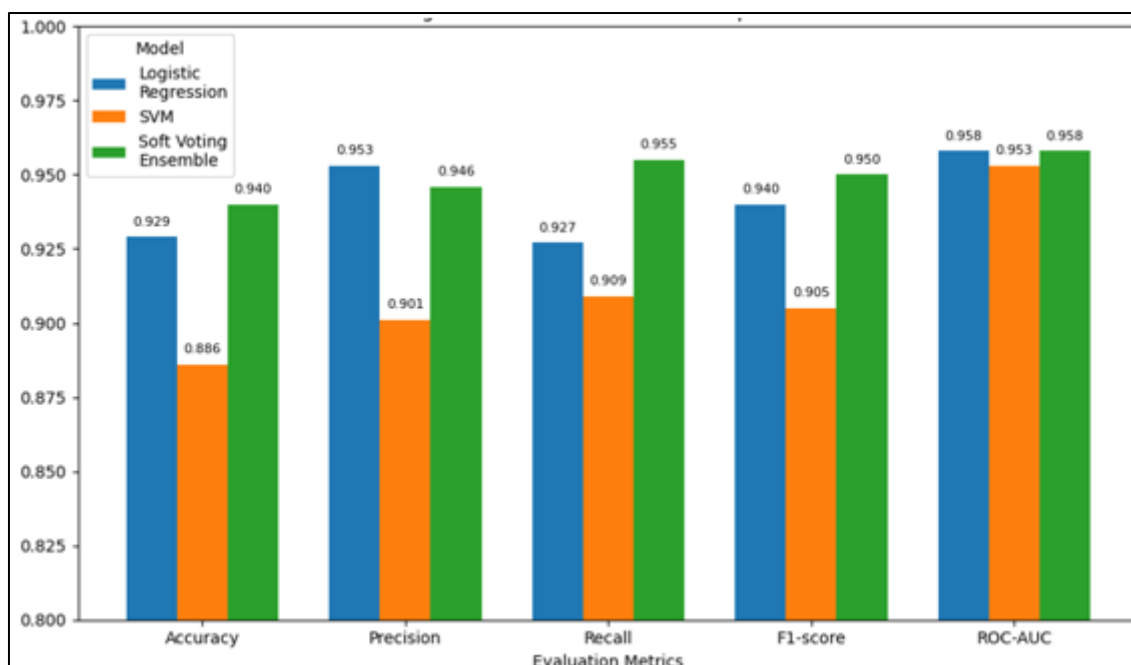


Figure 2: Model Performance Comparison

4.2 Hyperparameter tuning

Grid search with stratified 5-fold cross-validation was used to tune each model, optimizing for ROC-AUC. The optimal Logistic Regression configuration used an L2 penalty with $C = 1$. The optimal SVM configuration used a linear kernel with $C = 0.1$. The optimal ensemble configuration weighted the Logistic Regression and SVM base learners 2:1. Grid search with cross-validation was selected as the tuning approach because it exhaustively evaluates all specified hyperparameter combinations while using cross-validation to produce a more robust estimate of model performance, reducing the risk of overfitting to a single train/validation split

4.3 ROC curve analysis

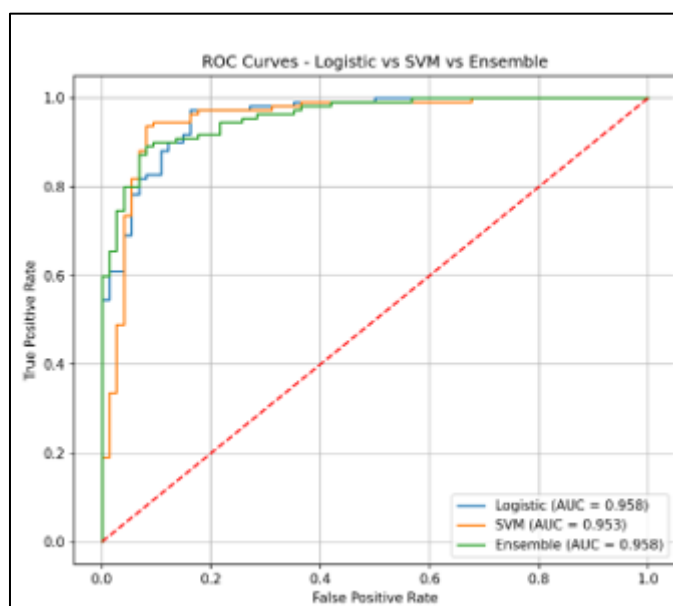


Figure 3: ROC curve comparison

Receiver operating characteristic (ROC) curve analysis confirmed the pattern observed in the performance table. The Logistic Regression and ensemble ROC curves closely tracked one another near the top-left corner of the plot, both achieving an AUC of approximately 0.958, while the SVM curve trailed slightly with an AUC of approximately 0.953. Across a broad range of decision thresholds (approximately 0.3–0.6), both the Logistic Regression and ensemble models

maintained the lowest misclassification error, while the SVM model showed comparatively higher error across most thresholds.

4.4 Final model selection

Although the tuned ensemble model achieved the strongest pure predictive performance, Logistic Regression was selected as the primary model for hypothesis testing and clinical interpretation. This decision reflects a deliberate prioritization of interpretability: Logistic Regression provides directly interpretable coefficients and odds ratios for individual predictors, which are essential for addressing the study's research hypothesis regarding the role of sex in heart disease risk among high-cholesterol patients, and for identifying which clinical variables are statistically significant drivers of risk. The ensemble model, by contrast, remains the stronger choice when the objective is to maximum predictive performance alone, since combining two base learners with complementary decision boundaries reduces individual model weaknesses and yields a marginally more robust classifier for pure prediction purposes.

4.5 Hypothesis testing results

To test the study's central hypothesis, analysis was restricted to the subgroup of patients with high cholesterol (≥ 240 mg/dL). Within this subgroup, the selected Logistic Regression model was estimated using the final eight predictors, with HeartDisease as the outcome.

Among high-cholesterol patients, three variables, Age, ChestPainType, and Sex_Male, were statistically significant predictors at the 5% significance level. The coefficient for Sex_Male was positive and highly significant, indicating that male patients with high cholesterol had substantially higher odds of heart disease than female patients with high cholesterol, even after controlling for other clinical variables in the model. These results support rejecting the null hypothesis in favor of the alternative hypothesis: among high-cholesterol patients, male sex is associated with meaningfully elevated odds of heart disease.

5 DISCUSSION

The findings indicate that, among patients with high cholesterol, male sex functions as a strong and independent predictor of heart disease, with male patients showing several times higher odds of heart disease than female patients even after accounting for other clinical factors included in the model. This result moves the analysis beyond general-purpose risk classification by identifying a specific, clinically actionable, sex-differentiated risk pattern within a defined clinical subgroup, and is consistent with broader epidemiological evidence that men experience higher rates of cardiovascular mortality than women.

From a practical standpoint, these findings suggest that clinicians and health system administrators may benefit from prioritizing targeted screening and preventive interventions for male patients presenting with elevated cholesterol, rather than applying uniform screening protocols across sexes. The high frequency of asymptomatic chest pain (ASY) observed in the sample, particularly among male patients, reinforces the clinical importance of proactive, risk-based screening: because ASY presentations can mask underlying disease, relying on symptom-driven detection alone may miss high-risk patients who would otherwise be identified through a structured risk score.

More broadly, the study illustrates the value of combining predictive accuracy with interpretability in clinical machine learning applications. Although the ensemble model achieved marginally stronger pure predictive metrics, the decision to select Logistic Regression as the primary analytic model reflects a broader principle relevant to health technology design: predictive models intended to support, rather than replace, clinical judgment should be evaluated not only on statistical performance but also on whether their outputs, in this case, interpretable odds ratios, can be meaningfully understood and acted upon by the clinicians who will use them.

Study limitations and lessons learned

Several limitations should be considered when interpreting these findings. First, the study relies on a relatively small dataset of 918 patients, which may limit the generalizability of the results to broader or more diverse patient populations. Second, the provenance and original data collection process underlying the dataset could not be independently verified, introducing some uncertainty regarding data quality and representativeness, and raising the possibility of sampling bias toward the population from which the data were originally drawn. Third, the study's scope was limited to a defined set of clinical variables, and it did not include potentially relevant lifestyle or hereditary factors such as smoking status, body mass index (BMI), family history, or dietary patterns, which are known contributors to cardiovascular risk.

A key lesson learned during the study was the importance of data quality: correcting biologically implausible cholesterol values (zeros and values above 300 mg/dL) through median imputation meaningfully improved the reliability of the

resulting model, underscoring that careful data preparation is often as consequential to model quality as model selection itself. A second lesson was the value of triangulating multiple feature selection methods: comparing PCA loadings against univariate feature selection results provided greater confidence in the retained predictor set, particularly for the two variables central to the study's hypothesis.

Future work should address these limitations by incorporating larger or external datasets, adding lifestyle and hereditary variables such as smoking, BMI, and family history to improve model generalizability, and expanding the set of models evaluated to further validate the robustness of the observed sex-cholesterol interaction.

6 CONCLUSION

This study developed and evaluated a machine learning model to predict heart disease risk from routine clinical variables and tested a gender-specific hypothesis regarding the relationship between cholesterol and heart disease risk. Using a triangulated feature selection approach and comparing Logistic Regression, SVM, and a soft voting ensemble classifier, all tuned via grid search with cross-validation, the study found strong overall predictive performance (ROC-AUC 0.953–0.958 across models) and selected Logistic Regression as the primary model based on its combination of strong performance and interpretability. Restricting analysis to high-cholesterol patients, the study found that male sex was a significant and independent predictor of heart disease, with male patients showing substantially higher odds of heart disease than female patients. These findings support targeted, sex-specific screening and preventive strategies for high-cholesterol male patients and demonstrate the value of interpretable predictive modeling as a foundation for clinical decision-support tools. Future research should extend this work using larger, externally validated datasets and an expanded set of clinical and behavioral variables.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Disclosure of conflict of interest

The author declares no financial or non-financial conflicts of interest.

Statement of ethical approval

This study analyzed a publicly available secondary clinical dataset and did not involve direct participant recruitment or intervention by the author.

REFERENCES

- [1] CDC. Heart Disease [Internet]. 2026. Heart Disease Facts. Available from: <https://www.cdc.gov/heart-disease/data-research/facts-stats/index.html>
- [2] Kazi DS, Elkind MSV, Deutsch A, Dowd WN, Heidenreich P, Khavjou O, et al. Forecasting the Economic Burden of Cardiovascular Disease and Stroke in the United States Through 2050: A Presidential Advisory From the American Heart Association. *Circulation*. 2024 Jul 23;150(4). doi:10.1161/CIR.0000000000001258
- [3] Heart Disease: Still Leading Cause Of Death – And Preventable | Premier Health [Internet]. Available from: <https://www.premierhealth.com/your-health/articles/health-topics/heart-disease-still-leading-cause-of-death-and-preventable>
- [4] Lv Y, Cao X, Yu K, Pu J, Tang Z, Wei N, et al. Gender differences in all-cause and cardiovascular mortality among US adults: from NHANES 2005–2018. *Front Cardiovasc Med*. 2024 Feb 14;11. doi:10.3389/fcvm.2024.1283132
- [5] Silent Heart Attack: Signs, Symptoms & Recovery [Internet]. Available from: https://my.clevelandclinic.org/health/diseases/21630-silent-heart-attack?utm_source=chatgpt.com
- [6] Hirol S. Navigating Outliers for Accurate Data Analysis & Decisions [Internet]. 2023. Available from: <https://statusneo.com/navigating-outliers-for-accurate-data-analysis-decisions/>
- [7] Syam SS. Understanding and Handling Outliers in Data Analysis. Medium [Internet]. Available from: <https://medium.com/@heysan/understanding-and-handling-outliers-in-data-analysis-727a768650fe>
- [8] Cangelosi R, Goriely A. Component retention in principal component analysis with application to cDNA microarray data. *Biol Direct*. 2007 Jan 17;2(1):2. doi:10.1186/1745-6150-2-2
- [9] Harris JK. Primer on binary logistic regression. *Fam Med Community Health*. 2021;9(Suppl 1):e001290.

- [10] Mishra A. Logistic Regression: The baseline classification model Explained. Medium [Internet]. 2021 Dec 12 [cited 2026 Aug 8]. Available from: <https://medium.com/@ashishmishra2k19/logistic-regression-the-baseline-classification-model-explained-89915f5b13b2>
- [11] Saini A. Support vector machine (svm): A complete guide for beginners. Anal Com. 2021.