JOURNAL OF
EMERGING INVESTIGATORS

Article

Predictive modeling of cardiovascular disease using exercise-based electrocardiography

Sahil Hazari[1], Sunil Hazari[2]
[1] Biomedical STEM Program, Walton High School, Marietta, Georgia
[2] University of West Georgia, Carrollton, Georgia

SUMMARY
Cardiovascular disease is the leading cause of death globally, highlighting the need for research focused on early detection and risk management to enhance patient outcomes. This study investigated the association between eleven cardiovascular risk factors and the presence of heart disease, hypothesizing a significant relationship between these factors and heart disease. The goal of this study was to create a model that could predict cardiovascular disease outcomes using exercise-based electrocardiography. We hypothesized that there would be significant associations between various risk factors and the presence of heart disease. Specifically, we aimed to investigate the relationships between demographic factors (age and sex), clinical symptoms (chest pain type), and physiological measurements (resting blood pressure, cholesterol level, fasting blood sugar, resting ECG results, maximum heart rate achieved, exercise-induced angina, oldpeak, and ST slope) with the presence or absence of heart disease. This approach allowed us to examine predictors across different categories that may affect cardiovascular health. We identified seven significant predictors of heart disease presence from the initial eleven risk factors, utilizing non-parametric data and multivariate analysis. Our findings were compared with contemporary research to evaluate the changes in heart disease risk factors over time. The results indicate that advancements in imaging, three-dimensional electrocardiography during exercise, and automated reporting systems are improving the diagnostic capabilities of traditional exercise-based electrocardiography. This research contributes to the ongoing efforts to improve early detection of cardiovascular disease.

INTRODUCTION
Cardiovascular disease has been the leading cause of death globally for decades. There were 14 million deaths worldwide related to cardiovascular disease in 2000 (1). This number increased to 18.5 million in 2019, with a nearly identical number of deaths among males and females (2). In 2020, there were 928,741 deaths related to cardiovascular disease in the United States, which accounted for 41.2% of deaths in the United States attributable to cardiovascular disease (2). Death rates from cardiovascular diseases have fallen in many countries due to the increase in awareness about cardiovascular health, advances in cardiovascular treatment, better screening and monitoring, and public policy initiatives (3). The decline in death rates can be attributed to the above factors that have made it possible to access affordable cardiovascular treatment (3).

The term cardiovascular disease is an umbrella term that includes various heart diseases such as arrhythmia, atherosclerosis, heart valve diseases, cardiomyopathy, pericarditis, myocardial infarction, hypertrophic cardiomyopathy, and congenital heart defects (4, 5). These cardiovascular diseases occur when blood vessels cannot supply blood to the heart muscle. As a result of this blockage, the patient experiences symptoms such as angina (chest pain), shortness of breath, or heart attack (6). Researches have identified some common risk factors for heart disease. These include lifestyle choices such as smoking and sedentary behavior, as well as cardiometabolic factors such as high blood pressure, high cholesterol, obesity, and diabetes (7, 8). However, additional research is needed on other risk factors that might increase the probability of heart disease.

Since innovative methods such as vascular imaging and three-dimensional echocardiography for diagnosing and treating cardiovascular diseases are not universally available, healthcare providers rely on traditional approaches such as exercise-based electrocardiograph guidelines (9). Public policy can be developed to educate the population on early detection and risk management. These initiatives have been identified as important components in reducing the incidence of cardiovascular diseases to improve patient outcomes (10, 11). Preventative measures to combat risk factors have significantly reduced the mortality rate, but more research is needed to diagnose and prevent cardiovascular diseases (12).

The use of age and sex variables in combination with clinical factors can yield a better overall analysis and understanding of cardiovascular health (13, 14). At the fundamental level, these variables can provide a non-invasive observational approach to assessing cardiovascular risk. The progression of age represents the passage through different developmental stages, which can put older individuals at risk of cardiovascular disease (15). Also, the biological differences between males and females can play a crucial role in shaping and influencing both physical and psychological processes in individuals, so research is needed to further investigate the impact factors such as age and sex in diagnosis of cardiovascular disease (16).

Chest pain, often referred to as angina, is a common symptom of heart disease and is an indication that the heart is not receiving oxygen-rich blood (17). Researchers have used chest pain as a key diagnostic tool, because it may signal that

https://doi.org/10.59720/24-239

the heart is under stress and indicate a serious cardiovascular problem. Understanding the specific relationships between different chest pain types and diseases can improve diagnostic accuracy and allow patients to receive better treatment (18). Chest pain perception can vary between sexes, meaning that sex-specific associations of chest pain type and heart disease can lead to a more accurate prediction of the presence or absence of heart disease (19). In the dataset used in this study, four types of chest pains (typical angina, atypical angina, non-anginal pain, and asymptomatic chest pain) were used as biomarkers of heart disease.

The first objective of this study was to develop a model for predicting the outcome of cardiovascular disease using exercise-based electrocardiography (EBE). A standard electrocardiograph (ECG) measures heart activity at rest, while EBE measures heart activity during physical exertion. The second objective was to establish a baseline for comparison with more recent data. By using a comparative approach, we could assess the evolution of heart disease risk factors, which could contribute to ongoing efforts to improve the early detection of cardiovascular disease. Most of the existing literature on cardiovascular disease focuses on current data, often ignoring historical trends that can provide valuable insights (20). Comparative analysis can help clinicians and medical researchers track the evolution of cardiovascular risk factors over time and identify predictors that have remained relevant and important in current cardiovascular medical research (21).

Predictive modeling using data and statistical tools has been used extensively in medical research to predict outcomes based on patterns in data (22, 23). Predictive modeling can provide insights for early detection, risk assessment, and personalized treatment strategies. For cardiac studies, variables with biological, physiological, and developmental implications can affect a wide range of outcomes (24). By using this data, clinicians can make informed decisions about patient care management (25). We used logistic regression in this study because it is well suited for binary classification of dichotomous outcomes (presence or absence of heart disease).

We hypothesized that there are significant associations between various risk factors and the presence of heart disease. Specifically, we aimed to investigate the relationships between demographic factors (age and sex), clinical symptoms (chest pain type), and physiological measurements (resting blood pressure, cholesterol level, fasting blood sugar, resting ECG results, maximum heart rate achieved, exercise-induced angina, oldpeak, and ST slope) with the presence or absence of heart disease. This approach allowed us to examine predictors across different categories that may affect cardiovascular health.

RESULTS

We hypothesized that there would be a significant association between cardiovascular risk factors and heart disease. To test this hypothesis, we analyzed a dataset which included data about 918 patients from 5 locations. Our research focused on examining the impact of various risk factors that could help diagnose cardiovascular disease. In the heart disease dataset that was used in this study, there were 193 females (21%), 725 males (79%), and the age range of participants was from 28 to 77 years (**Table 1**).

Category	Subcategory	Count	%
Sex	Male	725	78.98
	Female	193	21.02
Age Group	<30	3	0.33
	30-39	64	6.97
	40-49	220	23.96
	50-59	334	36.38
	60-69	238	25.93
	70+	59	6.43
Chest Pain Type	Typical Angina	39	4.25
	Atypical Angina	235	25.6
	Non-Anginal pain	193	21.02
	Asymptomatic	451	49.13
Heart Disease	Absent	412	44.88
	Present	506	55.12

Table 1: Demographic information of the patients from four locations sourced from UCI Machine Learning Repository (n = 918). From 1190 original cases, we curated 918 after eliminated cases that contained missing or inaccurate information.

Sex and age categories taken together are not significantly associated with heart disease

We used age and sex as initial variables for investigation in this study. An investigation of distribution by sex and age for the presence of heart disease shows that the prevalence of heart disease is higher for males across all age categories (**Figure 1**). To analyze these variables, we used a chi-square test of independence to examine the association between age categories and sex. There was no significant association between sex and age category combined regarding the presence of heart disease ($\chi^2 = 6.719$, $df = 918$, $p = 0.081$), but we observed in the dataset that males had a higher prevalence of heart disease compared to females. Therefore, sex and age categories together did not have a significant combined effect on the presence of heart disease.

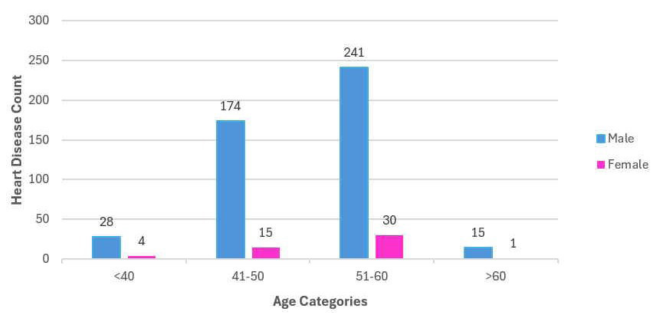

Figure 1: Heart disease by age distribution and sex (n = 508). Males (n = 458) across all age categories had a higher prevalence of heart disease than females (n = 50). The 51-60 age category showed the highest prevalence of heart disease for males as well as females. A chi-square test showed males had a statistically significant higher prevalence of heart disease compared to females $\chi^2(1, n = 918) = 1004.74$, $p < 0.001$

https://doi.org/10.59720/24-239

Chest pain types have a significant association with heart disease

Next, we wanted to identify whether certain types of chest pain are more strongly associated with heart disease, which can aid in diagnosis and patient management. The types of pains examined were typical angina, atypical angina, non-anginal pain, and asymptomatic pain. Typical angina refers to chest discomfort such as pressure or squeezing as a result of exertion or stress. Atypical angina presents as chest discomfort in the form of a sharp pain that may not be related to physical exertion. Non-anginal pain may not be related to cardiac disease as it can be caused by gastric, muscle, or pulmonary issues. Asymptomatic pain (also called silent heart attack) is when a patient has a heart condition but there are no noticeable symptoms.

We used a chi-square test of independence to examine the relationship between chest pain type and heart disease. There was a statistically significant association between chest pain type and heart disease ($\chi^2 = 268.067$, $df = 3$, $p < 0.001$) for both males and females. This suggests that the type of chest pain experienced is significantly related to the presence or absence of heart disease. A closer look at the data reveals different patterns between chest pain type and sex. Males with asymptomatic chest pain had a higher prevalence of heart disease as compared to males with typical angina who showed a lower prevalence of heart disease (**Figure 2**). These findings suggest that chest pain is an important factor to consider in the diagnosis and risk assessment of heart disease.

Predictors of heart disease

We next investigated the effect of all 11 predictor variables on the heart disease classification outcome. Based on the types of variables (nominal and continuous) that were present in this study, binary logistic regression was used for analysis. Binary logistic regression calculates the probability of success over the probability of failure by providing the results in the form of an odds ratio (26). The expected outcome was heart disease, coded as 1, and the absence of heart disease was coded as 0. The objective was to identify the most significant factors contributing to heart disease. The coefficients, standard error, z-statistic, Wald, p-values, and Exp (Coefficients) collectively provide information about predictive power of the statistical model (**Table 2**).

The significant predictors of heart disease, in order of impact based on odds ratio values Exp(B), were ST slope, sex, exercise angina, fasting blood sugar, chest pain type, oldpeak, and serum cholesterol. ST slope in an ECG measures the interval between the stages of electrical activity in the ventricles of the heart. Oldpeak is a measurement of the ST segment depression that is induced by exercise relative to rest. Serum cholesterol is a measurement of the total cholesterol concentration per deciliter of blood. The non-significant variables were resting ECG, age, resting blood pressure, and maximum heart rate. Resting ECG refers to the results of an ECG test, classified as normal or abnormal. One unit change in the coefficient (B) predictor value would cause a change in the Exp(B) log odds outcome probability. The heart disease target group probability would be equal to the non-target normal group probability if the odds ratio was equal to 1. For all predictors with an odds ratio greater than 1, the event would be likely to occur. As an example, the odds of a person having heart disease due to high fasting blood sugar would be 3.013 times higher than a normal person who does not have high fasting blood sugar.

In the null model, no predictors are included, whereas the estimated model includes the predictors and the outcome. High loglikelihood and low deviance parameters show a strong alignment of the model with the data points (27). The model met both conditions in this study (**Table 3**). The model fit summary also shows the R-square values that range from 0.0472 to 0.639. The R-square value (also called the coefficient of determination) indicates how well the model is able to make predictions based on the dataset. Nagelkerke's R-square is the adjusted version of Cox & Snell's R-square, which showed 63.9% of the change in the criterion variable could be accounted for by the predictor variables in the model (28). The quality of the statistical model was further confirmed by reviewing the AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion). Lower values in the

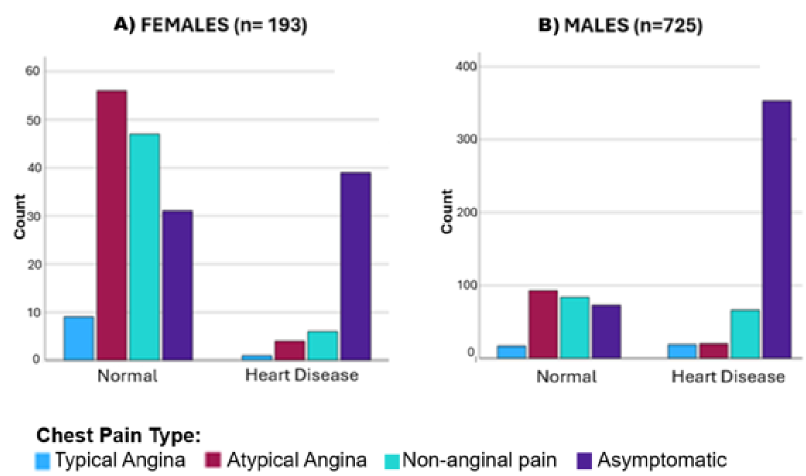

Figure 2: Number of individuals who were reported to experience one of four types of chest pain (n = 918). A) Females with no heart disease are more likely to have atypical angina and non-anginal pain, but females with heart disease are more likely to be asymptomatic. **B)** Males with asymptomatic chest pain have a higher prevalence of heart disease. The chi-square test showed significant association between sex and chest pain type ($\chi2 = 36.879$, $p < 0.001$).

https://doi.org/10.59720/24-239

	Coefficients B	SE	z-statistic	Wald	p-value	Exp Coefficients Exp (B)
ST Slope	1.597	0.202	7.927	62.835	1.0 x 10^{-308}*	4.940
Age	0.015	0.012	1.189	1.414	0.234	1.015
Chest Pain Type	0.727	0.112	6.464	41.778	1.0 x 10^{-4}*	2.069
Cholesterol	-0.003	0.001	-3.069	9.420	0.002*	0.997
Exercise Angina	1.053	0.231	4.564	20.833	1.0 x 10^{-308}*	2.866
Fasting Blood Sugar	1.103	0.259	4.251	18.074	9.0 x 10^{-4}*	3.013
Max Heart Rate	-0.008	0.005	-1.774	3.146	0.076	0.992
Old Peak	0.350	0.114	3.064	9.387	0.002*	1.419
Resting Blood Pressure	0.005	0.006	0.850	0.722	0.396	1.005
Resting ECG	0.065	0.124	0.520	0.271	0.603	1.067
Sex	1.301	0.254	5.123	26.242	1.0 x 10^{-308}*	3.674
Intercept	-6.300	1.417	-4.446	19.766	0.000	0.002

Table 2: Coefficient Matrix of 11 predictor variables (n = 918). Model parameters showed seven significant predictors (* shows significance $p < 0.01$).

	Null Model	Estimated Model
LogLikelihood	-631.1	-333.4
Deviance	1262.1	666.9
AIC	1264.1	690.9
BIC	1269	748.8
df	917	906
Cox and Snell R^2		0.477
Nagelkerke R^2		0.639
McFadden R^2		0.472

Table 3: Model fit summary of Null model compared to Estimated model for prediction of outcome of heart disease. The Null model is a reference model with no predictors. The estimated model includes predictors to explain the data. The estimated model provided a better model fit for loglikelihood, deviance, AIC (Akaike Information Criterion), and BIC (Bayesian Information Criterion).

estimated model compared to the null model indicate better goodness of fit and performance of the logistic regression for the given set of data (29). The result confirmed our model's reliability and the validity of its predictions.

Predictive model performance

To evaluate the performance of the model, we used a classification table, which is also referred to as a confusion matrix (**Table 4**). For the normal group (no heart disease), the model was able to correctly predict 247 (84.3%) cases, and 63 cases were predicted incorrectly. For the target group of individuals who were diagnosed with heart disease, the model correctly predicted 435 (85.63%) of the cases and incorrectly predicted 73 cases. In both scenarios (normal and heart disease classification), the model prediction was correct the majority of the time. The overall percentage accuracy of the model in classification was calculated to be 85.18%.

DISCUSSION

The purpose of this study was to develop a predictive model for heart disease using 11 variables that were associated with EBE. The results of this study showed the value of using EBE historical data along with other demographic, physiological, and diagnostic factors to create a predictive model for detecting coronary artery disease. By comparing the results of this study with more recent data, it is possible to assess historical heart disease risk factors that have remained consistent. Researchers can further investigate these factors, which can help with early detection of cardiovascular disease. To conduct a comparison with the dataset that was used in this study, we used four current, comprehensive, and highly cited studies that included at least some of the variables of this study.

We compared comparative data analysis that showed evolving trends and patterns for cardiovascular risk factors over time with trends reported in several recent studies. Researchers have attributed an increase in cardiovascular disease to an aging society (2). Advancements in medical and health interventions have counteracted the detrimental effects of an aging population on cardiovascular disease (30). Additional factors such as sedentary behavior, alcohol and tobacco consumption, and high blood pressure also have a high impact on the increase of cardiovascular disease. It is interesting to note that cardiovascular disease showed a declining trend among young adults, but among older adults, a higher prevalence of cardiovascular disease was observed (31). These patterns showing an increase in cardiovascular disease can be explained by an improvement of health standards in contemporary medicine, with guideline-directed therapy reducing the risk of mortality due to cardiovascular disease (32). Previous studies have used demographic variables such as sex and age categories as predictors (4, 8). However, our study did not find a significant combined effect on the presence of heart disease. There is a need to reassess the impact of these demographic factors in risk assessment models. Our study found that males had a higher prevalence of heart disease as compared to females in all age categories. This finding suggests attention needs to be paid to male-specific risk factors, prevention strategies, and biological or lifestyle factors that could have contributed to this disparity. Researchers have recognized personalized medicine an effective approach to enhance patient care as it is able to leverage genetic information, advanced imaging techniques, and computational modeling to provide individualized treatment strategies (33). The broader categories of age and sex can combine with other, more specific, risk factors to show an interaction where personalized medicine can lead to earlier diagnosis. Interaction results in certain risk factors having different levels of influence rather than additive effects for all variables. The significant factors identified in this study can assist medical practitioners in detecting coronary artery disease and guide personalized treatment approaches, which can improve patient outcomes.

Several of the 11 variables used in this study that are relevant in cardiovascular research to predict disease risk were also previously identified as significant risk factors (30). Although males had a higher prevalence of heart disease, the combined effect of age and sex as predictors has diminished in predictive accuracy, perhaps as a result of advancements in medical interventions and changes in lifestyle patterns (30). Our findings that that age and sex were important but not significant predictors contrasts with another report that showed disparities in cardiovascular disease mortality between sexes across regions (2). This may have been due to the characteristics of the sample in our dataset, which may have caused sampling bias. In general, there were higher death rates in males compared to females based on worldwide statistics. The recent finding that cardiovascular risk factors differ by sex with men generally having higher risks is consistent with the results of this study (31). Previous

https://doi.org/10.59720/24-239

	# of cases Predicted value = 0 (normal)	# of cases Predicted value = 1 (heart disease)	% Correct
target = 0	347	63	84.63
target = 1	73	435	85.63

Table 4: Classification table matrix (n = 918). The model was able to correctly identify at a high rate the cases of heart disease and those without (normal). Target 0 indicates no heart disease; target 1 indicates heart disease is present.

studies have highlighted sex disparities in cardiovascular disease prevalence, and have recommended using age as a modulator for more nuanced models that account for the interactions between age and sex (2, 32).

For chest pain type as a predictor of cardiovascular disease, the results of this study found a significant association between all chest pain types and the presence of heart disease. Males experiencing asymptomatic chest pain showed the highest prevalence of heart disease of all the chest pain types. Asymptomatic chest pain means that the chest pain is silent and does not cause anginal symptoms. Asymptomatic chest pain can be used as an indicator of the prevalence of heart disease. This can be diagnosed using screening tests such as ECG. For females, the association between chest pain type and the presence of heart disease was also significant. As compared to males, females showed a stronger association between typical angina, atypical angina, and non-anginal pain and the absence of heart disease. For asymptomatic pain, both sexes show an association with heart disease, but it was stronger for males (82.9%) than females (55.7%). Previous studies have found a higher prevalence of heart disease in individuals with typical angina (30). This counterintuitive finding shows the complexity of predicting the outcome of cardiovascular diseases solely based on chest pain characteristics. Using 2022 data, it had been noted that the complexity of relying on chest pain type to diagnose heart disease needs further research into chest pain characteristics (32). The other studies did not specifically address the issue of chest pain type, so the results of this study can provide valuable insights into developing broad risk assessment models of chest pain type as one of the predictors for diagnosis. Clinical practitioners can educate patients that the absence of heart pain does not indicate the absence of heart disease.

The predictive model used in this study included a combination of parameters under the demographic, symptom/clinical, physiological, and diagnostic categories. The data analysis showed good accuracy in identifying factors that significantly impact the presence of cardiovascular disease. The potential of using EBE data with other factors can enhance the non-invasive detection of coronary artery disease. This combination will help with more timely and informed patient diagnosis and treatment. The results of this study showed the value of using EBE historical data along with other factors to create a predictive model for detecting heart disease. The significant predictors found for the outcome of heart disease were ST slope, sex, exercise angina, fasting blood sugar, chest pain, oldpeak, and cholesterol. Although age was not a significant predictor in the logistic regression model for this dataset, other studies have found that age may interact with other factors in ways that may require further research (34, 35). It is possible that the age variable may be modulated by ethnicity and genetic predisposition (36). As this study and other researchers have reported, the significant role of sex in predicting heart disease suggests that differences such as hormonal, genetic, or behavioral may play a crucial role in heart disease risk (37).

Another predictor found in this study, ST slope, was not shown to have a predictive significance of ST-level changes in a sample of Finnish nationals (38). However, the data was not collected as part of EBE. Rather, the study was conducted using normal electrocardiography. This result could also have been due to variability or contextual differences between the samples. Further research is needed to understand the underlying factors contributing to these differences. For the oldpeak variable that was significant in this study, a review of recent literature revealed an absence of studies that mention this predictor. This gap in the current literature suggests a need to investigate this variable further to confirm the findings of this study.

In addition to blood pressure, cholesterol, and fasting blood sugar, which were significant predictors in this study, new variables such as body mass index and hemoglobin A1C have been identified as significant predictors in other studies (31). For future studies, researchers could use other recent predictors in EBE, such as exercise capacity, diastolic function, and myocardial strain (39, 40). Another study made similar suggestions regarding modern imaging and computational modeling to improve diagnostic accuracy (2). Over the years, predictive models for heart disease have evolved to include different variables due to advances in medical science. Twenty years ago, many experts considered age, sex, cholesterol level, and smoking status to be the main factors that contributed to the risk of heart disease (7). Risk models today incorporate other variables besides previously established risk factors, including biomarkers (e.g., high-sensitivity C-reactive protein), imaging diagnostics, and lifestyle factors (8). We found that cholesterol and fasting blood sugar continue to be important factors, but the data also showed that ST slope and exercise-induced angina are significant and consistent with current findings. This means that while traditional predictors are still useful, additional variables measured through EBE could further improve the assessment of heart disease risk. Newer imaging technologies and biomarkers could further improve diagnostic accuracy. This integrative approach can lead to more personalized treatment strategies and better patient outcomes.

We used a secondary dataset for this study. Secondary data is extensively used in clinical research as it provides advantages such as the ability to source data from hard-to-reach participants (41-43). Cardiovascular disease is a broad term that includes a range of diagnoses related to risk factors, including those that may be congenital or genetic. Also, the dataset used in this study did not look at risk factors that could affect specific diagnoses.

Although the dataset used in this study had cases from multiple regions (Cleveland, Long Beach (VA), Hungary, Switzerland, and the UCI Heart Dataset), individual locations were not identified for each case. Having this information included and coded in the dataset would have enabled further comparative analysis. Since different locations can have unique characteristics such as environmental and socioeconomic factors, these variables could have influenced cardiovascular disease parameters for the cross-sectional

https://doi.org/10.59720/24-239

data collected and compiled in the dataset.

Our study bridged a research gap by comparing historical and contemporary data on cardiovascular disease predictors by using EBE, which continues to be used today for the diagnosis of cardiovascular disease. The continued relevance of significant variables found in this study, such as resting blood pressure, cholesterol, and fasting blood sugar, shows the complexity of cardiovascular disease risk assessment. New insights are needed into sex-specific risk factors, and the diagnostic significance of chest pain types highlights areas for further research and clinical focus. Specifically related to EBE, we found that recent advancements in imaging, three-dimensional echocardiography during exercise, and automated reporting systems are enhancing the use of a traditional exercise-based echocardiography for diagnostics (40). Using modern findings along with historical data can enhance our understanding of heart disease predictors, which can contribute valuable knowledge by aiding practitioners and researchers in diagnosing cardiovascular disease.

METHODS

Data collection

The heart disease dataset was sourced from Kaggle, a repository of datasets in different fields (44). The dataset was located by using keywords such as cardiovascular disease, risk factors, angina, and electrocardiography. These search terms resulted in several datasets, from which we chose the heart disease dataset since it curated multiple other data sets that were available independently but not combined before. The dataset included a sample of patients with and without cardiovascular disease from five different locations. Data used in this study was collected using electrocardiography, a widely used non-invasive diagnostic tool to examine heart function (45).

Variables

The variables from six different locations included demographic factors (age and sex), symptom/clinical factors (chest pain type and exercise angina), physiological factors (resting blood pressure, serum cholesterol, fasting blood sugar, maximum heart rate), and diagnostic factors (resting electrocardiograph, oldpeak (ST depression induced by exercise relative to rest), and slope of ST segment). The dataset included readings for four different types of chest pain which were used for analysis. These types of pains were typical angina, atypical angina, non-anginal pain, and asymptomatic pain.

Data preprocessing

Data was imported into SPSS (version 29), which was the statistical analysis software used in the study. Initial exploration of the dataset identified several problems that required attention. There were 1190 cases in the original dataset. A check was done to determine if there were any duplicate cases across the predictor and outcome variables. It was found that the dataset had 272 duplicate cases, so these were eliminated, and the remaining 918 unique cases were used for further analysis. Additionally, it was found that there were 172 cases in the dataset that had a recorded cholesterol value of 0, which is not a valid value for cholesterol levels. Per SPSS guidelines in tagging missing values, these values were recoded as -1 to signify missing data and to avoid misrepresentation during further analysis. Three other cases were removed due to an invalid ST Slope, BP recorded as 0, and missing data for ST Slope and BP. By making these corrections, it was confirmed that the dataset of 918 was reliable and ready for analysis.

Statistical analysis

Multivariate characteristics in the form of categorical, continuous values and attributes of predictor variables, and the binary outcome variable were included in the dataset. The next step was to check for multicollinearity among the predictor variables. Multicollinearity is present when two variables are highly correlated, which makes it difficult to determine the individual effect of each variable. Multicollinearity is determined by examining the correlation matrix of the independent variables. The correlation matrix of predictor variables showed values for under acceptable levels of 0.7, indicating that multicollinearity as not an issue for this dataset (46). Chi-square analysis for non-parametric analysis and logistic regression procedure were used to test the hypotheses. We utilized predictive variables that have been historically associated with cardiovascular disease (47, 48).

ACKNOWLEDGMENTS

The student author would like to thank his Walton High School STEM Academy teachers, Ms. Tobie Hendricks and Dr. Tina Link, for their encouragement, motivation, and activities in the STEM program that helped him develop an interest in Biomedical science and research. He would also like to thank his mother, Sangita Hazari, for inspiring him to always do his best and showing him the importance of service.

Received: July 17, 2024
Accepted: November 15, 2024
Published: June 28, 2025

REFERENCES

1. Dattani, S., et al. "Cardiovascular Diseases." *Our World in Data*, 2023, https://ourworldindata.org/cardiovascular-diseases.
2. "World Heart Report. Confronting the World's Number One Killer." *World Heart Federation,* Geneva, Switzerland, 2023.
3. McClellan, M., et al. "Call to Action: Urgent Challenges in Cardiovascular Disease: A Presidential Advisory from the American Heart Association." *Circulation*, vol. 139, no. 9, 2019, pp. e44-e54, https://doi.org/10.1161/CIR.0000000000000652.
4. Van Praagh, R. "Terminology of Congenital Heart Disease. Glossary and Commentary." *Circulation*, vol. 56, no. 2, 1977, pp. 139-143, https://doi.org/10.1161/01.CIR.56.2.139.
5. Edwards, W. D. "Classification and Terminology of Cardiovascular Anomalies." *Moss & Adams Heart Disease in Infants, Children, and Adolescents Including the Fetus and Young Adult*, Lippincott Williams and Wilkins, Philadelphia, 2001, pp. 118-137.
6. Keteepe-Arachi, T. and Sharma, S. "Cardiovascular Disease in Women: Understanding Symptoms and Risk Factors." *European Cardiology Review*, vol. 12, no. 1,

2017, p. 10, https://doi.org/10.15420/ecr.2016:32:1.
7. Wood, D. "Established and Emerging Cardiovascular Risk Factors." *American Heart Journal*, vol. 141, no. 2, 2001, pp. S49-S57, https://doi.org/10.1067/mhj.2001.109951.
8. Teo, K. K. and Rafiq, T. "Cardiovascular Risk Factors and Prevention: A Perspective from Developing Countries." *Canadian Journal of Cardiology*, vol. 37, no. 5, 2021, pp. 733-743, https://doi.org/10.1016/j.cjca.2021.02.009.
9. Flora, G. D. and Nayak, M. K. "A Brief Review of Cardiovascular Diseases, Associated Risk Factors and Current Treatment Regimes." *Current Pharmaceutical Design*, vol. 25, no. 38, 2019, pp. 4063-4084, https://doi.org/10.2174/1381612825666190925163827.
10. Heidenreich, P. A., et al. "Forecasting the Future of Cardiovascular Disease in the United States: A Policy Statement from the American Heart Association." *Circulation*, vol. 123, no. 8, 2011, pp. 933-944, https://doi.org/10.1161/CIR.0b013e31820a55f5.
11. Pearson-Stuttard, J., et al. "Reducing US Cardiovascular Disease Burden and Disparities through National and Targeted Dietary Policies: A Modeling Study." *PLoS Medicine*, vol. 14, no. 6, 2017, p. e1002311, https://doi.org/10.1371/journal.pmed.1002311.
12. Li, Z., et al. "Global, Regional, and National Death, and Disability-Adjusted Life-Years (DALYs) for Cardiovascular Disease in 2017 and Trends and Risk Analysis from 1990 to 2017 Using the Global Burden of Disease Study and Implications for Prevention." *Frontiers in Public Health*, vol. 9, 2021, p. 559751, https://doi.org/10.3389/fpubh.2021.559751.
13. Kasabe, R. and Narang, G. "Heart Disease Prediction Using Machine Learning." *International Journal of Engineering Research & Technology*, vol. 9, no. 8, 2020, pp. 353-355, https://doi.org/10.17577/IJERTV9IS080128.
14. Benfallah, Y., et al. "Multivariate Analysis of the Impact of Demographic and Clinical Factors on Cardiovascular Health." *medRxiv*, 2024, p. 2024-05, https://doi.org/10.1101/2024.05.19.24307595.
15. Leifer, G. and Fleck, E. *Growth and Development Across the Lifespan: A Health Promotion Focus*. Elsevier Health Sciences, 2012.
16. Miller, V. "Why Are Sex and Gender Important to Basic Physiology and Translational and Individualized Medicine?" *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 306, no. 6, 2014, pp. 781-788, https://doi.org/10.1152/ajpheart.00994.2013.
17. Mehta, P. K., et al. "Gender in Cardiovascular Medicine: Chest Pain and Coronary Artery Disease." *European Heart Journal*, vol. 40, no. 47, 2019, pp. 3819-3826, https://doi.org/10.1093/eurheartj/ehz784.
18. LaCroix, A., et al. "Chest Pain and Coronary Heart Disease Mortality among Older Men and Women in Three Communities." *Circulation*, vol. 81, no. 2, 1990, pp. 437-446, https://doi.org/10.1161/01.CIR.81.2.437.
19. Cho, D., et al. "Gender Differences in the Presentation of Chest Pain in Obstructive Coronary Artery Disease: Results from the Korean Women's Chest Pain Registry." *The Korean Journal of Internal Medicine*, vol. 35, no. 3, 2020, pp. 582-592, https://doi.org/10.3904/kjim.2018.320.
20. Biswas, P., et al. "Role of Exercise Echocardiography to Predict Coronary Artery Disease." *University Heart Journal*, vol. 17, no. 2, 2021, pp. 114-117, https://doi.org/10.3329/uhj.v17i2.54368.
21. He, J., et al. "Trends in Cardiovascular Risk Factors in US Adults by Race and Ethnicity and Socioeconomic Status 1999-2018." *JAMA*, vol. 326, no. 13, 2021, pp. 1286-1298, https://doi.org/10.1001/jama.2021.15187.
22. Wang, V. Y., et al. "Image-Based Predictive Modeling of Heart Mechanics." *Annual Review of Biomedical Engineering*, vol. 17, no. 1, 2015, pp. 351-383, https://doi.org/10.1146/annurev-bioeng-071114-040609.
23. Sanchez-Pinto, L. N., et al. "Comparison of Variable Selection Methods for Clinical Predictive Modeling." *International Journal of Medical Informatics*, vol. 116, 2018, pp. 10-17, https://doi.org/10.1016/j.ijmedinf.2018.05.006.
24. Rodgers, J. L., et al. "Cardiovascular Risks Associated with Gender and Aging." *Journal of Cardiovascular Development and Disease*, vol. 6, no. 2, 2019, p. 19, https://doi.org/10.3390/jcdd6020019.
25. Baumeister, S. E., et al. "Predictive Modeling of Health Care Costs: Do Cardiovascular Risk Markers Improve Prediction?" *European Journal of Preventive Cardiology*, vol. 17, no. 3, 2010, pp. 355-362, https://doi.org/10.1097/HJR.0b013e328333a0b7.
26. Field, A. "Discovering statistics using IBM SPSS statistics." 2013, pp. 115-21.
27. O'Connell, A. A. and Amico, K. R. "Logistic Regression." *The Reviewer's Guide to Quantitative Methods in the Social Sciences*, 2010, pp. 221-239.
28. Smith, T. J. and McKenna, C. M. "An Examination of Ordinal Regression Goodness-of-Fit Indices under Varied Sample Conditions and Link Functions." *Multiple Linear Regression Viewpoints*, vol. 38, no. 1, 2012, pp. 1-7.
29. Kuha, J. "AIC and BIC: Comparisons of Assumptions and Performance." *Sociological Methods & Research*, vol. 33, no. 2, 2004, pp. 188-229, https://doi.org/10.1177/0049124103262065.
30. Pearson-Stuttard, J., et al. "Modeling Future Cardiovascular Disease Mortality in the United States: National Trends and Racial and Ethnic Disparities." *Circulation*, vol. 133, no. 10, 2016, pp. 967-978, https://doi.org/10.1161/CIRCULATIONAHA.115.019904.
31. He, J., et al. "Trends in Cardiovascular Risk Factors in US Adults by Race and Ethnicity and Socioeconomic Status, 1999-2018." *Journal of American Medical Association*, vol. 326, no. 13, 2021, pp. 1286-1298, https://doi.org/10.1001/jama.2021.15187.
32. Tsao, C. W., et al. "Heart Disease and Stroke Statistics—2023 Update: A Report from the American Heart Association." *Circulation*, vol. 147, no. 8, 2023, pp. e93-e621, https://doi.org/10.1161/CIR.0000000000001137.
33. Brand, A., et al. "Placebogenomics: A New Concept and Tool for Personalized Medicine and Public Health." *Journal of Integrative Biology*, vol. 25, no. 2, 2021, pp. 76-78, https://doi.org/10.1089/omi.2020.0221.
34. Steenman, M. and Lande, G. "Cardiac Aging and Heart Disease in Humans." *Biophysical Reviews*, vol. 9, no. 2, 2017, pp. 131-137, https://doi.org/10.1007/s12551-017-0255-9.
35. Lakatta, E. and Levy, D. "Arterial and Cardiac Aging: Major Shareholders in Cardiovascular Disease Enterprises." *Circulation*, vol. 107, no. 2, 2003, pp. 346-354, https://doi.org/10.1161/01.CIR.0000048893.62841.F7.
36. Pencina, M. J., et al. "Quantifying Importance of Major

Risk Factors for Coronary Heart Disease." *Circulation*, vol. 139, no. 13, 2019, pp. 1603-1611, https://doi.org/10.1161/CIRCULATIONAHA.117.031855.

37. Connelly, P. J., et al. "The Importance of Gender to Understand Sex Differences in Cardiovascular Disease." *Canadian Journal of Cardiology*, vol. 37, no. 5, 2021, pp. 699-710, https://doi.org/10.1016/j.cjca.2021.02.005.
38. Istolahti, Tiia, et al. "Long-term prognostic significance of the ST level and ST slope in the 12lead ECG in the general population." *Journal of Electrocardiology*, vol. 58, 2020, pp. 176-183, https://doi.org/10.1016/j.jelectrocard.2019.12.010.
39. Buckley, B. J., et al. "Exercise Based Cardiac Rehabilitation and All Cause Mortality among Patients with Atrial Fibrillation." *Journal of the American Heart Association*, vol. 10, no. 12, 2021, p. e020804, https://doi.org/10.1161/JAHA.121.020804.
40. Tegegne, T. K., et al. "Effects of Exercise-Based Cardiac Rehabilitation Delivery Modes on Exercise Capacity and Health-Related Quality of Life in Heart Failure: A Systematic Review and Network Meta-Analysis." *Open Heart*, vol. 9, no. 1, 2022, p. e001949, https://doi.org/10.1136/openhrt-2021-001949.
41. Chatfield, S. L. "Recommendations for Secondary Analysis of Qualitative Data." *The Qualitative Report*, vol. 25, no. 3, 2020, pp. 833-842, https://doi.org/10.46743/2160-3715/2020.4092.
42. Graham, B. "Kaggle Diabetic Retinopathy Detection Competition Report." *University of Warwick*, vol. 22, no. 9, 2015.
43. Pederson, L. L., et al. "Use of Secondary Data Analyses in Research: Pros and Cons." *Journal of Addiction Medicine and Therapeutic Science*, vol. 6, no. 1, 2020, pp. 058-060, https://doi.org/10.17352/2455-3484.000039.
44. Janosi, A., et al. "Heart Disease." *UCI Machine Learning Repository*, 1988, https://doi.org/10.24432/C52P4X.
45. Vieillard-Baron, A., et al. "A Decade of Progress in Critical Care Echocardiography: A Narrative Review." *Intensive Care Medicine*, vol. 45, 2019, pp. 770-788, https://doi.org/10.1007/s00134-019-05604-2.
46. Bayman, E. O. and Dexter, F. "Multicollinearity in Logistic Regression Models." *Anesthesia & Analgesia*, vol. 133, no. 2, 2021, pp. 362-365, https://doi.org/10.1213/ANE.0000000000005593.
47. Metz, L. D., et al. "The Prognostic Value of Normal Exercise Myocardial Perfusion Imaging and Exercise Echocardiography: A Meta-Analysis." *Journal of the American College of Cardiology*, vol. 49, no. 2, 2007, pp. 227-237, https://doi.org/10.1016/j.jacc.2006.08.048.
48. Oxborough, D., et al. "Exercise-Induced Cardiac Fatigue - A Review of the Echocardiographic Literature." *Echocardiography*, vol. 27, no. 9, 2010, pp. 1130-1140, https://doi.org/10.1111/j.1540-8175.2010.01251.x.