Cadiovascular Disease Prediction

Xavier White [1], Trey Newton [1], Blake Nichols [1]
Xiaomao Liu[2], Janmejay Mohanty[2], Biswajit Biswal[2] , Nikunja Swain[2]
South Carolina State University
Computer Science and Mathematics Department
1 – Students 2- Faculty Mentors

INTRODUCTION

Cardiovascular disease (CVD) remains one of the leading causes of death worldwide, posing a significant challenge to global health systems. Early detection and prevention are critical to reducing its impact, yet traditional diagnostic methods often fall short in identifying at-risk individuals before serious complications arise. This project focuses on developing a machine learning-based cardiovascular disease prediction model that can analyze key health indicators and predict individual risk with greater accuracy and efficiency. This aligns closely with the center's mission to advance health equity through innovative, data-driven solutions. By combining medical knowledge with artificial intelligence, this work embraces the center's commitment to improving public health outcomes, expanding access to preventive care, and empowering communities through technology-driven healthcare innovation.

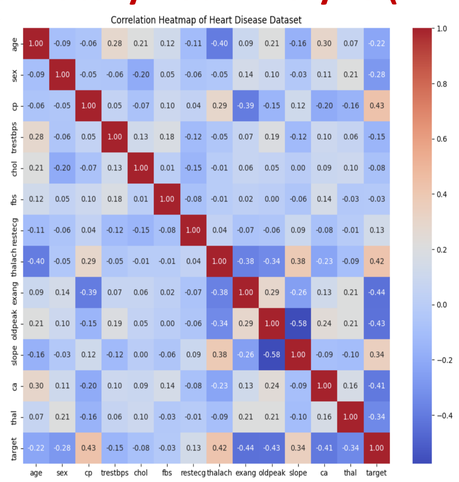
RESEARCH FOCUS & METHODS

This research represents a novel application of machine learning in preventive cardiology by integrating clinical data with advanced predictive modeling to improve early detection of cardiovascular risk. Unlike traditional diagnostic tools that often rely on limited indicators, our model processes a wide array of patient data—such as blood pressure, cholesterol, BMI, and lifestyle metrics—enabling a more comprehensive and personalized assessment. Preliminary results demonstrate high predictive accuracy across multiple algorithms, including Random Forest and Support Vector Machines, indicating strong potential for real-world clinical impact. This work builds on a foundation of interdisciplinary collaboration between data science and biomedical research (intra-thrust), while also fostering connections with behavioral science and health policy groups (inter-thrust) to address broader determinants of health. The insights gained from this study could transform preventive care practices, inform public health strategies, and contribute meaningfully to the growing field of AI-driven precision medicine.

ABSTRACT

This project presents a machine learning-based approach for predicting the risk of cardiovascular disease using patient health data. By leveraging algorithms such as Logistic Regression, Random Forest, and Support Vector Machines, the system analyzes features including age, blood pressure, cholesterol levels, BMI, and lifestyle factors to assess cardiovascular risk. The dataset, preprocessed for missing values and scaled for consistency, is split into training and testing sets to evaluate model accuracy. Performance metrics such as precision, recall, and ROC-AUC scores are used to determine the most effective model. This predictive tool aims to support early diagnosis and preventive healthcare decisions, offering an accessible, data-driven method for identifying at-risk individuals.

Exploratory Data Analysis (EDA)

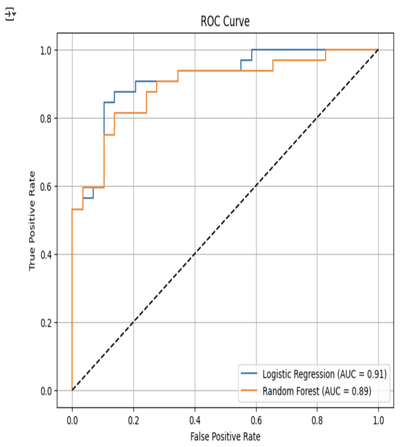


The heatmap shows the strength of correlation between different health features, helping identify which variables are most related to cardiovascular risk. The ROC curve illustrates the model's ability to distinguish between patients with and without heart disease, with a curve closer to the top-left indicating better performance.

RESULTS

```
Model loaded successfully.
Please provide the following information:
Age: 23
Sex (0 for female, 1 for male): 1
Chest Pain Type (0-3): 0
Resting Blood Pressure (in mm Hg): 120
Cholesterol (in mg/dl): 50
Fasting Blood Sugar > 120 mg/dl (0 for false, 1 for true): 0
Resting Electrocardiographic Results (0-2): 0
Maximum Heart Rate Achieved: 205
Exercise Induced Angina (0 for no, 1 for yes): 0
ST depression induced by exercise relative to rest: 0
Slope of the peak exercise ST segment (0-2): 0
Number of major vessels (0-3) colored by flourosopy: 0
Thal (0, 1, 2, or 3): 0

--- Prediction Results ---
Based on the provided information, the model predicts a higher likelihood of having heart disease.
Probability of having heart disease: 0.68
```



CONCLUSION

This project showed how machine learning can be used to predict the risk of cardiovascular disease using basic health data. By training and testing different models, we found effective ways to identify individuals at risk early on. The results highlight the potential of technology to support preventive healthcare and improve patient outcomes. With further development, this tool could be used in real-world settings to help doctors make faster, more informed decisions.

ACKNOWLEDGEMENT

The funding for this work was provided by (a) AIM for Composites, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science (award# DE-SC0023389), (b) RII Track-1: ADAPT in SC: AI-enabled Devices for the Advancement of Personalized and Transformative healthcare in South Carolina, National Science Foundation (NSF), Award # 2242812, (c) Preparing Cyber Warfare Professionals by Integration of Curriculum, Experiences, and Internships, Office of Naval Research (ONR), Award Number: N00014-23-1-2245, and BSRA SRNL grant.