

# Heart Disease Risk Assessment: A Machine Learning Approach

Minju Kim, Ling Li, Lei Chen

**Abstract**—Heart disease, particularly heart attacks (myocardial infarction), stands as a leading cause of mortality in the United States, accounting for a significant portion of deaths annually. Early detection of heart attacks is crucial due to the high fatality rates and the limited window for effective intervention. The difference between women and men has also not been fully illustrated. This study leverages the 'Cardiovascular Disease' dataset from Kaggle, encompassing a wide array of attributes like age, gender, blood pressure, and lifestyle factors, to explore the potential of machine learning in addressing this challenge. Utilizing models such as Logistic Regression, Naive Bayesian, Random Forest, and Decision Tree classifiers, the research evaluates their effectiveness in predicting cardiovascular diseases. Each model's performance is assessed based on accuracy, precision, recall, F1-Score, and ROC-AUC, with a focus on gender-specific variations in predictive accuracy. This study aims to advance the methodologies for early detection of heart disease, thereby contributing to the broader objective of reducing mortality rates associated with heart attacks through timely and data-driven intervention strategies.

## I. INTRODUCTION

In the United States, heart disease stands as a predominant cause of both male and female mortality, accounting for 35% of total female deaths in 2019, as per (1). Within the broad spectrum of heart disease, heart attacks, medically referred to as myocardial infarctions, emerge as a significant category. A heart attack transpires when a segment of the heart muscle experiences an insufficiency in blood flow, thereby making it a pivotal contributor to the overarching heart disease burden. Shockingly, the annual incidence of heart attacks surpasses 750,000 cases, as reported by (2). Regrettably, more than one-fifth of individuals grappling with heart attacks succumb to this condition, with an even more disheartening statistic indicating that almost half of those fatalities transpire suddenly, often occurring prior to their arrival at a hospital, as documented by (3) and (4).

Considering the alarming fatality rates and the narrow window for effective intervention in cases of heart attacks, early detection of such events emerges as a pivotal prerequisite for preventive therapeutic measures. Several well-documented risk factors, including but not limited to high blood pressure, diabetes, smoking, age, and gender, have been identified in prior studies, as elucidated in (5). However, the contemporary landscape presents a formidable challenge: the burgeoning volume of real-world data, replete with diverse features, makes traditional analytical approaches increasingly inadequate. The modern landscape poses a significant challenge, as the growing volume of diverse real-world data renders conventional analytical methods insufficient. Consequently, in this data-rich era, the rise of data mining techniques has empowered machine learning to play a central role in medicine. This intricate fusion of data-driven methodologies with the medical field holds the potential to revolutionize our approach to heart attack detection and prevention.

Compelling evidence underscores the significance of well-known risk factors like high blood pressure, high cholesterol, diabetes, obesity, unhealthy eating, a lack of physical activity, and smoking in the development of ischemic heart disease. Yet, there are equally important but often neglected risk factors, encompassing psychological, social, economic, and cultural factors, influenced by gender, that seem to contribute to cardiovascular disease, especially in women ((6); (7)). The Jackson Heart Study involving 5,301 African American participants has shown a stronger link between adult socioeconomic status and cardiovascular disease risk in women compared to men ((8)). To advance the development of innovative strategies for early detection and precise management of heart attacks in women, it is crucial to prioritize the creation of new prediction models that are highly sensitive to the female population.

The primary aim of this study is to anticipate the occurrence of heart attacks in women as opposed to men, employing a range of distinct attributes. We focus on addressing two key inquiries:

- 1) Are there variations in heart attack frequency based on gender?
- 2) Is there differentiation in the heart attack prediction utilizing medical assessments and lifestyle factors between men and women?

## II. RELATED WORK

Data scientists have harnessed the power of machine learning to create numerous models for diagnosing, detecting, and predicting a range of diseases, including heart attacks. As an illustration, (9) employed machine learning to integrate clinical data with quantitative imaging-based variables, resulting in a substantial enhancement in the accuracy of predicting myocardial infarctions and cardiac mortality. Additionally, (10) integrated age, sex, and paired high-sensitivity cardiac troponin I concentrations from patients, enabling the prediction of acute myocardial infarction likelihood.

Nonetheless, there remains a dearth of risk-assessment models that account for risk factors unique to females. Several conditions exclusive to women, such as obstetric and gynecological history encompassing gestational hypertension, gestational diabetes, and polycystic ovary syndrome ((11)), contribute to an elevated risk of cardiovascular disease. Therefore, to foster innovative advancements in early heart attack detection for women, greater emphasis must be placed on unraveling the influence of diverse risk factors on sex-specific disparities. Our project is dedicated to the prediction of heart attacks, particularly in women, employing a range of machine learning algorithms.

## III. METHODOLOGY

### A. Exploratory Data Analysis (EDA)

Our analysis began by examining the key target variable 'cardio', indicative of cardiovascular disease presence in the

dataset. Figure 1 illustrates the distribution of this variable using binary encoding: '0' for absence and '1' for presence of the disease. The comparable height of the bars in this figure indicates a balanced dataset in terms of the target variable, which is beneficial for model training. This balance ensures exposure to a nearly equal number of cases with and without cardiovascular disease, enhancing the potential for accurate generalization in predictions.

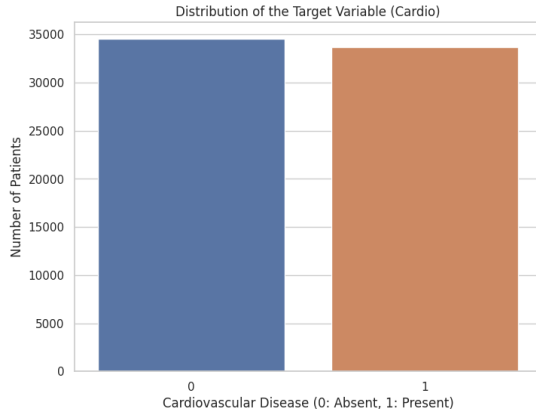


Fig. 1. Distribution of the target variable

Then we explore the distribution of categorical features in Figure 2 including gender, cholesterol, gluc, smoke, alco, active, and bp category encoded.

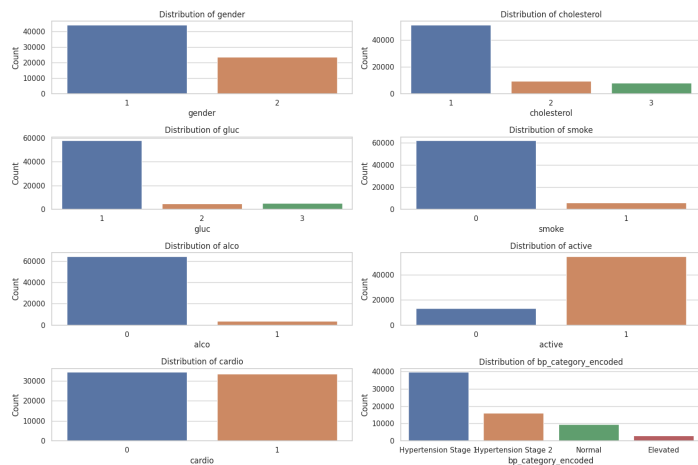


Fig. 2. Distribution of categorical features

- **Gender:** Within the dataset, we observe an imbalance in gender representation, with females appearing more frequently than males. The dataset does not provide additional context for the specific gender assignment for each type, leading to a noticeable gender distribution discrepancy in our analysis.
- **Cholesterol:** The data reveals that a substantial proportion of individuals exhibit Cholesterol level 1, corresponding to a normal or healthy range. In contrast, Cholesterol levels 2 and 3 are less prevalent, with level 3 being slightly more frequent than level 2. This underscores the predominance of individuals with normal cholesterol levels in the dataset.
- **Glucose (gluc):** The majority of individuals within the dataset exhibit a glucose level of 1, indicating a normal range. Levels 2 and 3 are less common, with level 3 being the least frequent. This suggests that most individuals have glucose levels within the normal range.
- **Smoke:** There is a significant disparity in smoking habits in the dataset. The vast majority do not smoke, while a

smaller proportion are identified as smokers. This highlights the predominance of non-smokers.

- **Alcohol (alco):** Similar to smoking, most individuals abstain from alcohol consumption, with only a smaller fraction categorized as alcohol consumers. This signifies a significant imbalance in alcohol consumption within the dataset.
- **Physical Activity (active):** The dataset categorizes individuals into two groups based on their level of physical activity. A substantial majority are classified as "active," indicating a physically active lifestyle, while a smaller proportion are "inactive." This highlights the prevalence of active lifestyles.
- **Blood Pressure Category (bp category encoded):** The distribution across blood pressure categories shows "Hypertension Stage 1" as the most common, followed by "Normal," and "Hypertension Stage 2." Fewer individuals fall into the remaining categories. This provides insights into the varying prevalence of different blood pressure stages, with "Hypertension Stage 1" being the most prevalent.

Now we differentiate the feature distribution based on the presence of cardiovascular disease, as shown in Figure 3.

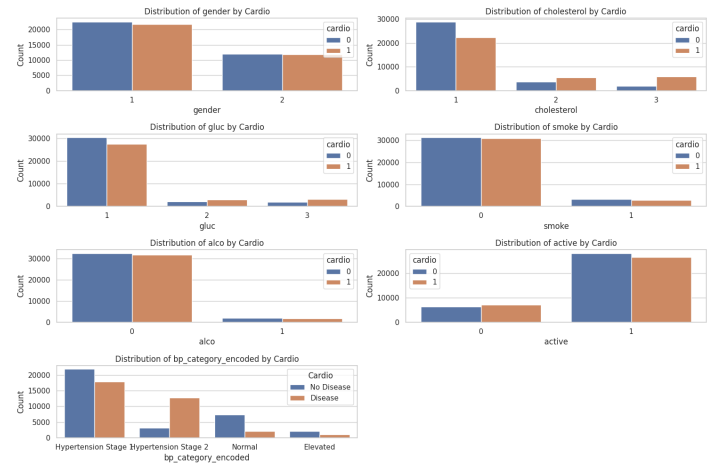


Fig. 3. Distribution of categorical features by the presence of cardiovascular disease

In our dataset, various factors play a role in the presence of cardiovascular diseases. Gender types show a relatively balanced distribution, with both types having individuals both with and without cardiovascular diseases, albeit with a slightly higher prevalence in gender type 1. The relationship between cholesterol levels and cardiovascular diseases is intriguing. Individuals with cholesterol level 1 exhibit a fairly even distribution of cardiovascular disease presence. However, for cholesterol levels 2 and 3, there is a noticeable increase in the prevalence of cardiovascular disease. A similar pattern is observed with glucose levels; level 1 displays a balanced distribution, while levels 2 and 3 show a slightly higher prevalence of the disease. The balance between smokers and non-smokers in terms of cardiovascular disease prevalence is also noteworthy. Whether individuals smoke or not, there is an equitable distribution. Likewise, cardiovascular disease distribution is comparable for alcohol consumers and non-consumers. Active individuals have a marginally lower prevalence of cardiovascular diseases compared to non-active ones. Furthermore, the data reveals intriguing insights when exploring blood pressure categories.

Individuals categorized under "Hypertension Stage 1" exhibit an even distribution of cardiovascular disease presence. In

contrast, those in the "Normal" category display a lower prevalence of the disease. However, individuals in the "Hypertension Stage 2" and other categories show a higher prevalence of cardiovascular diseases. These nuanced patterns within our dataset offer valuable insights into the interplay between various factors and the presence of cardiovascular diseases, forming a foundation for in-depth analysis and potential interventions. The main insights are as follows. Cholesterol levels seem to be a significant indicator of cardiovascular disease risk. As the cholesterol level increases, the risk of having the disease also appears to rise. Activity level also seems to play a role. Being active might be associated with a lower risk of cardiovascular diseases. While smoking and alcohol consumption are often considered risk factors for many health conditions, in this dataset, they don't show a pronounced difference in cardiovascular disease prevalence between their categories. The blood pressure category provides insights into the relationship between blood pressure stages and cardiovascular diseases. Those in the "Normal" range seem to have a lower risk, while individuals in higher stages of hypertension show a higher risk.

We then explored the gender-specific distribution of categorical variables. Figure 4 presents these trends, highlighting that males typically show higher counts of normal and elevated cholesterol, as well as glucose levels. Conversely, lifestyle factors like smoking and alcohol consumption are more common among females. Physical activity patterns indicate greater activity in males, and a similar trend is noted in blood pressure, where males predominantly fall within 'Hypertension Stage 1' and 'Hypertension Stage 2' categories.

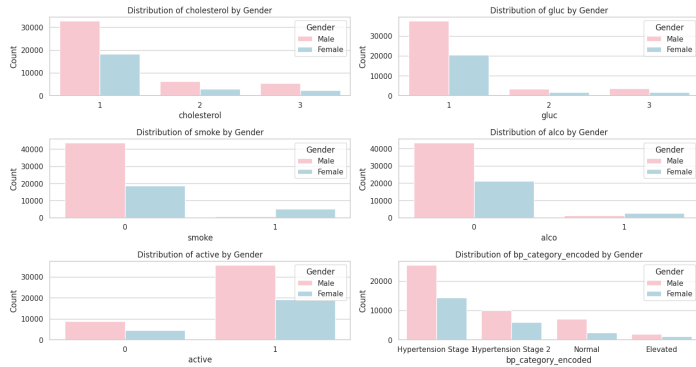


Fig. 4. Distribution of categorical features by gender

Further, we analyzed the gender-based distribution of numeric variables. This analysis, shown in Figure 5, reveals that males generally have higher height and weight, while blood pressure measurements show similar central tendencies but more variance in males. BMI distributions for both genders mostly peak within the healthy range.

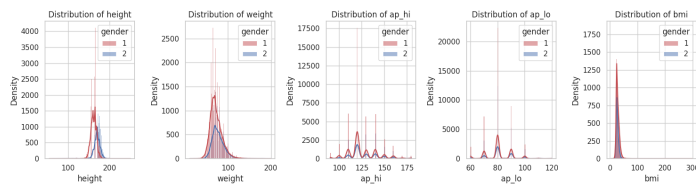


Fig. 5. Distribution of continuous features by gender

- 1) **High Positive Correlations:** Notably, 'age' and 'age years' exhibit a perfect positive correlation with a coefficient of 1. This strong correlation is unsurprising, as 'age years' is likely a derived value from 'age,' resulting from the conversion of days to years. Similarly, 'weight'

and 'bmi' (Body Mass Index) share a robust positive correlation, indicating that as weight increases, BMI tends to rise as well. This correlation is expected, given that BMI is inherently a function of both weight and height.

- 2) **Moderate Positive Correlations:** Within the dataset, 'ap\_hi' (systolic blood pressure) displays a moderate positive correlation with 'cardio,' suggesting that higher systolic blood pressure values may be associated with the presence of cardiovascular diseases. Additionally, 'cholesterol' and 'cardio' demonstrate a moderate positive correlation, implying that elevated cholesterol levels could be linked to cardiovascular diseases.
- 3) **Low to Moderate Negative Correlations:** A negative correlation is observed between 'active' (physical activity) and 'cardio,' though it is not particularly strong. This suggests that being physically active may be associated with a lower risk of cardiovascular diseases, aligning with earlier visual observations.

These correlation findings provide valuable guidance for further analysis. 'ap\_hi' and 'cholesterol' are potential significant predictors for the target variable 'cardio,' given their noticeable correlations. The perfect correlation between 'age' and 'age years' implies redundancy, where one feature would suffice for modeling. While 'smoke' and 'alco' are often regarded as risk factors for cardiovascular diseases, their linear correlations with 'cardio' are minimal in this dataset. However, their potential impacts might be better captured through non-linear relationships or interactions with other features. It is essential to remember that correlation does not imply causation, and further analyses, like regression or causal inference methods, are necessary to understand the nature and causality of these relationships thoroughly.

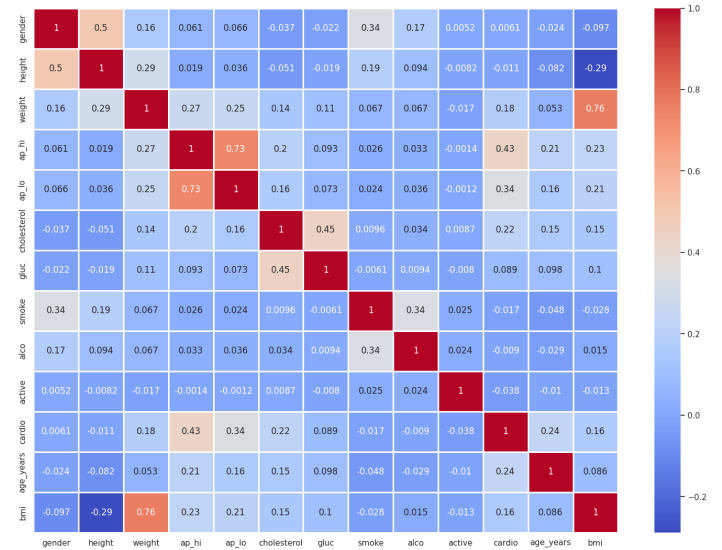


Fig. 6. Correlation between variables

## B. Preprocessing

Our initial assessment of outliers was conducted through boxplot visualizations for several key metrics, including "height," "weight," "ap\_hi" (systolic blood pressure), "ap\_lo" (diastolic blood pressure), and "bmi" (Body Mass Index). Our analysis revealed that the outliers present in the "ap\_hi" and "ap\_lo" measurements were not substantial enough to warrant adjustment. These variables, indicative of blood pressure, often exhibit natural variability and are less prone to measurement error or data entry anomalies that typically produce outliers.

Conversely, for the variables "height," "weight," and "bmi," we observed a more pronounced presence of significant outliers. Given the high correlation among these three metrics—where an aberration in one is likely to affect the others—we deemed it necessary to apply our outlier handling methodology to these specific features.

Also, we choose to that Standard Scaling (Z-score normalization) for the following reasons. First, it centers the data by subtracting the mean from each feature, making it a preferred choice for machine learning algorithms, especially linear models like logistic regression and support vector machines. This centered data aids in achieving better model performance. Second, Standard Scaling imparts unit variance to the features by dividing each feature by its standard deviation, thereby standardizing the variance to 1. This ensures that all features contribute equally to distance-based computations, which is particularly valuable in applications like k-means clustering and k-nearest neighbors. Additionally, it offers robustness against outliers compared to other scaling methods, striking a balance between accommodating extreme values and preserving the overall dataset's scaling integrity. Furthermore, its versatility and general-purpose nature make it a reliable choice when specific data characteristics are not explicitly known, and it allows for meaningful feature coefficient comparisons in linear models, enhancing model interpretability. Table 3 presents the description of data characteristics after preprocessing, normalization and scaling.

### C. Model Evaluation Metrics

In evaluating our machine learning models, we employed six key metrics for a comprehensive performance assessment:

- 1) **Accuracy:** Measures the proportion of correctly predicted observations to the total observations.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Observations}}$$

- 2) **Precision:** Indicates the ratio of correctly predicted positive observations to the total predicted positive observations.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

- 3) **Recall (Sensitivity):** Assesses the ratio of correctly predicted positive observations to all observations in the actual class.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

- 4) **F1 Score:** The weighted average of Precision and Recall, considering both false positives and false negatives.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- 5) **ROC-AUC Score:** Represents the area under the Receiver Operating Characteristic (ROC) curve, measuring the model's ability to distinguish between classes.

- The ROC curve plots the true positive rate against the false positive rate at various threshold settings.

These metrics were applied to evaluate the performance of our models on the entire dataset, as well as on subsets for women's and men's datasets, providing insights into the model's effectiveness across different segments.

### D. Learning Curves and Permutation Feature Importances Analysis

In our model evaluation process, we focused on learning curves and permutation feature importances to gauge model performance and the impact of various features.

1) **Learning Curves Analysis::** Learning curves demonstrate the evolution of a model's performance with the increase in training data, plotting training and validation scores against the volume of training data.

- Formulation:

$$\text{Train Scores Mean} = \frac{1}{n} \sum_{i=1}^n \text{Train Scores}_i$$

$$\text{Validation Scores Mean} = \frac{1}{n} \sum_{i=1}^n \text{Validation Scores}_i$$

- These calculations help determine if the model is underfitting or overfitting as more data is introduced.

2) **Permutation Feature Importances Analysis::** Permutation feature importance is used to assess the contribution of each feature to the model's predictive accuracy.

- Formulation:

$$\text{Feature Importance} = \text{Model Score}_{\text{original}} - \text{Model Score}_{\text{shuffled}}$$

- This process is repeated to obtain an average importance score for each feature, indicating its influence on model predictions.

These analytical techniques played a pivotal role in evaluating the performance of our models and facilitating a comparative analysis across different datasets, including the entire dataset, as well as segmented data for women and men.

## IV. BASELINE MODEL

We use the Logistic model as our baseline model. As shown in Table I, As delineated in Table I, the model demonstrates an accuracy range of 72-74% across both the entire dataset and gender-specific subsets. However, precision consistently surpasses recall, suggesting cautious predictions but missing some positive cases. Further exploration indicates that the model performs slightly better for women in terms of accuracy and recall when compared to its performance for men. This observation might suggest that certain features within the dataset hold stronger predictive power for women or that the distribution of specific risk factors differs by gender To enhance performance, adjusting thresholds or employing techniques like SMOTE could improve the model's ability to identify high-risk individuals without compromising precision. Overall, while the model serves as a reliable baseline for predicting cardiovascular diseases, there's potential to optimize its performance by improving recall without sacrificing precision, particularly in identifying high-risk cases.

TABLE I  
LOGISTIC MODEL ASSESSMENT

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Overall	0.7250	0.7503	0.6570	0.7006	0.7236
Female	0.7366	0.7573	0.6845	0.7191	0.7358
Male	0.7239	0.7658	0.6563	0.7069	0.7249

With an extensive array of features at our disposal, we conducted an exploratory analysis employing Principal Component Analysis (PCA) to effectively reduce the dimensionality

of our dataset. Visual representation in the Figure 7 elucidates the variance elucidated by each principal component. Nevertheless, our efforts employing the initial 7 components and subsequent training of the base model yielded subpar results compared to the performance of the original, unabridged model. One compelling rationale for abstaining from PCA is its limitation in providing insights into feature importance, crucial for comprehending the underlying data dynamics. Hence, we have opted against pursuing PCA as it does not afford us the necessary clarity on the most influential features within our dataset, and the technique did not enhance the model's predictive capabilities as anticipated.

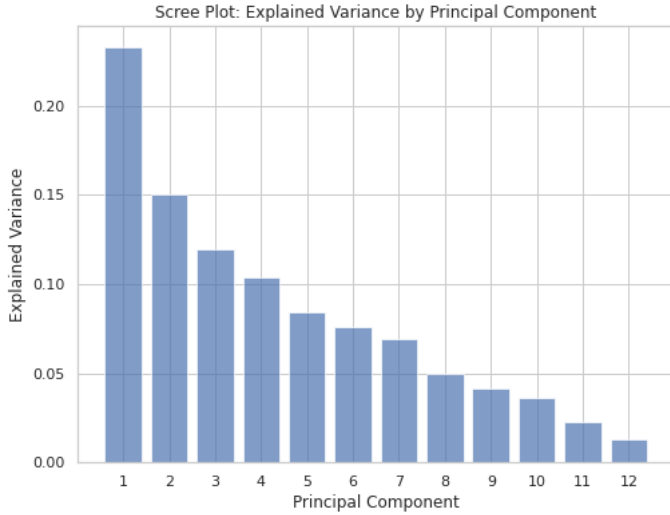


Fig. 7. Explained variance by principal component

## V. EXPERIMENTS AND RESULTS

We experimented with various classification models including Naive Bayesian, Random Forest, Decision Tree, SVM, KNN, and XGBoost. Each model underwent a triad of training iterations: one on the complete dataset, another solely on women's data, and a third on men's data. Following this, meticulous fine-tuning of the models ensued, involving parameter adjustments aimed at elevating their performance. We rigorously evaluated these models, examining not only accuracy scores, recall rates, and learning curves but also delving deeper into additional performance indicators. This comprehensive analysis aimed to comprehend their overall efficacy. Seeking deeper understanding, we delved into the significance of diverse features within these models. Moreover, we scrutinized how these models demonstrate distinct behaviors in relation to different genders.

**Naive Bayesian:** It is widely used in classification tasks. Based on Bayes' theorem, it calculates the probability of a particular event or class occurring based on prior knowledge of conditions associated with that event.

TABLE II  
NAIVE BAYESIAN ASSESSMENT

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Overall	0.7101	0.7297	0.6332	0.6780	
Female	0.7202	0.7334	0.6531	0.6910	0.7175
Male	0.7131	0.7511	0.6320	0.6865	0.7126

As shown in Table II. The Naive Bayes model showed an overall accuracy of 71.01% on the entire dataset. When

examining gender-specific data, it performed better for women with an accuracy of 72.02% and slightly less effectively for men at 71.31%. For women, the model had higher precision and recall compared to men, but men had a slightly higher precision. The analysis revealed differences in predictive outcomes between genders, notably in true negatives, false positives, false negatives, and true positives within the confusion matrices.

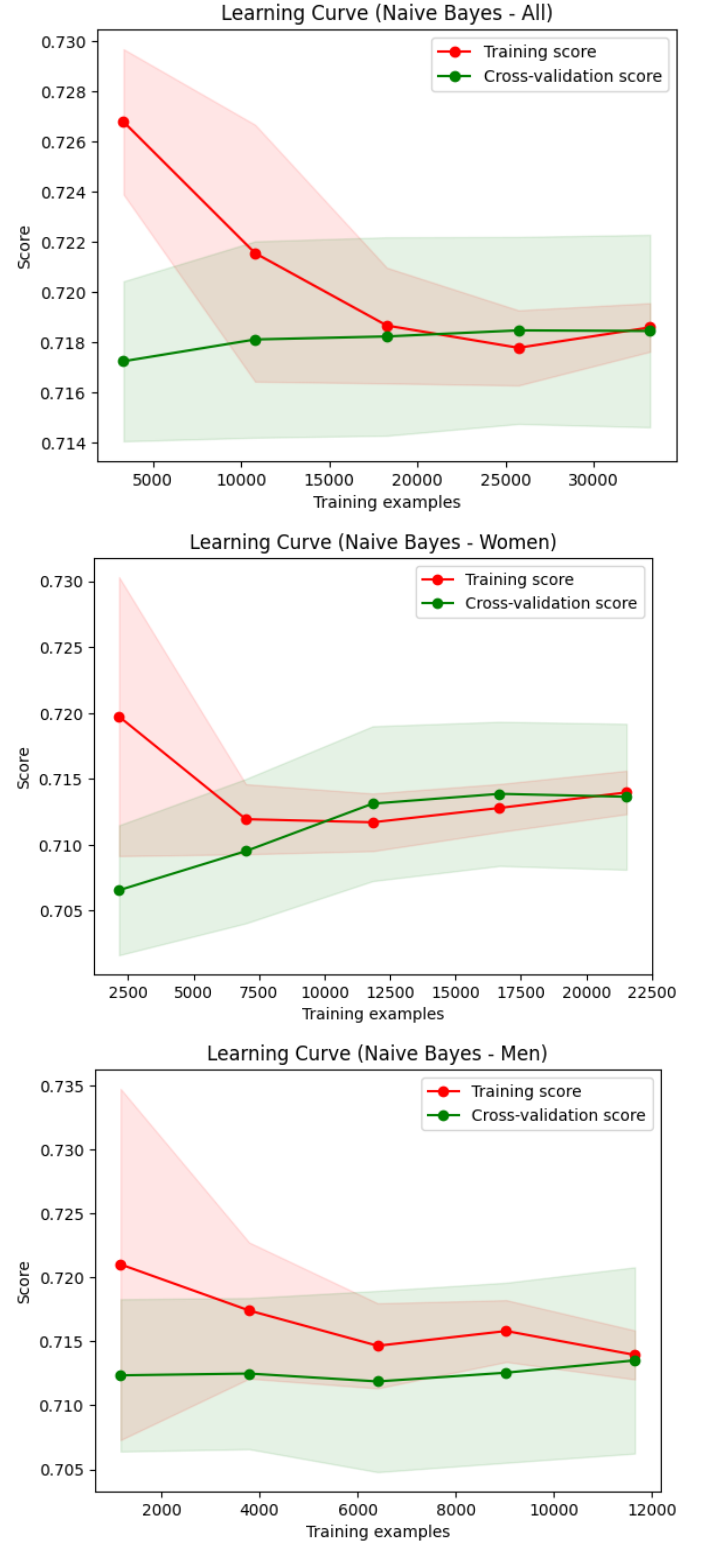


Fig. 8. Learning curves of Naive Bayesian models

The learning curve showcases an intriguing progression where the training score initially declines before rising, while the cross-validation score consistently ascends, culminating in a convergence to a single point. This pattern suggests an initial struggle of the model with limited data or high complexity,



resulting in the training score decreasing as the model overfits or fails to generalize well. However, as more data is incorporated or adjustments are made to simplify the model, the training score gradually increases, signifying improved generalization. Simultaneously, the cross-validation score steadily rises, indicating the model's consistent ability to generalize to unseen data. Ultimately, both scores converge to a singular point, denoting an optimal balance between model complexity and generalization, portraying a well-performing model that neither underfits nor overfits the data.

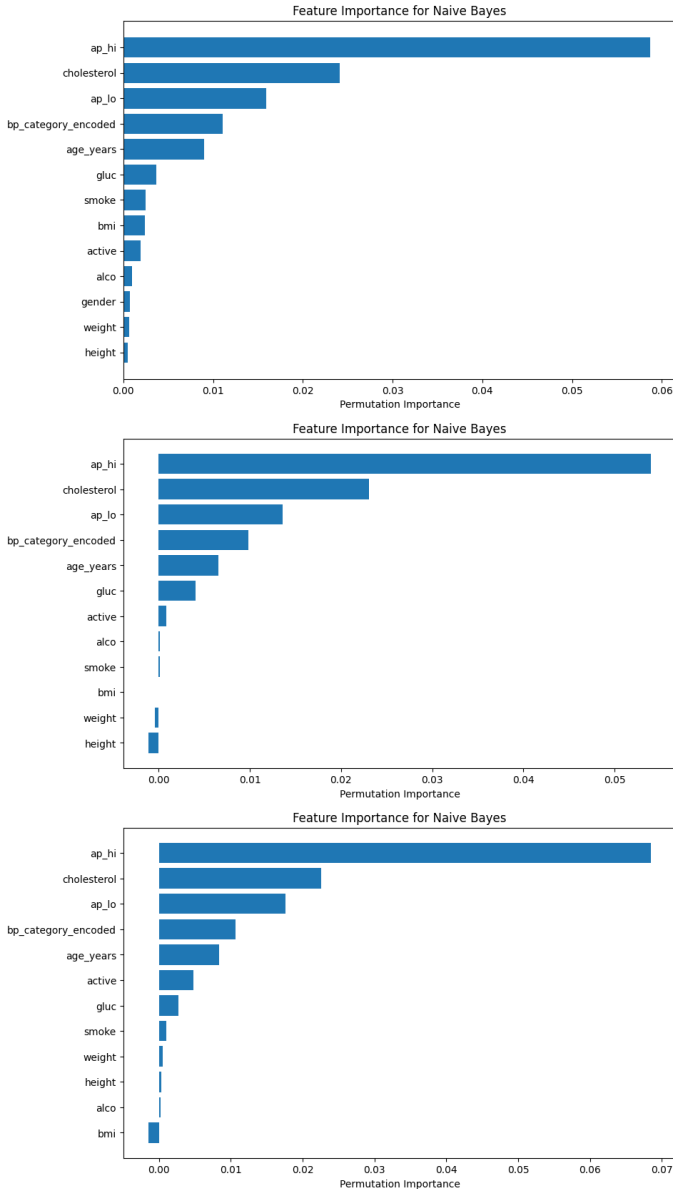


Fig. 9. Feature importance in the Naive Bayesian models

The feature importance analysis within the Naive Bayes predictive models for heart attack prediction reveals distinct gender-specific influential factors. Among women, systolic blood pressure, cholesterol level, diastolic blood pressure, overall blood pressure, and age emerge as the top five significant determinants. Conversely, in the case of men, systolic blood pressure, age, weight, cholesterol level, and height stand out as the most critical predictors. These findings underscore the nuanced differences in risk factor importance between genders in predicting heart attacks, emphasizing the varying impacts of physiological parameters on heart health among women and men. Such gender-specific insights not only enhance our understanding of cardiovascular risk assessment

but also pave the way for tailored preventive measures and targeted interventions to mitigate heart attack risks based on individualized risk profiles.

**Other classification models:** The assessment matrix, learning curve and insights of other machine learning models including Random Forest, Decision Tree, SVM, KNN, and XGBoost are described in the Appendix.

**Ensemble model:** Upon thorough evaluation of our six machine learning models, we've selected the top performers: Random Forest, SVM, XGBoost, and Logistic Regression, to create an ensemble using voting and weighted combination. Collectively, across various datasets and gender divisions, these models achieve an overall accuracy of approximately 73%, indicating a modest enhancement compared to their individual performances. Notably, the recall scores consistently reach approximately 66% to 67%, signifying the models' capacity to capture a significant portion of true positive instances for both genders.

TABLE III  
RANDOM FOREST ASSESSMENT AFTER PARAMETER TUNING

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Overall	0.7320	0.7556	0.6602	0.7047	0.7298
Female	0.7342	0.7467	0.6697	0.7061	0.7314
Male	0.7332	0.7510	0.6701	0.7083	0.7312

**Deep learning models** In our final phase, we trained deep learning models and found that the Multi-Layer Perceptron (MLP) generally outperformed the Perceptron across various evaluation metrics, encompassing both the entire dataset and gender-segregated subsets. Despite the MLP's enhanced performance, it did not surpass the ensemble learning models. This could be attributed to reasons such as the dataset's non-linearity, high learning rates, or insufficient learning time for these models. Upon comparing the Perceptron and MLP for binary classification, nuanced performance differences emerged across the dataset and gender subsets. The Perceptron achieved 69.50% overall accuracy, with varied precision, recall, and F1-Score in women's (58.08%) and men's (68.30%) data. Conversely, the MLP displayed a slightly higher accuracy of 69.51%, notably improving outcomes for women (70.83%) versus men (67.83%). Consistently outperforming the Perceptron, the MLP exhibited superior balance between precision and recall in women's data and maintained similar performance in men's data, indicating its overall superiority in handling diverse classification scenarios.

In analyzing feature importance, we've discerned specific factors that are distinctive to each gender. These results emphasize the varying significance of certain physiological factors in predicting outcomes for women and men, while also illuminating shared factors that apply across genders. Overall, the ensemble model demonstrates a balanced and consistent performance across both genders, indicating its potential usefulness for prediction tasks on this dataset.

## VI. CONCLUSION

This research demonstrates the effectiveness of machine learning in predicting heart attacks. The Logistic Regression model, serving as the baseline, showed consistent performance across gender subsets with an accuracy of around 72-74%, suggesting strong generalization capabilities. Then, the Naive Bayesian model displayed notable precision and was particularly effective in predicting outcomes for women compared with

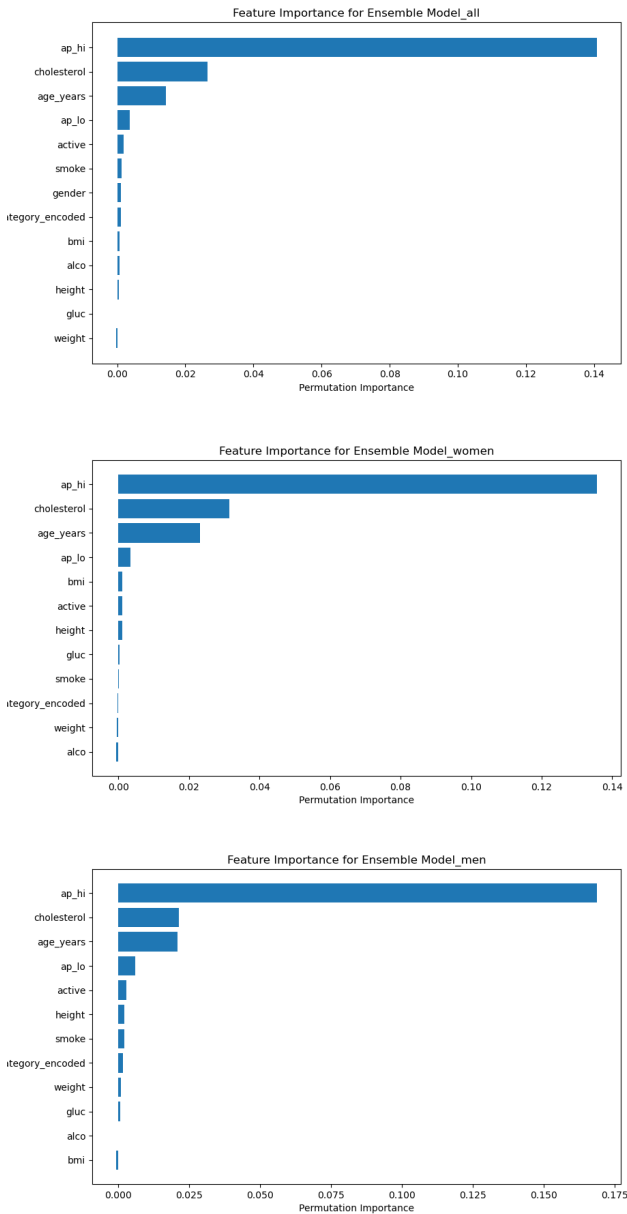


Fig. 10. Feature Importance of Ensemble Model

that for men. The Random Forest model, after parameter tuning, emerged as robust and accurate, with improved performance metrics across all subsets and a balanced precision-recall trade-off. The Decision Tree classifier performed with moderate accuracy, showing slight variations in predictive patterns between genders. In addition, the Ensemble model which combines Random Forest, SVM, XGBoost, and Logistic models exactly improved the prediction of heart attacks. What's more, the feature importance analysis in our research revealed the main significant features contributing to the heart attack prediction model like blood pressure, cholesterol level, and age. Overall, these models underscore the potential of machine learning in revolutionizing heart attack detection and prevention, especially when considering gender-specific risk factors. The study highlights the importance of continued exploration in this field, advocating for more nuanced, data-driven approaches to healthcare.

## VII. LIMITATIONS AND FUTURE DEVELOPMENT

Our research still has some limitations. First and foremost, although this dataset has involved various features that have

been recognized as directly relevant to cardiovascular disease, the model here might not capture certain gender-specific features that could be crucial for accurate predictions such as obstetric and gynecological history encompassing gestational hypertension. According to the feature importance analysis, high blood pressure contributes significantly to the prediction of cardiovascular disease. In our future model, we may need to consider women's gestational hypertension history. Therefore, a lack of gender-specific information or insufficient features might limit the model's ability to distinguish between genders effectively. It will be better to collect more diverse data that includes a broader representation of gender-related factors to improve the model's understanding.

Secondly, gender imbalance is another potential limitation in our datasets, which means inherent biases related to gender. The model we used might inadvertently learn and perpetuate these biases, impacting the fairness in predictions of cardiovascular disease.

Additionally, the feature importance analysis in this research revealed that certain features have a negative effect on the model's performance or introduce biases. We didn't perform feature extraction or modification, which might be a valid approach to improve the model.

Last but not least, for the ensemble model, although it is powerful, it could be complex and less interpretable compared to individual base models. Understanding the decision-making process might be challenging. Results may be improved by using simpler models or techniques that offer better interpretability while maintaining reasonable performance. To understand and interpret ensemble model predictions, model-agnostic interpretability techniques (e.g., SHAP values, LIME) will also be useful.

## REFERENCES

- [1] J. Hagenauer, E. Offer, and L. Papke. Iterative decoding of binary block and convolutional codes. *IEEE Trans. Inform. Theory*, vol. 42, no. 2, pp. 429–445, Mar. 1996.
- [2] T. Mayer, H. Jenkac, and J. Hagenauer. Turbo base-station cooperation for intercell interference cancellation. *IEEE Int. Conf. Commun. (ICC)*, Istanbul, Turkey, pp. 356–361, June 2006.
- [3] J. G. Proakis. *Digital Communications*. McGraw-Hill Book Co., New York, USA, 3rd edition, 1995.
- [4] F. R. Kschischang. Giving a talk: Guidelines for the Preparation and Presentation of Technical Seminars. <http://www.comm.toronto.edu/frank/guide/guide.pdf>.

## REFERENCES

2030. *The Lancet*, 397(10292):2385–2438, 2021.

- [1] Kevin S Ikuta, Lucien R Swetschinski, Gisela Robles Aguilar, Fablina Sharara, Tomislav Mestrovic, Authia P Gray, Nicole Davis Weaver, Eve E Wool, Chieh Han, Anna Gershberg Hayoon, et al. Global mortality associated with 33 bacterial pathogens in 2019: a systematic analysis for the global burden of disease study 2019. *The Lancet*, 400(10369):2221–2248, 2022.
- [2] Emelia J Benjamin, Michael J Blaha, Stephanie E Chiuve, Mary Cushman, Sandeep R Das, Rajat Deo, Sarah D De Ferranti, James Floyd, Myriam Fornage, Cathleen Gillespie, et al. Heart disease and stroke statistics—2017 update: a report from the american heart association. *circulation*, 135(10):e146–e603, 2017.
- [3] Alan S Go, Dariush Mozaffarian, Véronique L Roger, Emelia J Benjamin, Jarett D Berry, William B Borden, Dawn M Bravata, Shifan Dai, Earl S Ford, Caroline S Fox, et al. Heart disease and stroke statistics—2013 update: a report from the american heart association. *Circulation*, 127(1):e6–e245, 2013.
- [4] Joseph P Ornato and Mary M Hand. Warning signs of a heart attack. *Circulation*, 104(11):1212–1213, 2001.
- [5] Richard C Becker. Heart attack and stroke prevention in women. *Circulation*, 112(17):e273–e275, 2005.
- [6] Gregory A Roth, Catherine Johnson, Amanuel Abajobir, Foad Abd-Allah, Semaw Ferede Abera, Gebre Abyu, Muktar Ahmed, Baran Aksut, Tahiya Alam, Khurshid Alam, et al. Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. *Journal of the American college of cardiology*, 70(1):1–25, 2017.
- [7] Kristi Rahrig Jenkins and Mary Beth Ofstedal. The association between socioeconomic status and cardiovascular risk factors among middle-aged and older men and women. *Women & health*, 54(1):15–34, 2014.
- [8] Samson Y Gebreab, Ana V Diez Roux, Allison B Brenner, DeMarc A Hickson, Mario Sims, Malavika Subramanyam, Michael E Griswold, Sharon B Wyatt, and Sherman A James. The impact of lifecourse socioeconomic position on cardiovascular disease events in african americans: the jackson heart study. *Journal of the American Heart Association*, 4(6):e001553, 2015.
- [9] Frederic Commandeur, Piotr J Slomka, Markus Goeller, Xi Chen, Sebastien Cadet, Aryabod Razipour, Priscilla McElhinney, Heidi Gransar, Stephanie Cantu, Robert JH Miller, et al. Machine learning to predict the long-term risk of myocardial infarction and cardiac death based on clinical risk, coronary calcium, and epicardial adipose tissue: a prospective study. *Cardiovascular research*, 116(14):2216–2225, 2020.
- [10] Martin P Than, John W Pickering, Yader Sandoval, Anoop SV Shah, Athanasios Tsanas, Fred S Apple, Stefan Blankenberg, Louise Cullen, Christian Mueller, Johannes T Neumann, et al. Machine learning to predict the likelihood of acute myocardial infarction. *Circulation*, 140(11):899–909, 2019.
- [11] Birgit Vogel, Monica Acevedo, Yolande Appelman, C Noel Bairey Merz, Alaide Chieffo, Gemma A Figtree, Mayra Guerrero, Vijay Kunadian, Carolyn SP Lam, Angela HEM Maas, et al. The lancet women and cardiovascular disease commission: reducing the global burden by