International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 OPEN ACCESS
Available Online at : www.ijsrcseit.com
Volume 10, Issue 7, May-June-2024 | Published Online : 20th June 2024

Exploring Heart Disease : Perspectives and Visualizations

R Suganya[1], Baswaraju Swathi[2], Sai Ashwin Palukuru[3], Tene Eshani[4]

[1]Associate Professor, Department of Computer Science and Engineering(Data Science), New Horizon College of Engineering

[2]Associate Professor & Head, Department of Computer Science and Engineering(Data Science) New Horizon College of Engineering

[2]UG Studen, Department of Computer Science and Engineering(Data Science), New Horizon College of Engineering

ABSTRACT

Heart disease is a major global health concern, necessitating extensive research into its complex clinical and demographic components. This study intends to reveal critical insights for developing preventative and treatment techniques through extensive data analysis and visualization. This research initiative intends to thoroughly investigating cardiac disease utilizing modern data visualization approaches. The objective of this research is to find intricate patterns and connections that will shed light on the disease's dynamics by analyzing clinical and demographic datasets in depth. Using visualization methods such as scatter plots, heatmaps, and interactive charts, the project aims to create a full visual story of heart disease, providing essential views for both healthcare practitioners and researchers. The findings are intended to help enhance knowledge and management of cardiac disease, ultimately allowing for more informed healthcare decision-making.

Keywords- Heart disease, Perspectives, Visualization, Key risk factors, Decision-making

I. INTRODUCTION

Heart disease, which includes illnesses including coronary artery disease, myocardial infarction, and heart failure, remains a major global health concern. It accounts for a sizable proportion of global mortality and morbidity, putting enormous strain on healthcare systems and economies alike (Global Burden of Disease Collaborative Network, 2020). According to the World Health Organization (WHO), cardiovascular diseases (CVDs) account for approximately 17.9 million deaths per year, emphasizing the vital need for better prevention and treatment techniques. Understanding the various components that contribute to heart disease is critical for designing focused therapies and improving patient outcomes. Epidemiological studies have found a number of risk factors, including hypertension, diabetes, obesity, smoking, and sedentary lifestyles, that increase the likelihood of getting cardiovascular disease [1,2]. These findings highlight the necessity of comprehensive approaches that address both individual risk factors and larger systemic concerns impacting heart health.

In addition to individual risk factors, socioeconomic determinants such as healthcare access, education, and wealth disparity have a significant impact on cardiovascular health inequalities [3,4]. Addressing these inequities necessitates a thorough understanding of how social and environmental variables interact with biological determinants to affect heart disease prevalence and outcomes. Furthermore, advances in medical technology and therapeutic options have transformed the management of heart disease, increasing survival rates and quality of life for many patients. However, inequities in access to these breakthroughs persist, emphasizing the ongoing importance of equitable healthcare policies and actions [5-7]. Despite advances in knowledge and treatment, heart disease continues to be a severe socioeconomic burden worldwide. The economic costs of cardiovascular disease include direct medical bills, productivity losses owing to disability and premature death, and strain on healthcare systems. To address these difficulties, healthcare practitioners, academics, politicians, and community stakeholders must work together interdisciplinary, in addition to improving public health policies.

The goals of the research on heart disease

- ✓ Investigate worldwide and regional changes in the prevalence, mortality, and risk factors for heart disease.
- ✓ Identify key risk factors. Consider lifestyle factors (e.g., food, exercise), medical history (e.g., hypertension, diabetes), and socioeconomic factors that contribute to heart disease.
- ✓ Investigate Socioeconomic differences: Investigate differences in heart disease incidence and outcomes based on access to healthcare, education, and economic determinants.
- ✓ Data Visualization Techniques: To illustrate complicated patterns in heart disease data, use advanced visualization tools such as scatter plots, heatmaps, and interactive charts.
- ✓ Inform Prevention and Treatment methods: Offer evidence-based insights to improve preventive measures and treatment methods for reducing the impact of heart disease on public health.

These aims are intended to guide research toward a more complete understanding of cardiac illness, enabling for better healthcare policy and practice decisions.

II. LITERATURE SURVEY

In recent years, the healthcare industry has witnessed notable advancements through the integration of data mining and machine learning techniques, particularly within the realm of medical cardiology. Heart disease remains a prominent global cause of mortality [8-12], underscoring the critical need to identify risk factors and early indicators through robust research efforts. These advanced methodologies offer promising avenues for early detection and preventive strategies. [13] sought to enhance the precision of the Framingham risk score (FRS) by developing a quantum neural network-based system for predicting cardiovascular disease (CVD). Validated with data from 689 individuals and compared against the FRS using Framingham research, their model achieved an impressive 98.57% accuracy in predicting CVD risk, surpassing existing methods. This approach holds potential for assisting healthcare providers in devising more effective treatment plans and enabling early disease diagnosis. [14] utilized the Cleveland heart disease dataset to construct a machine learning model for cardiovascular disease prediction. Employing supervised classification methods such as naive Bayes, decision tree, random forest, and k-nearest neighbor (KKN), they found KKN to achieve the highest accuracy at 90.8%. This study underscores the effectiveness of machine learning in disease prediction and underscores the importance of selecting appropriate models.

[17] applied machine learning techniques to identify significant risk factors for cardiovascular disease among patients with metabolic-associated fatty liver disease (MAFLD). Their model, incorporating multiple logistic regression and principal component analysis (PCA), effectively identified high-risk patients with an

accuracy of 85.11% and an AUC of 0.87. This study highlights the potential of ML in detecting widespread CVD risk based on straightforward patient criteria. [15] explored various ML algorithms, including decision tree and support vector machine (SVM), for predicting heart failure using data from the Cleveland Clinic Foundation. Their findings revealed that the decision tree achieved the highest accuracy of 93.19%, offering valuable insights into ML's application in predicting heart disease. [16] conducted a comparative study to identify optimal feature selection methods for anticipating cardiovascular disease. Their research concluded that combining the XGBoost classifier with the wrapper technique yielded the most accurate results, achieving an accuracy of 73.74%. This study underscores the importance of robust feature selection in enhancing predictive analytics for cardiovascular illness. Despite these advancements, previous research often suffers from limited datasets that may lead to overfitting. In contrast, our study leverages a substantial dataset comprising 70,000 patients and 11 features, thereby reducing the risk of overfitting.

III. KNOWLEDGE THROUGH STATISTICAL VISUALIZATION

Statistical visualization in relation to heart disease involves using graphs, charts, and visual representations to analyze and present data on cardiovascular health. These tools help researchers and healthcare professionals gain deeper insights into aspects such as risk factors, disease progression, treatment outcomes, and epidemiological trends. These visuals, like scatter plots for correlations and heatmaps for prevalence, make complex data easier to understand and communicate, aiding in decision-making for healthcare strategies and policies.

Explore the dataset.Identify patterns, distributions, and relationships in the data. Perform an intensive exploratory data analysis (EDA). Dive into bivariate relationships with the target to examine how variables such as age, gender, chest pain type, blood pressure, cholesterol levels, and other clinical indications influence the presence or absence of heart disease. The dataset for this study was gathered from Kaggle, which offers a plethora of information for cardiovascular health research.

The dataset consists of various demographic, clinical, and diagnostic attributes related to heart disease. Key variables include age, representing the patient's age in years, and sex, denoting gender (0 for male, 1 for female). The variable "cp" indicates the type of chest pain experienced by patients, categorized into four types: typical angina, atypical angina, non-anginal pain, and asymptomatic. Resting blood pressure (trestbps) and serum cholesterol (chol) levels are measured in mm Hg and mg/dl, respectively. Fasting blood sugar (fbs) levels above 120 mg/dl are categorized as true (1) and false (0). Resting electrocardiographic results (restecg) detail normal readings, ST-T wave abnormalities, and probable or definite left ventricular hypertrophy.

Thalach denotes the maximum heart rate achieved during a stress test. Other variables include exercise-induced angina (exang), ST depression induced by exercise relative to rest (oldpeak), slope of the peak exercise ST segment (slope), number of major vessels colored by fluoroscopy (ca), and results from the thalium stress test (thal), classified as normal, fixed defect, reversible defect, or not described. The target variable (target) indicates the presence (1) or absence (0) of heart disease, forming the basis for predictive modeling and analysis in cardiovascular research.

Fig. 1 Dataset overview of significant cardiovascular health factors in patients.

Fig. 2 Presents cardiovascular health measures for heart disease patients.

Fig. 3 Overview of Categorical Characteristics in Heart Disease Dataset

Fig. 4 Analyzing the Distribution of Continuous Variables

Through Histograms

This investigation comprehensively explores key cardiovascular health factors using a dataset that includes demographic, clinical, and diagnostic attributes relevant to heart disease. Figure 1 provides an overview of critical metrics such as age, blood pressure, cholesterol levels, maximum heart rate achieved during stress tests, and ST depression induced by exercise relative to rest. It summarizes these variables to highlight their significance in cardiovascular health assessment. Figure 2 zooms in on these metrics specifically among patients diagnosed with heart disease, illustrating how these health measures vary within this subset of individuals. This focused analysis helps discern patterns and trends specific to those affected by heart disease. Moving to categorical variables, Figure 3 presents an overview of variables like sex, chest pain type, fasting blood sugar levels, resting electrocardiographic results, and others. It provides a summary of their distributions within the dataset, offering insights into the categorical characteristics prevalent among the study population. Figures 4 and 5 employ histograms and bar graphs, respectively, to analyze the distribution patterns of continuous and categorical variables. Figure 4 uses histograms to depict the frequency distributions of age, blood pressure, cholesterol levels, and other continuous features, revealing their spread across the dataset.

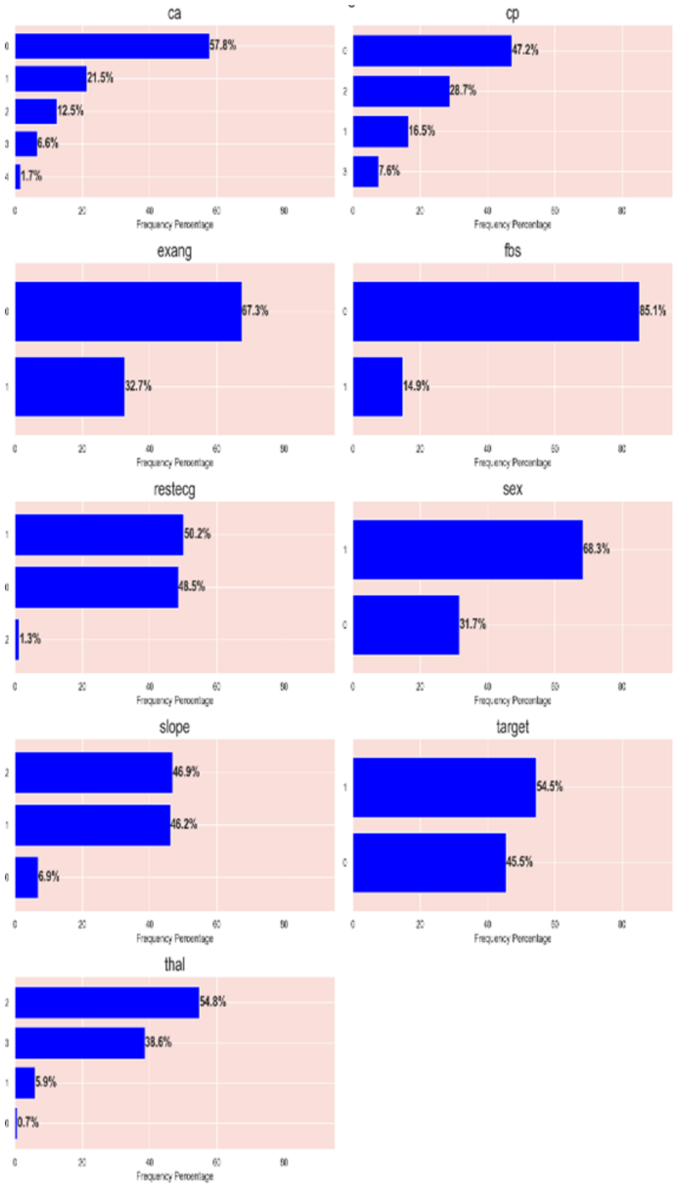

Fig. 5 Exploring Categorical Variables Distribution Using Bar Graph

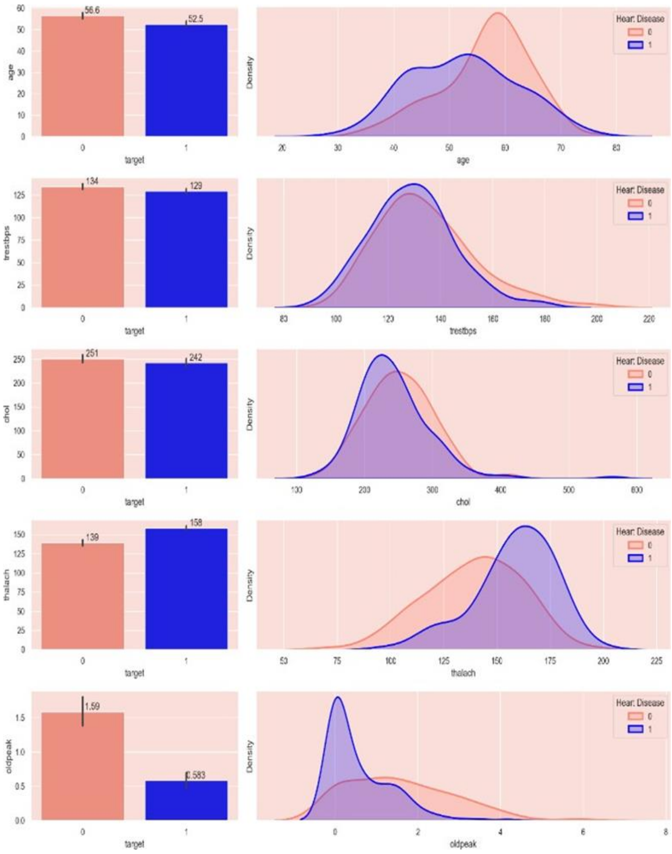

Fig. 6 Analyzing Distribution of Continuous Features in Relation to Heart Disease Target

Figure 5 employs bar graphs to visually explore how categorical variables such as sex and chest pain type are distributed among different categories within the dataset. This comparative visualization aids in understanding the prevalence and distribution of categorical variables relevant to heart disease. Figure 6 delves into the relationship between continuous features and the target variable (presence or absence of heart disease). By analyzing variables like age, blood pressure, and cholesterol levels in relation to heart disease outcomes, it elucidates their potential as predictors or indicators of the disease.

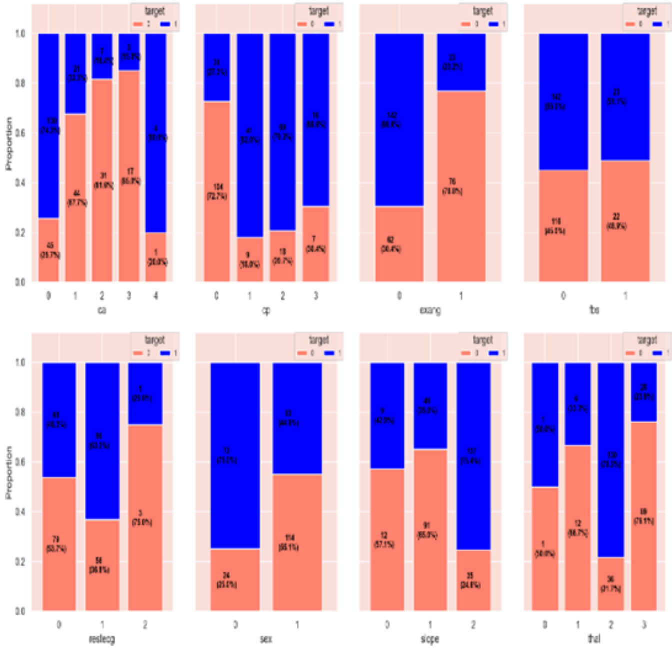

Fig. 7 Comparing Categorical Features with Heart Disease Target Using Stacked Barplots

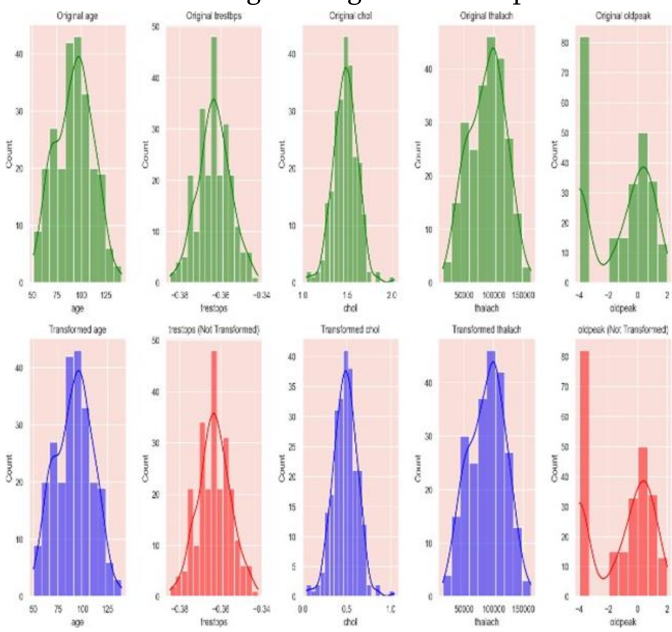

Fig. 8 Examining the Distribution of Continuous Features and Implementing Box-Cox Transformation

Figure 7 utilizes stacked bar plots to compare the distribution of categorical features across the presence and absence of heart disease. It visually contrasts variables like sex and chest pain type to discern their impact on heart disease outcomes, providing insights into how these factors influence disease prevalence. Finally, Figure 8 examines the distribution of continuous features while implementing Box-Cox transformation. This transformation stabilizes variance and enhances normality, thereby improving the suitability of these variables for statistical analyses and predictive modeling related to heart disease. Together, these figures offer a comprehensive exploration of cardiovascular health factors, leveraging visual analytics to uncover patterns, distributions, and relationships crucial for understanding and managing heart disease.

V. CONCLUSION

The findings underscore the critical importance of continuous monitoring and early intervention in managing cardiovascular health. Key insights include the significant impact of age, blood pressure, and cholesterol levels on heart disease outcomes. The categorization of chest pain types and the assessment of electrocardiographic results further enhance our understanding of diagnostic indicators. These insights are pivotal for healthcare providers in optimizing treatment strategies and preventive measures. In the realm of cardiovascular health research, future advancements hold promise in several key areas. Integrating advanced machine learning techniques, such as deep learning models, can enhance the predictive accuracy and efficiency of diagnosing heart disease. Longitudinal studies that track patients over extended periods will provide deeper insights into disease progression and treatment efficacy across diverse populations. Moreover, incorporating genetic profiling and lifestyle data into predictive models can enable personalized risk assessment and intervention strategies. Interactive visualization tools will continue to evolve, offering healthcare providers intuitive platforms for data interpretation and decision-making. Collaborative efforts across disciplines and global datasets will further enrich our understanding of cardiovascular diseases, paving the way for innovative treatments and preventive measures that can significantly impact public health outcomes worldwide.

II. REFERENCES

[1]. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2020. Available from: http://ghdx.healthdata.org/gbd-results-tool.

[2]. World Health Organization. Cardiovascular diseases (CVDs) Fact Sheet. Updated May 2021. Available from: https://www.who.int/news-room/fact-sheets/detail/cardiova scular-diseases-(cvds).

[3]. Benjamin EJ, Virani SS, Callaway CW, et al. Heart Disease and Stroke Statistics—2018 Update: A Report From the American Heart Association. Circulation. 2018;137(12). doi:10.1161/CIR.0000000000000558.

[4]. Yusuf S, Hawken S, Ounpuu S, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet. 2004;364(9438):937-952. doi:10.1016/S0140-6736(04)17018-9.

[5]. Mozaffarian D, Benjamin EJ, Go AS, et al. Heart Disease and Stroke Statistics—2016 Update: A Report From the American Heart Association. Circulation. 2016;133(4) doi:10.1161/CIR.0000000000000350.

[6]. Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. Lancet. 2006;367(9524):1747-1757. doi:10.1016/S0140-6736(06)68770-9.

[7]. Centers for Disease Control and Prevention (CDC). Heart Disease Facts. Updated August 2020. Available from: https://www.cdc.gov/heartdisease/facts.htm.

[8]. Waigi, R.; Choudhary, S.; Fulzele, P.; Mishra, G. Predicting the risk of heart disease using advanced machine learning approach. Eur. J.

Mol. Clin. Med. 2020, 7, 1638–1645. [Google Scholar]

[9]. Breiman, L. Random forests. Mach. Learn. 2001, 45, 5–32. [Google Scholar] [CrossRef]

[10]. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the KDD '16: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794. [Google Scholar] [CrossRef]

[11]. Gietzelt, M.; Wolf, K.-H.; Marschollek, M.; Haux, R. Performance comparison of accelerometer calibration algorithms based on 3D-ellipsoid fitting methods. Comput. Methods Programs Biomed. 2013, 111, 62–71. [Google Scholar] [CrossRef]

[12]. K, V.; Singaraju, J. Decision Support System for Congenital Heart Disease Diagnosis based on Signs and Symptoms using Neural Networks. Int. J. Comput. Appl. 2011, 19, 6–12. [Google Scholar] [CrossRef]

[13]. Narin, A.; Isler, Y.; Ozer, M. Early prediction of Paroxysmal Atrial Fibrillation using frequency domain measures of heart rate variability. In Proceedings of the 2016 Medical Technologies National Congress (TIPTEKNO), Antalya, Turkey, 27–29 October 2016. [Google Scholar] [CrossRef]

[14]. Shah, D.; Patel, S.; Bharti, S.K. Heart Disease Prediction using Machine Learning Techniques. SN Comput. Sci. 2020, 1, 345. [Google Scholar] [CrossRef]

[15]. Alotaibi, F.S. Implementation of Machine Learning Model to Predict Heart Failure Disease. Int. J. Adv. Comput. Sci. Appl. 2019, 10, 261–268. [Google Scholar] [CrossRef]

[16]. Hasan, N.; Bao, Y. Comparing different feature selection algorithms for cardiovascular disease prediction. Health Technol. 2020, 11, 49–62. [Google Scholar] [CrossRef]

[17]. Drożdż, K.; Nabrdalik, K.; Kwiendacz, H.; Hendel, M.; Olejarz, A.; Tomasik, A.; Bartman, W.; Nalepa, J.; Gumprecht, J.; Lip, G.Y.H. Risk factors for cardiovascular disease in patients with metabolic-associated fatty liver disease: A machine learning approach. Cardiovasc. Diabetol. 2022, 21, 240