



# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



**Impact Factor: 8.699**

**Volume 15, Issue 3, March 2026**

# **HEART DISEASE ANALYSIS: Risk Prediction & Insights**

**Pathan Zuveriya Tabassum<sup>1</sup>, Krosuru Chandra Kiran<sup>2</sup>, Mudamanchu Ashok<sup>3</sup>,  
Lingampalli Hemanth Kumar<sup>4</sup>, A Lakshmi Rajeswari<sup>5</sup>**

U.G Students, Department of CSE (Artificial Intelligence & Machine Learning), Kallam Haranadhareddy Institute of  
Technology, Guntur, Andhra Pradesh, India<sup>1-4</sup>

Assistant Professor, Kallam Haranadhareddy Institute of Technology, Guntur, Andhra Pradesh, India<sup>5</sup>

**ABSTRACT:** Heart disease is one of the leading causes of mortality worldwide, emphasizing the importance of early risk assessment and effective data analysis. This project, “*Heart Disease Analysis: Risk Prediction & Insights*,” focuses on using data analytics techniques to identify key factors associated with heart disease and provide meaningful insights. The study begins with dataset preprocessing, including data cleaning, handling missing values, removing inconsistencies, and transforming variables to ensure data quality and reliability.

Exploratory Data Analysis (EDA) is performed to examine patterns, trends, and correlations among various medical attributes such as age, cholesterol levels, blood pressure, and other clinical parameters. The results are presented through interactive dashboards and data storytelling techniques, enabling users to easily interpret complex datasets and derive actionable insights.

By leveraging visualization tools and analytical methods, this project highlights critical risk indicators and supports informed decision-making in healthcare. Overall, it demonstrates the significance of data analytics in understanding heart disease patterns and improving awareness for preventive care and better health outcomes.

**KEYWORDS:** Heart Disease Analysis, Data Analytics, Data Visualization, Data Preprocessing, Dashboard Development, My sql, Flask Web Application.

## **I. INTRODUCTION**

The Heart disease is one of the leading causes of mortality across the globe, making it a major public health concern. The increasing prevalence of cardiovascular conditions highlights the need for effective analysis and early identification of risk factors. This project, “*Heart Disease Analysis: Risk Prediction & Insights*,” aims to explore and analyze healthcare data using data analytics techniques to better understand the patterns associated with heart disease. The primary objective of this study is to examine various clinical and demographic attributes such as age, cholesterol levels, blood pressure, and other health indicators to identify key factors contributing to heart disease. The project begins with comprehensive data preprocessing, including data cleaning, handling missing values, and transforming variables to ensure consistency and reliability of the dataset. Following preprocessing, Exploratory Data Analysis (EDA) is conducted to uncover hidden patterns, trends, and relationships within the data. Various statistical and visualization techniques are used to present the findings in a clear and meaningful way. Interactive dashboards and data storytelling approaches are also incorporated to enhance understanding and communication of insights. By focusing on data-driven analysis rather than complex modeling, this project emphasizes the importance of visualization and interpretation in healthcare analytics. The outcomes provide valuable insights that can support informed decision-making, increase awareness, and contribute to preventive healthcare strategies for reducing the impact of heart disease.

## **II. LITERATURE SURVEY**

Heart disease has been widely studied in the field of healthcare analytics due to its high global mortality rate. Numerous research works emphasize the importance of data-driven approaches to understand risk factors and support early diagnosis. Traditional studies primarily focused on clinical observations, while recent advancements highlight the role of data analytics in extracting meaningful insights from large healthcare datasets.

Several researchers have applied Exploratory Data Analysis (EDA) to identify significant patterns and relationships among attributes such as age, cholesterol, blood pressure, and lifestyle factors. These studies demonstrate that proper data preprocessing, including handling missing values and outliers, plays a crucial role in improving the quality and reliability of analysis. Effective preprocessing ensures that the dataset is suitable for generating accurate and meaningful insights.

In addition, many works have focused on visualization techniques to simplify complex medical data. Tools such as dashboards and graphical representations have been used to present trends and correlations in an interactive and user-friendly manner. Data storytelling has also emerged as an important approach, enabling stakeholders to better understand analytical results and make informed decisions.

Recent literature also highlights the growing importance of healthcare analytics in supporting preventive care. By identifying key risk indicators, researchers have shown that data analytics can assist healthcare professionals in recognizing high-risk patients and improving overall treatment strategies. Furthermore, the integration of analytics with visualization tools enhances accessibility and usability for both technical and non-technical users.

Overall, the reviewed studies confirm that data analytics, combined with effective preprocessing and visualization, plays a vital role in understanding heart disease patterns. These contributions provide a strong foundation for this project, which focuses on generating actionable insights through analytical techniques and interactive dashboards.

### III. METHODOLOGY

The proposed *Heart Disease Analysis: Risk Prediction & Insights* system follows a structured data analytics pipeline consisting of data collection, data preprocessing, exploratory data analysis, visualization, and dashboard development. This methodology focuses on extracting meaningful insights from healthcare data to understand risk factors associated with heart disease and support informed decision-making.

#### Data Collection:

The dataset used in this study consists of structured medical records containing attributes such as age, gender, cholesterol levels, blood pressure, chest pain type, fasting blood sugar, and other clinical parameters, along with a target variable indicating the presence or absence of heart disease. Each record represents an individual patient, enabling detailed analysis of health indicators and their relationship with heart disease. The dataset is obtained from reliable public sources such as the Kaggle repository, which provides standardized and anonymized healthcare data for analysis. Since the dataset includes multiple medical attributes, it offers a comprehensive basis for identifying patterns and trends related to cardiovascular conditions.

#### Data Preprocessing:

Data preprocessing plays a crucial role in ensuring the quality and consistency of the dataset. The raw data is first cleaned by handling missing values, removing duplicates, and correcting inconsistencies. Categorical variables such as chest pain type and gender are transformed into numerical formats for easier analysis. Numerical attributes like cholesterol and blood pressure are checked for outliers and normalized if necessary to maintain uniformity. These steps ensure that the dataset is accurate, structured, and ready for further analytical processes.

#### Exploratory Data Analysis (EDA):

EDA is conducted to examine relationships and patterns among various health indicators. Statistical summaries and correlation analysis are used to identify key attributes that influence heart disease. Trends such as the impact of age on heart conditions or the relationship between cholesterol levels and disease occurrence are analyzed. This step helps in understanding the distribution of variables and detecting significant risk factors.

#### Data Visualization:

Visualization techniques are applied to represent the analyzed data in a clear and intuitive manner. Graphs such as bar charts, histograms, scatter plots, and heatmaps are used to illustrate patterns, correlations, and trends. These visual representations simplify complex data and make it easier to interpret insights.

**Dashboard Development and Storytelling:**

Interactive dashboards are created using tools like Tableau or Power BI to present insights dynamically. These dashboards allow users to filter and explore data based on different attributes, enhancing usability. Data storytelling techniques are incorporated to communicate findings effectively, enabling both technical and non-technical users to understand key insights.

Overall, this methodology emphasizes data cleaning, analysis, and visualization to derive meaningful insights from healthcare data, highlighting critical risk factors and supporting better awareness and preventive healthcare strategies.

**IV. EXPERIMENTAL RESULTS**

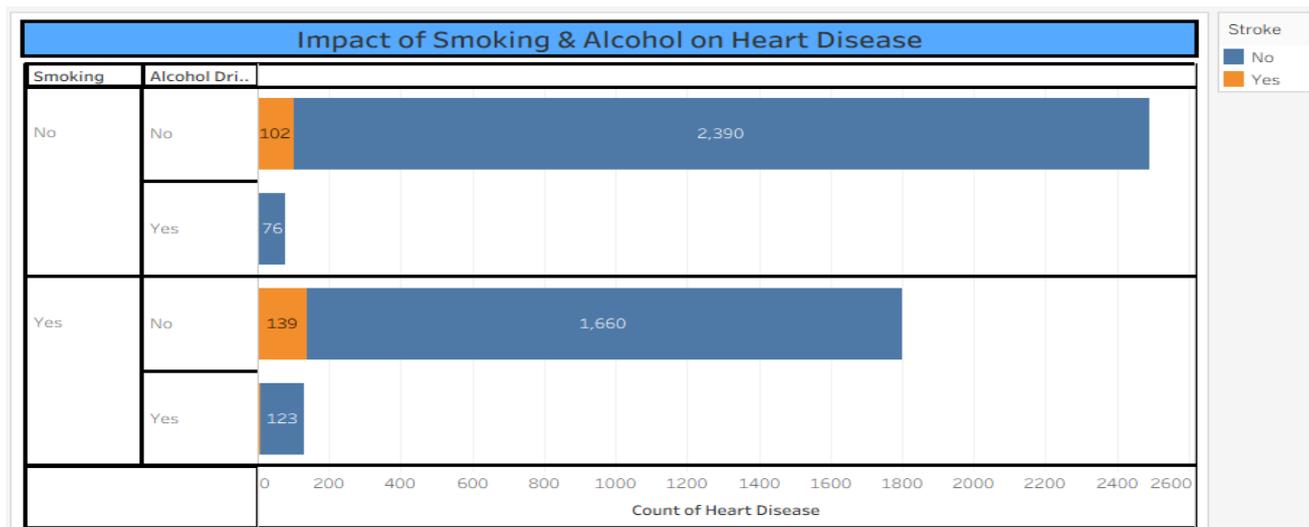
The analytical performance of the heart disease dataset is evaluated using statistical analysis, correlation measures, and visualization techniques to identify key risk factors and patterns. The results are presented through various charts, dashboards, and summary statistics to assess the relationships between clinical attributes and the presence of heart disease.

Among the analyzed factors, attributes such as cholesterol levels, blood pressure, chest pain type, and age show strong correlations with heart disease occurrence. Correlation analysis and heat maps indicate that certain features have a higher influence on the target variable, helping to identify critical health indicators. Visualization results demonstrate clear patterns, such as increased risk in higher age groups and abnormal clinical measurements.

Interactive dashboards further enhance the understanding of these patterns by allowing users to explore data dynamically. Comparative analysis across different attributes highlights variations in risk levels among patient groups. Overall, the results confirm that data analytics techniques effectively reveal meaningful insights, supporting better understanding, awareness, and preventive strategies for heart disease management.

**Fig. 1** illustrates the impact of smoking and alcohol consumption on heart disease occurrence. The chart compares different groups based on smoking and alcohol habits and their association with heart disease cases. Individuals who neither smoke nor consume alcohol show the highest number of non-heart disease cases, with relatively fewer affected individuals.

However, an increase in heart disease cases is observed among smokers, even without alcohol consumption. The risk further increases for individuals who both smoke and consume alcohol, indicating a combined negative effect of these lifestyle factors. Overall, the figure highlights that smoking significantly contributes to heart disease risk, and alcohol consumption further amplifies this effect, emphasizing the importance of healthy lifestyle choices.



A visual comparison of heart disease patterns using analytical visualization techniques created in Tableau. The chart highlights variations in key attributes and their relationship with heart disease occurrence, enabling clear comparison across different categories.

From the visualization, it can be observed that certain attributes show noticeable differences in their contribution to heart disease cases. The graphical representation helps in identifying trends, variations, and possible risk indicators within the dataset. Instead of model-based evaluation, the focus is on interpreting patterns through visual analytics.

Overall, the figure demonstrates how Tableau dashboards effectively transform complex healthcare data into meaningful insights, allowing users to easily understand relationships between variables and supporting data-driven decision-making in heart disease analysis.

Fig. 2: Home Page

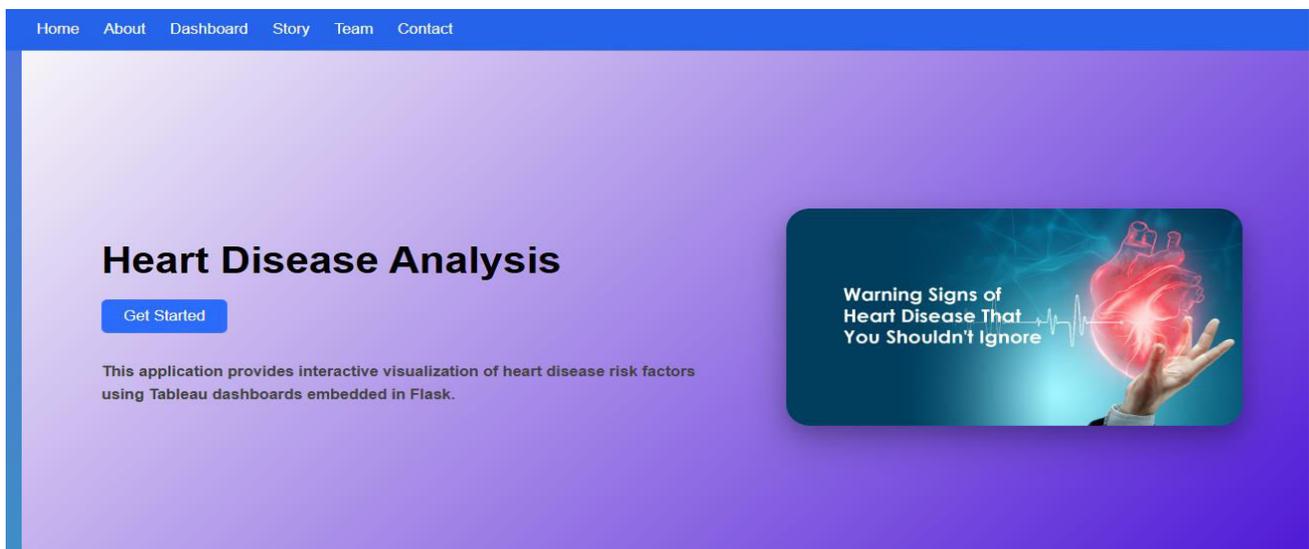


Fig. 3: Predict Page



**Fig. 4: Result Page**

## V. CONCLUSION

This project, “*Heart Disease Analysis: Risk Prediction & Insights*,” demonstrates the effective use of data analytics techniques to understand patterns and key factors associated with heart disease. By focusing on data preprocessing, exploratory data analysis, and visualization, the study successfully transforms raw healthcare data into meaningful and interpretable insights.

The preprocessing stage ensured data quality by handling missing values, removing inconsistencies, and preparing the dataset for analysis. Through Exploratory Data Analysis (EDA), significant relationships between attributes such as age, cholesterol levels, blood pressure, and lifestyle factors were identified. These findings highlight how certain clinical indicators and habits contribute to increased risk of heart disease.

The use of Tableau dashboards and data storytelling played a crucial role in presenting complex data in an interactive and user-friendly manner. Visualizations enabled clear identification of trends, comparisons, and patterns, making the analysis accessible to both technical and non-technical users.

Overall, the project emphasizes the importance of data analytics in healthcare for improving awareness and supporting preventive measures. By providing clear insights into risk factors, this study contributes to better understanding and decision-making in heart disease management. Future work can focus on integrating real-time data and expanding datasets to enhance the depth and applicability of analysis.

## VI. FUTURE SCOPE

The *Heart Disease Analysis: Risk Prediction & Insights* project can be further enhanced by expanding both the dataset and analytical capabilities. Future work may include integrating larger and more diverse healthcare datasets from multiple sources to improve the depth and reliability of insights. Incorporating real-time data from wearable devices and health monitoring systems can provide continuous analysis and more dynamic dashboards.

Advanced visualization techniques and interactive features can be added to Tableau dashboards to enable deeper user exploration and personalized insights. The project can also be extended to include geographic and demographic analysis to better understand regional trends and population-specific risk factors.

In addition, integrating predictive analytics in the future can help forecast potential heart disease risks based on historical patterns. Deployment of the system as a web-based or cloud platform can improve accessibility for healthcare professionals and researchers. Overall, these enhancements can make the system more scalable, insightful, and impactful in supporting preventive healthcare strategies.

#### REFERENCES

- [1] D. Dua and C. Graff, "UCI Repository: Heart Disease Dataset," University of California, Irvine, 2019.
- [2] Kaggle Dataset, "Heart Disease Data for Analysis," Available: [https://drive.google.com/file/d/190Qmq27LeZZ\\_nWricP3Obl7ys\\_5otEsp/view?usp=sharing](https://drive.google.com/file/d/190Qmq27LeZZ_nWricP3Obl7ys_5otEsp/view?usp=sharing)
- [3] S. Few, "Information Dashboard Design: Displaying Data for At-a-Glance Monitoring," Analytics Press, 2006.
- [4] C. Stolte, D. Tang, and P. Hanrahan, "Polaris: A System for Query, Analysis, and Visualization of Multidimensional Databases," IEEE Transactions on Visualization and Computer Graphics, vol. 8, no. 1, pp. 52–65, 2002.
- [5] Tableau Software, "Tableau: Data Visualization and Business Intelligence Platform," Available: <https://www.tableau.com/>
- [6] M. Friendly, "Data Visualization: A Practical Introduction," Sage Publications, 2014.
- [7] E. Tufte, "The Visual Display of Quantitative Information," Graphics Press, 2001.
- [8] N. Knaflic, "Storytelling with Data: A Data Visualization Guide for Business Professionals," Wiley, 2015.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details