



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 Prediction and Analysis using Data Analytics and Visualization Techniques

Mahipathi Rohitha Priya Sai Lakshmi¹, Kakarla Vamsidhar², Koka Gopi Raju³, Ontipuli Sai Kalyan⁴,

Dr T Venkateswarlu⁵

U.G. Students, Department of CSE (Artificial Intelligence & Machine Learning), Kallam Haranadhareddy

Institute of Technology, Guntur, Andhra Pradesh, India¹⁻⁴

Assistant Professor, Kallam Haranadhareddy Institute of Technology, Guntur, Andhra Pradesh, India⁵

ABSTRACT: Heart disease is one of the leading causes of death worldwide, making early detection essential. Analyzing medical data to identify risk factors is challenging due to its complexity and multiple influencing parameters.

This paper presents a data analytics-based Heart Disease Analysis System. The dataset is preprocessed to ensure data quality and stored in the MySQL database. SQL queries are used to extract insights, and Tableau dashboards are created to visualize relationships between health parameters such as age, blood pressure, and cholesterol.

The analysis indicates that factors like increasing age, high cholesterol, and high blood pressure are strongly associated with heart disease. The system is integrated into a Flask-based web application, enabling easy access to visual insights and supporting data-driven healthcare decision-making.

KEYWORDS: Heart Disease, Data Analytics, Healthcare Data, SQL Analysis, Data Visualization, Tableau, Flask Web Application, Risk Factors.

I. INTRODUCTION

Heart disease is one of the leading causes of death worldwide and poses a major challenge to healthcare systems. The increasing availability of healthcare data from medical records, clinical tests, and patient monitoring systems has created opportunities to analyze and understand heart-related conditions more effectively. However, identifying meaningful patterns from large and complex medical datasets remains a difficult task.

Traditional approaches to heart disease diagnosis depend on clinical expertise and basic statistical analysis. While these methods are effective, they may not always capture hidden relationships between multiple health parameters such as age, blood pressure, cholesterol levels, and lifestyle factors. This limitation makes it difficult to perform comprehensive analysis and early risk identification.

In this research, a data analytics-based Heart Disease Analysis System is proposed to study patient data and identify key risk factors. The system involves data preprocessing to ensure data quality, storage of structured data in a MySQL database, and the use of SQL queries for analytical operations. Tableau is used to create interactive dashboards that visually represent patterns and relationships among different health parameters.

The developed system is further integrated into a Flask-based web application, allowing users to access and explore the analysis easily. This approach helps in understanding heart disease patterns, identifying high-risk individuals, and supporting data-driven healthcare decision-making.

II. LITERATURE SURVEY

Heart disease analysis has gained significant attention due to the increasing availability of healthcare data and the growing need for early detection and prevention. Various data analytics and statistical techniques have been used to study medical datasets and identify key risk factors associated with heart disease. In [1], a study on healthcare data analysis highlighted the importance of data preprocessing and exploratory analysis in understanding patient health patterns.

Several studies emphasize the role of data cleaning and preparation in improving the quality of analysis. Handling missing values, removing inconsistencies, and ensuring data integrity are essential steps before performing any analytical operations. Properly prepared data helps in obtaining accurate and reliable insights from healthcare datasets.

Data visualization techniques have also been widely used in medical data analysis. In [2], researchers demonstrated how visualization tools can help in identifying relationships between parameters such as age, blood pressure, cholesterol levels, and heart disease occurrence. Graphical representations make complex medical data easier to understand and support better decision-making.

Database management systems play an important role in storing and processing large volumes of healthcare data. Structured storage using relational databases such as MySQL allows efficient querying and analysis using SQL operations like filtering, grouping, and aggregation.

Recent studies have focused on integrating data analytics with web-based applications to improve accessibility and usability. By combining data processing, visualization tools like Tableau, and web frameworks such as Flask, healthcare analytics systems can provide interactive and user-friendly platforms for analyzing patient data.

The work presented in this paper focuses on developing a Heart Disease Analysis System using data analytics and visualization techniques. The proposed system includes data preprocessing, SQL-based analysis, interactive dashboard creation, and web integration to provide meaningful insights for healthcare decision-making.

III. METHODOLOGY

The proposed Heart Disease Analysis System follows a structured data analytics pipeline consisting of data collection, data preprocessing, database storage, data analysis, visualization, and web integration. This methodology helps in identifying important risk factors and understanding patterns associated with heart disease.

Data Collection:

The dataset used in this study consists of structured patient health records containing attributes such as age, gender, blood pressure, cholesterol level, heart rate, and other clinical parameters. Each record represents an individual patient, along with a target variable indicating the presence or absence of heart disease. The dataset was obtained from a reliable online source [7] and is used for analytical purposes.

Data Preprocessing:

Data preprocessing is an essential step to ensure the quality and accuracy of analysis. The dataset is cleaned by handling missing values, removing duplicate records, and correcting inconsistencies. Data types are verified, and the dataset is organized into a structured format suitable for further processing. These steps help in transforming raw medical data into a clean and reliable dataset.

Database Storage:

The processed dataset is stored in a MySQL database to enable efficient data management and querying. Using a relational database structure allows organized storage of patient records and supports easy retrieval of data for analysis.

Data Analysis:

SQL queries are used to perform analytical operations on the dataset. Techniques such as filtering, grouping, aggregation, and conditional analysis are applied to study relationships between different health parameters and heart disease occurrence. This helps in identifying patterns and key risk factors.

Data Visualization:

Tableau is used to create interactive dashboards and charts that visually represent the analysis results. Visualizations such as age-wise distribution, gender-based analysis, and health parameter comparisons help in understanding trends and correlations within the dataset.

Web Integration:

The developed dashboards are integrated into a Flask-based web application to provide user-friendly access. The web interface allows users to view and interact with the visualizations easily. This integration enhances accessibility and enables users to explore heart disease analysis results effectively.

This methodology provides a complete data analytics framework for analyzing heart disease data and supports better understanding and decision-making in healthcare.

IV. EXPERIMENTAL RESULTS

The analysis of heart disease data was carried out using SQL queries and visualized through Tableau dashboards. Various charts and graphical representations were used to examine patterns and relationships among different health parameters.

The results indicate that the occurrence of heart disease increases with age, with a higher concentration of cases observed among middle-aged and elderly individuals. Age-wise analysis clearly shows that individuals above 50 years are at greater risk.

Further observations reveal that high cholesterol levels and elevated blood pressure are significant factors associated with heart disease. Patients with these conditions show a higher likelihood of being affected.

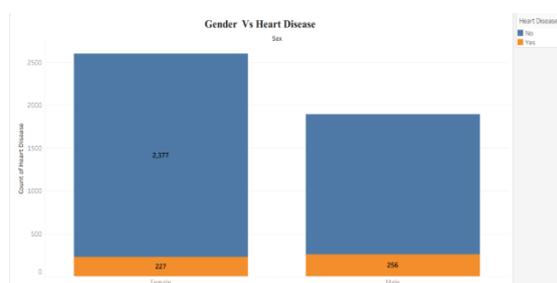
Gender-based analysis highlights variations in heart disease cases between male and female patients. The results indicate that heart disease occurrence is comparatively higher among male patients than female patients, providing important insights into demographic differences. In addition, lifestyle factors such as smoking and alcohol consumption contribute to increased risk levels.

The analysis also demonstrates the relationship between heart disease and other health conditions such as stroke and diabetes, indicating interconnected health risks.

Fig. 1 shows the Gender vs Heart Disease analysis, illustrating the distribution of heart disease cases among male and female patients. This visualization helps in understanding demographic differences in heart disease occurrence.

Overall, the results demonstrate that data analytics and visualization techniques are effective in analyzing complex healthcare data and identifying high-risk patient groups.

Fig. 1



Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503198|

The proposed Heart Disease Analysis System was implemented using data analytics and visualization techniques and evaluated using a healthcare dataset. The dataset was preprocessed to handle missing values, remove inconsistencies, and ensure data quality. SQL queries were applied to analyze the data and extract meaningful insights related to heart disease risk factors.

In addition to data analysis, a risk prediction feature was incorporated to estimate the likelihood of heart disease based on key parameters such as age, blood pressure, cholesterol levels, and lifestyle factors. This prediction is based on analytical patterns observed in the dataset, helping in identifying high-risk individuals.

The analysis results were visualized using Tableau dashboards, providing clear insights into relationships between different health parameters. These visualizations assist in understanding trends and identifying significant contributing factors.

The system was further integrated into a Flask-based web application to provide an interactive and user-friendly interface. The application allows users to explore dashboards and view predicted risk levels based on input data. The screenshots of the application are shown in Fig. 2, Fig. 3, Fig. 4, and Fig. 5.

Fig. 2 shows the home page. Fig. 3 displays the risk prediction input interface. Fig. 4 shows the risk prediction output, indicating the likelihood of heart disease as Low, Moderate, or High. Fig. 5 presents the dashboard interface with visual analytics.

Fig. 2: Home Page

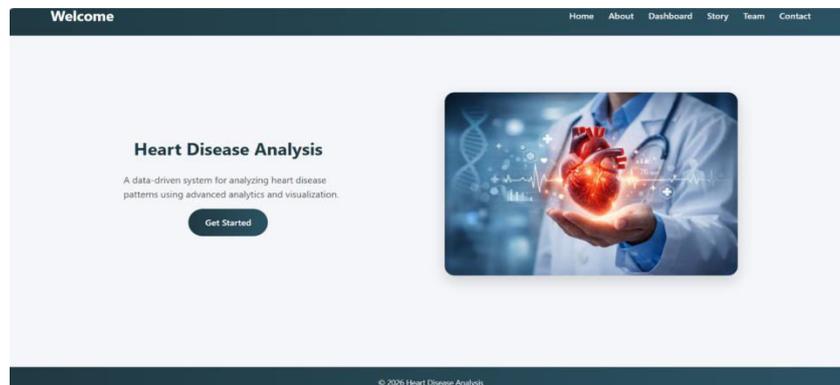


Fig. 3: Predict Page

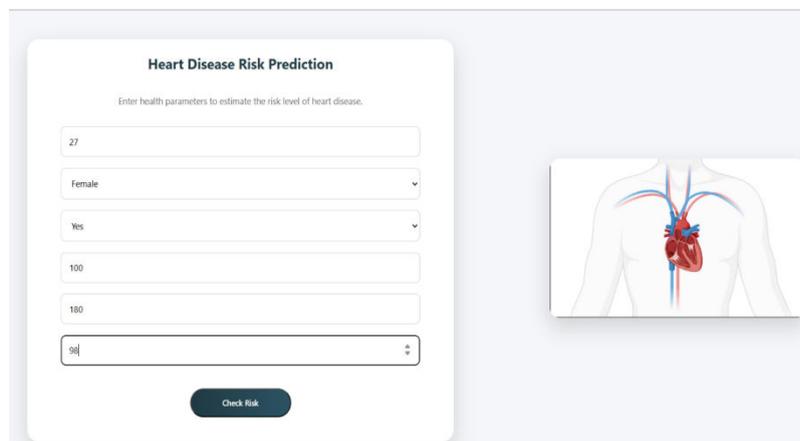


Fig. 4: Result Page

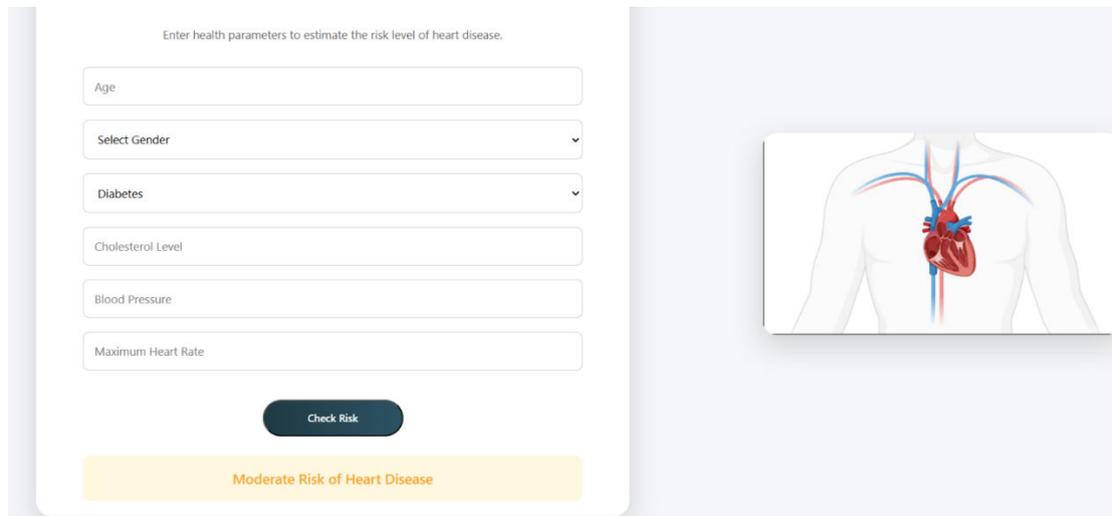
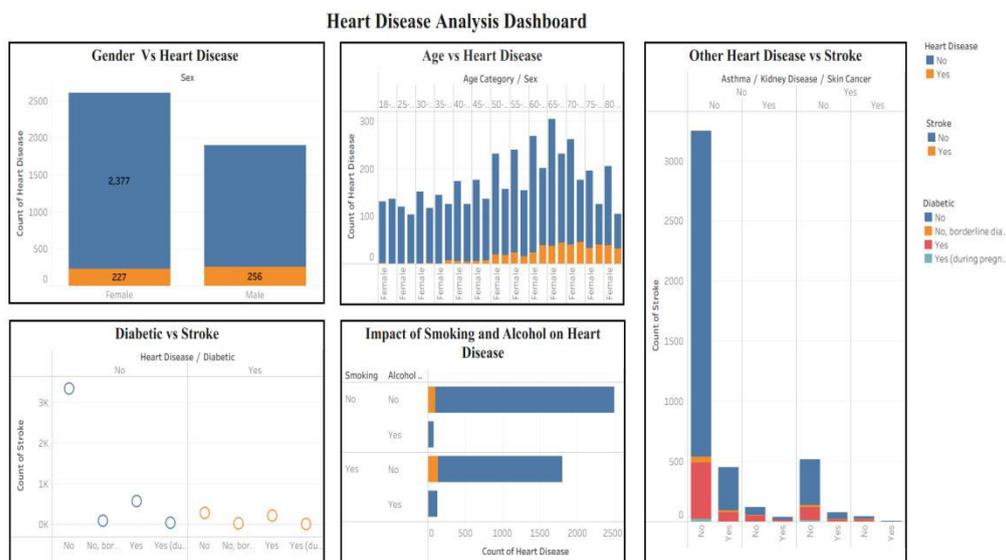


Fig. 5: Dashboard interface



V. CONCLUSION

This research presents a data analytics-based approach for analyzing heart disease data to identify key risk factors. With the increasing availability of healthcare data, analyzing patient information has become essential for early detection and prevention of heart disease.

The dataset was preprocessed to ensure data quality, and SQL queries were used to extract meaningful insights. Visualization techniques using Tableau were applied to represent relationships between health parameters such as age, blood pressure, cholesterol, and lifestyle factors.

The analysis results showed that factors such as increasing age, high cholesterol, high blood pressure, and lifestyle habits significantly contribute to heart disease occurrence. A risk prediction feature was also incorporated to classify individuals into Low, Moderate, or High risk categories based on input parameters.

Volume 15, Issue 3, March 2026**|DOI: 10.15680/IJRSET.2026.1503198|**

The system was integrated into a Flask-based web application, allowing users to explore visual insights and view risk predictions easily. The proposed system demonstrates the effectiveness of data analytics and visualization techniques in understanding heart disease patterns and supporting healthcare decision-making.

VI. FUTURE SCOPE

Although the proposed Heart Disease Analysis System provides effective insights using data analytics and visualization techniques, there are several opportunities for further improvement.

In the future, the system can be enhanced by using larger and more diverse healthcare datasets to improve the accuracy and depth of analysis. Additional health parameters such as genetic factors, diet, and physical activity can also be included for more comprehensive evaluation.

The risk prediction feature can be further refined by incorporating more detailed analytical rules and real-time patient data. Integration with wearable health devices can enable continuous monitoring and provide timely risk assessment.

Another possible improvement is the development of a real-time healthcare monitoring system that can track patient health conditions and generate alerts for high-risk individuals. Expanding the system into a full-scale healthcare platform can improve accessibility and usability for both patients and medical professionals.

Overall, future work can focus on enhancing data quality, expanding system capabilities, and improving real-time analysis for better healthcare decision-making.

REFERENCES

- [1] A. Raghupathi and V. Raghupathi, "Big Data Analytics in Healthcare: Promise and Potential," Health Information Science and Systems, vol. 2, no. 3, 2014.
- [2] S. Kumar and M. Singh, "Data Analytics in Healthcare: A Review," International Journal of Data Science and Analytics, 2020.
- [3] J. Han, M. Kamber, and J. Pei, "Data Mining: Concepts and Techniques," 3rd ed., Morgan Kaufmann, 2011.
- [4] R. Elmasri and S. B. Navathe, "Fundamentals of Database Systems," 7th ed., Pearson, 2016.
- [5] Tableau Software, "Tableau Data Visualization Tool," Available: <https://www.tableau.com/>
- [6] Oracle Corporation, "MySQL: Relational Database Management System," Available: <https://dev.mysql.com/doc/>
- [7] Pallets Projects, "Flask: A Python Micro Web Framework," Available: <https://flask.palletsprojects.com/>
- [8] Heart Disease Dataset, "Patient Health Records for Heart Disease Analysis," Available: https://drive.google.com/file/d/190Qmq27LeZZ_nWricP3Obl7ys_5otEsp/view



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details