IJIREEICE
ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.414 ⫶ Peer-reviewed & Refereed journal ⫶ Vol. 14, Issue 3, March 2026
DOI: 10.17148/IJIREEICE.2026.14365

An Optimized Machine Learning Framework for Heart Failure Patient Classification and Risk Prediction

K. Dinesh[1], K. Harika[2], M. Rohith[3], Sandi Sunanda[4]

B.Tech IV year, CSE Department, GNITC-Gurunanak Institutions Technical Campus Hyderabad[1-3]

Assistant Professor, CSE Department, GNITC-Gurunanak Institutions Technical Campus Hyderabad[4]

Abstract: Heart failure is a major global health problem, and early prediction of patient risk can significantly improve treatment outcomes. This project presents a machine learning-based system for classifying heart failure patients into high-risk and low-risk survival categories using clinical data. The dataset consists of approximately 5000 patient records with important medical features such as age, ejection fraction, serum creatinine, and blood pressure. Data preprocessing techniques including feature scaling, feature selection using SelectKBest, and class balancing using SMOTE were applied to improve model performance. Multiple machine learning algorithms were evaluated, including Logistic Regression and XGBoost. Among them, the XGBoost model demonstrated the best performance, achieving an accuracy of 99.70% with high precision and recall. A Flask-based web application was also developed to allow users to input patient data and obtain real-time risk predictions.

I. INTRODUCTION

Introduction: Heart failure is a major cardiovascular disease with high mortality worldwide. Early prediction of patient survival risk can support timely clinical intervention and better patient management. Traditional diagnostic methods often struggle to analyze large clinical datasets and complex medical relationships. Machine learning techniques help identify hidden patterns in healthcare data and improve disease prediction. This study proposes a machine learning framework to classify heart failure patients into high-risk or low-risk categories using clinical features such as age, serum creatinine, and ejection fraction, supporting accurate prediction and clinical decision-making

II. RELATED WORK

Several researchers have explored the application of machine learning techniques in the healthcare domain to improve early detection and prediction of heart-related diseases. Most studies utilize clinical datasets containing important medical parameters such as age, blood pressure, serum creatinine, and ejection fraction to build predictive models.

For instance, Chicco Davide and Jurman Giuseppe (2020) demonstrated that machine learning models can effectively predict heart failure survival using clinical data, emphasizing the significance of features like serum creatinine and ejection fraction. Similarly, other studies have implemented classification algorithms such as Logistic Regression, Decision Trees, and Support Vector Machines to categorize patients into high-risk and low-risk groups.

Recent research has increasingly focused on ensemble learning methods such as XGBoost due to their high accuracy and ability to handle complex, non-linear relationships in medical data. Additionally, techniques like SMOTE have been widely used to address class imbalance issues, improving the prediction of minority classes such as high-risk patients.

Feature selection methods like SelectKBest are also commonly applied to identify the most relevant clinical attributes, reducing dimensionality and enhancing model performance. Furthermore, web-based applications developed using frameworks such as Flask have been used to deploy these models, enabling real-time prediction and assisting healthcare professionals in decision-making.

These related works demonstrate the effectiveness of integrating machine learning algorithms, data preprocessing techniques, and web technologies to develop accurate and efficient heart failure prediction systems.

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.414 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 14, Issue 3, March 2026
DOI: 10.17148/IJIREEICE.2026.14365

III. LITERATURE SURVEY

[1] T. A. McDonagh *et al.*, *"2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure,"* European Heart Journal, vol. 42, no. 36, pp. 3599–3726, 2021.
This research paper provides comprehensive clinical guidelines for diagnosing and treating heart failure, covering both acute and chronic conditions. Conducted by Theresa A McDonagh *et al.* (2021), the study outlines standardized medical procedures, risk assessment strategies, and treatment protocols used by healthcare professionals worldwide. It highlights key clinical indicators such as ejection fraction, blood pressure, and serum biomarkers that are critical for evaluating heart failure severity. These guidelines serve as a foundational reference for selecting relevant features in machine learning models. The study also emphasizes early diagnosis and risk stratification, which aligns with predictive analytics approaches used in this project. Overall, it provides essential medical context that supports the development of accurate and clinically relevant prediction systems.

[2] G. Savarese and L. H. Lund, *"Global public health burden of heart failure,"* Cardiac Failure Review, vol. 3, no. 1, pp. 7–11, 2017.
This paper discusses the global impact of heart failure as a major public health issue. Conducted by Giuseppe Savarese and Lars H Lund (2017), the study highlights the increasing prevalence, mortality rates, and economic burden associated with heart failure worldwide. It explains how aging populations, lifestyle changes, and comorbidities contribute to the rising number of cases. The research emphasizes the need for early detection systems and efficient risk prediction models to reduce hospitalizations and mortality. This work justifies the importance of applying machine learning techniques in healthcare, particularly for predicting patient survival and improving treatment outcomes.

[3] S. S. Virani *et al.*, *"Heart disease and stroke statistics—2021 update: A report from the American Heart Association,"* Circulation, vol. 143, no. 8, pp. e254–e743, 2021.
This report provides detailed statistical insights into heart disease and stroke prevalence, risk factors, and mortality rates. Conducted by Salim S Virani *et al.* (2021), it presents large-scale epidemiological data that highlights the severity and widespread nature of cardiovascular diseases. The study identifies key risk factors such as age, hypertension, diabetes, and lifestyle habits, which are crucial for predictive modeling. These statistical findings support the selection of input features used in machine learning models for heart failure prediction. Additionally, the report underscores the importance of data-driven healthcare solutions to address growing cardiovascular risks.

[4] A. Rajkomar, J. Dean, and I. Kohane, *"Machine learning in medicine,"* New England Journal of Medicine, vol. 380, no. 14, pp. 1347–1358, 2019.
This research paper explores the transformative role of machine learning in modern healthcare. Conducted by Alvin Rajkomar, Jeff Dean, and Isaac Kohane (2019), the study explains how machine learning algorithms can analyze large-scale medical data to improve diagnosis, treatment, and patient outcomes. It highlights applications such as disease prediction, medical imaging analysis, and personalized medicine. The paper also discusses challenges like data quality, interpretability, and ethical concerns. This work provides a strong theoretical foundation for applying machine learning models like Logistic Regression and XGBoost in heart failure prediction systems.

[5] S. Angraal *et al.*, *"Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction,"* JACC: Heart Failure, vol. 8, no. 1, pp. 12–21, 2020.
This study focuses on predicting mortality and hospitalization risks in heart failure patients using machine learning techniques. Conducted by Saurabh Angraal *et al.* (2020), the research applies advanced predictive models to clinical datasets to identify high-risk patients. It highlights the importance of features such as ejection fraction, creatinine levels, and patient history in determining outcomes. The study demonstrates how machine learning models can outperform traditional statistical approaches in capturing complex relationships within medical data. It also emphasizes the role of predictive analytics in improving patient management and reducing hospital readmissions. This research directly supports the use of advanced algorithms like XGBoost in achieving high prediction accuracy in this project.

IV. PROPOSED WORK

MODULES:
1. Data Collection Module
Components: Clinical dataset containing patient records (age, ejection fraction, serum creatinine, blood pressure, etc.)
Purpose: Collect and organize patient medical data for analysis.

DOI: 10.17148/IJIREEICE.2026.14365

Functionality:

- Gathers structured healthcare data from hospital records or datasets.
- Ensures relevant clinical features are included for prediction.

2. Data Preprocessing Module
Purpose: Prepare raw data for machine learning models.
Functionality:

- Handles missing values and inconsistent data.
- Applies feature scaling (normalization/standardization).
- Balances dataset using SMOTE.

3. Feature Selection Module
Purpose: Identify the most important features affecting heart failure prediction.
Functionality:

- Uses SelectKBest to select top features.
- Reduces dimensionality and improves model performance.

4. Machine Learning Module
Purpose: Train models to classify patients into high-risk and low-risk categories.
Algorithms Used:

- Logistic Regression
- XGBoost

Functionality:

- Trains models on processed data.
- Learns patterns and relationships between clinical features and outcomes.

5. Model Evaluation Module
Purpose: Measure and compare model performance.
Metrics:

- Accuracy
- Precision
- Recall
- F1-Score

Functionality:

- Evaluates different models.
- Selects the best-performing model (XGBoost).

6. Flask Web Application Module
Purpose: Provide a user-friendly interface for prediction.
Technology Used: Flask
Features:

- Input patient details through a web form.
- Display real-time prediction (high-risk / low-risk).

7. Prediction Module
Purpose: Generate real-time heart failure risk predictions.
Functionality:

- Takes user input from the web app.
- Uses trained model to predict survival risk.
- Displays result instantly.

8. Data Visualization Module (Optional Enhancement)
Purpose: Provide insights into patient data and model performance.
Features:

- Graphs for feature importance.
- Model accuracy comparison charts.
- Risk distribution visualization.

9. Deployment Module
Purpose: Make the system accessible for real-world use.
Functionality:

- Deploys the Flask application on a server/cloud platform.
- Enables access from anywhere.

DOI: 10.17148/IJIREEICE.2026.14365

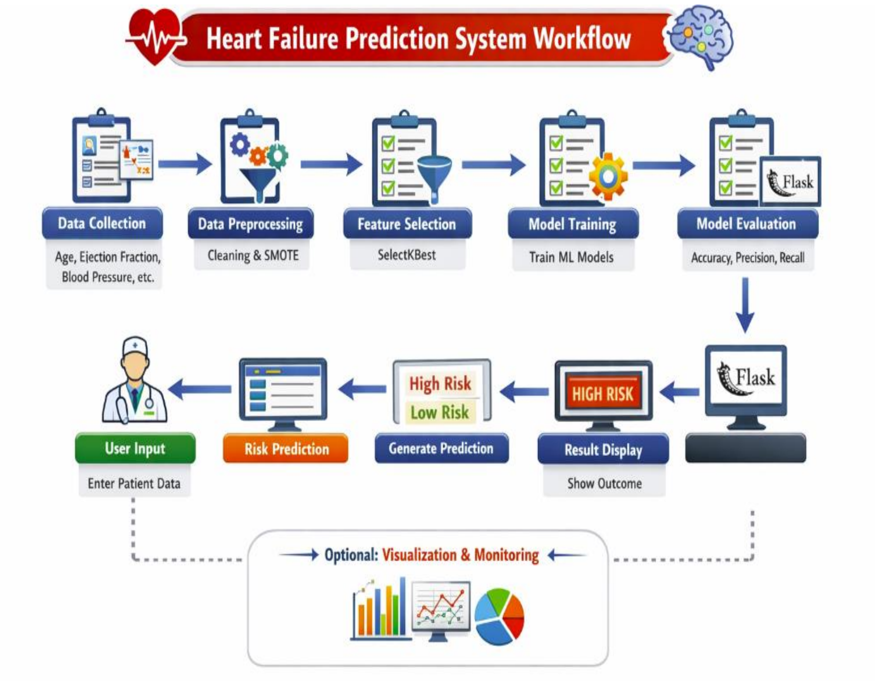

Fig: Workflow of Methodology

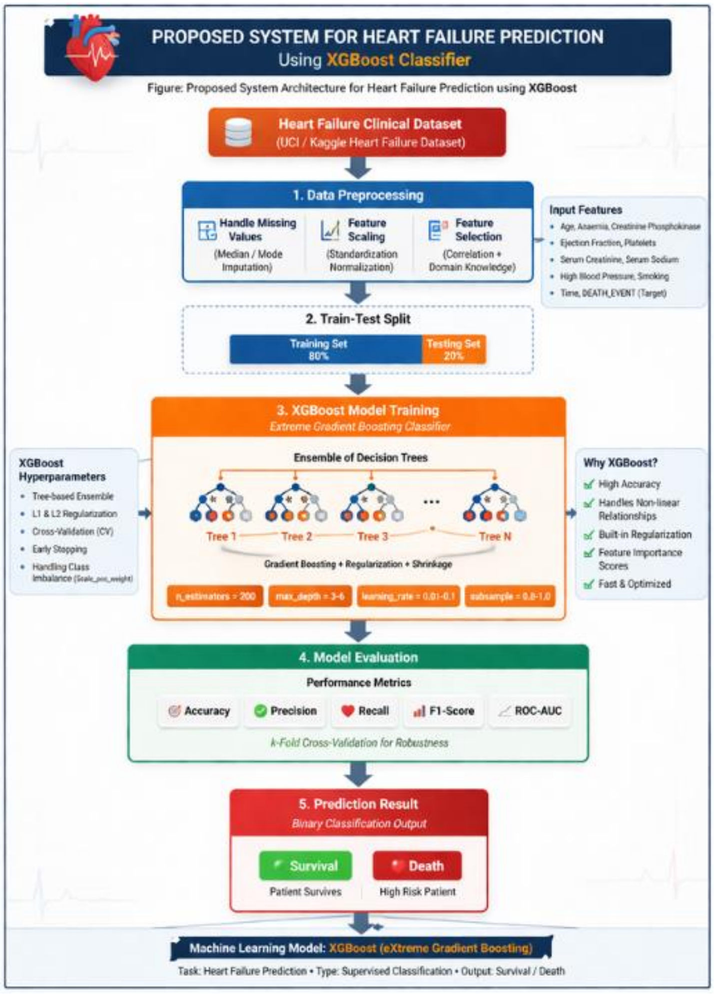

Fig: Proposed Model

DOI: 10.17148/IJIREEICE.2026.14365

V. RESULTS

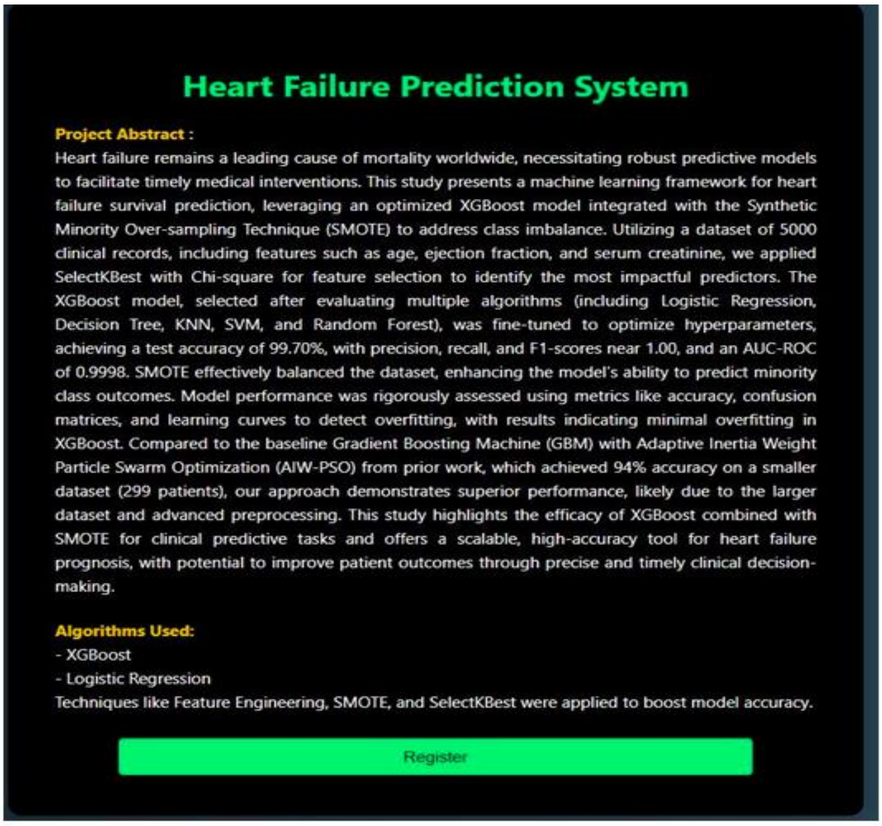

Fig.No:7.1: Home Page

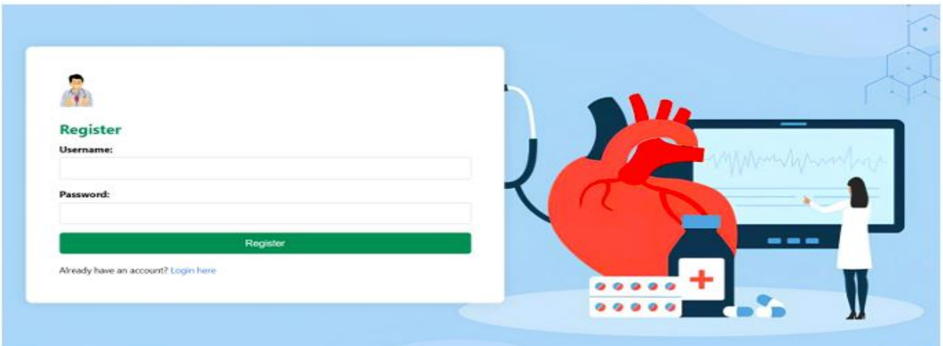

Fig.No:7.2: Registration Page

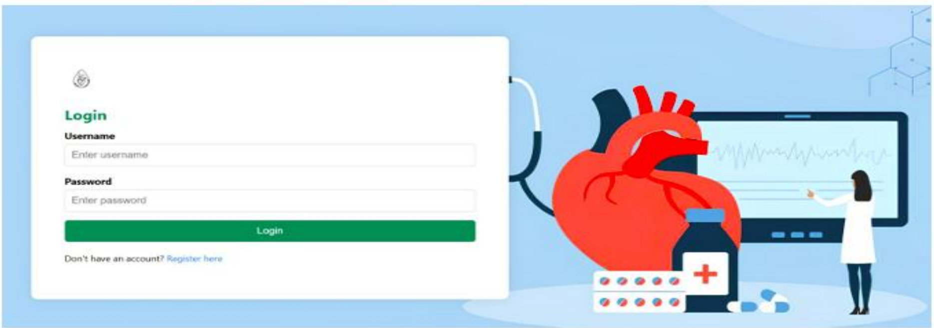

Fig.No:7.3: Login Page

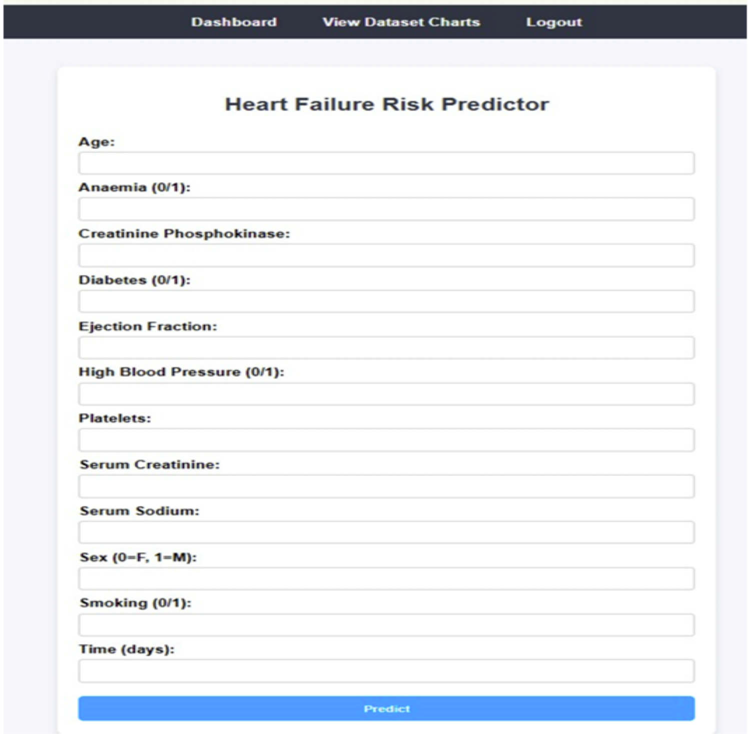

Fig.No:7.4: Dashboard

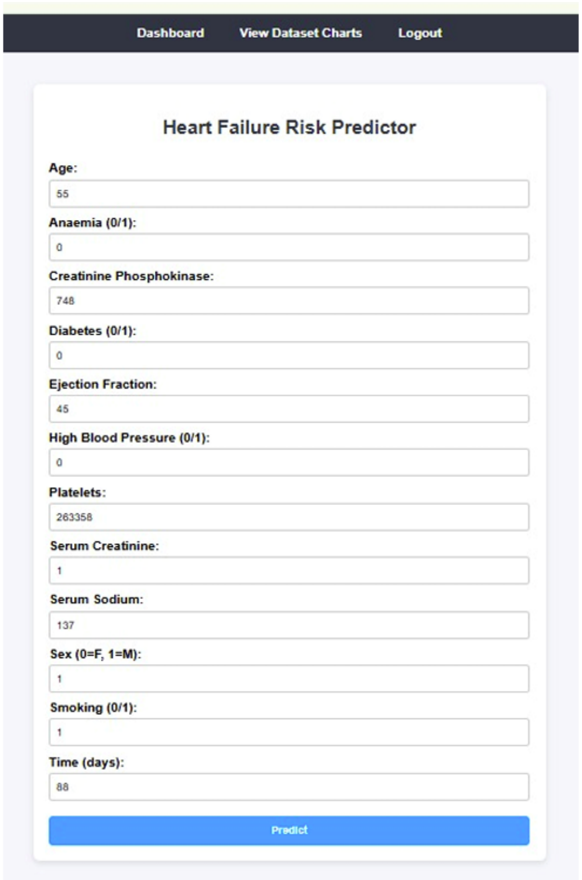

Fig.No:7.5: Input Page

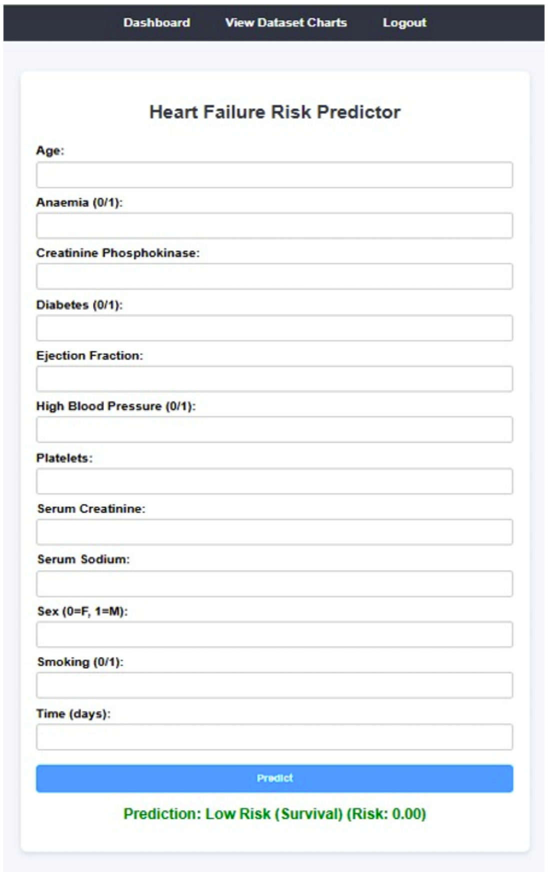

Fig.No:7.6: Result Page

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.414 ⁝⁝ Peer-reviewed & Refereed journal ⁝⁝ Vol. 14, Issue 3, March 2026
DOI: 10.17148/IJIREEICE.2026.14365

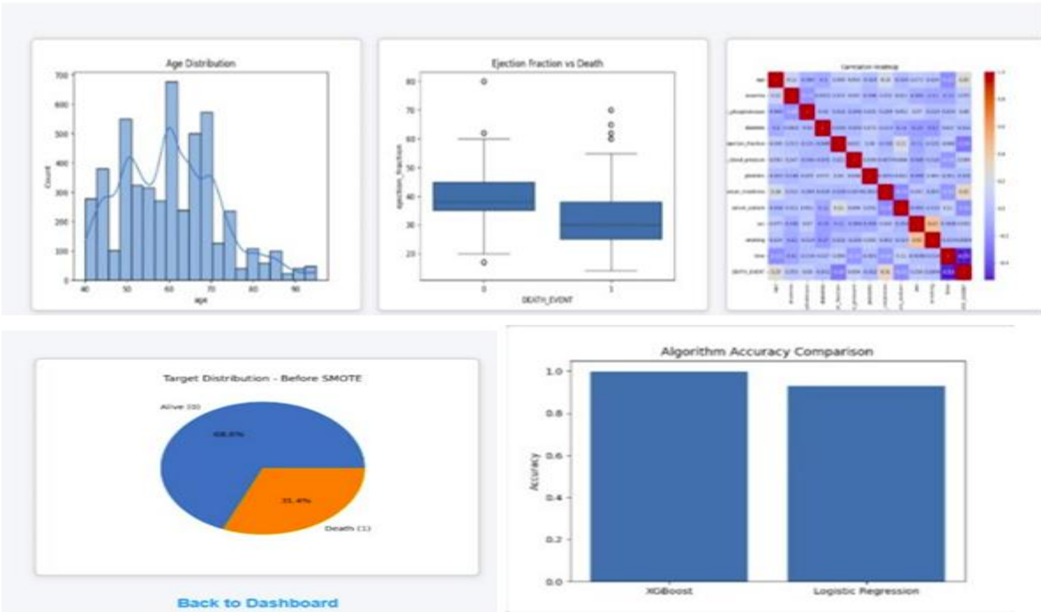

Fig.No:7.7: Dataset charts

Fig.No:7.1 – Home Page
Description:
This is the main landing page of the Heart Failure Prediction System. It provides an overview of the project, including its objective, algorithms used, and system capabilities.
Key Features:

- Displays project abstract and purpose
- Highlights algorithms like XGBoost and Logistic Regression
- Mentions techniques such as SMOTE and SelectKBest
- Contains a Register button to start using the system

Purpose:
Provides users with basic understanding and navigation entry point.

Fig.No:7.2 – Registration Page
Description:
This page allows new users to create an account in the system.
Key Features:

- Input fields for username and password
- Register button for account creation
- Option to navigate to login page

Purpose:
Ensures secure access and user management.

Fig.No:7.3 – Login Page
Description:
This page is used by registered users to log into the system.
Key Features:

- Username and password input
- Login button
- Link to registration page

Purpose:
Authenticates users before accessing prediction features.

DOI: 10.17148/IJIREEICE.2026.14365

Fig.No:7.5 – Input (Prediction) Page

Description:

This is the core page where users enter patient medical details for prediction.

Input Parameters Include:

- Age
- Anaemia
- Creatinine Phosphokinase
- Diabetes
- Ejection Fraction
- High Blood Pressure
- Platelets
- Serum Creatinine
- Serum Sodium
- Sex
- Smoking
- Time (days)

Result Page (Fig. 7.6)

The Result Page represents the Flask Web Application and Prediction Module. Users enter patient details and click "Predict". The system processes the input and gives an instant result showing whether the patient is at high risk or low risk of heart failure, along with a risk score.

Dataset Charts (Fig. 7.7)

The Dataset Charts represent the Data Visualization Module. They include graphs like age distribution, correlation heatmap, and model accuracy comparison. These charts help in understanding the data, identifying important features, and comparing model performance.

VI. CONCLUSION AND FUTURE SCOPE

This project successfully demonstrates the application of machine learning techniques in predicting heart failure patient outcomes. By combining advanced preprocessing techniques with the XGBoost algorithm, the system achieved very high predictive accuracy. The developed model can assist healthcare professionals in identifying high-risk patients early and making informed clinical decisions, potentially improving patient survival rates.

• Integration with hospital electronic health record systems

• Implementation of deep learning models such as ANN

• Real-time patient monitoring using wearable devices

• Deployment on cloud platforms for large-scale healthcare use

REFERENCES

[1]. T. A. McDonagh et al., ''2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure,'' Eur. Heart J., vol. 42, no. 36,pp. 3599–3726, Sep. 2021, doi: 10.1093/eurheartj/ehab368.

[2]. G. Savarese and L. H. Lund, ''Global public health burden of heart failure,'' Cardiac Failure Rev., vol. 3, no. 1, pp. 7–11, 2017, doi:10.15420/cfr.2016:25:2.

[3]. S. S. Virani et al., ''Heart disease and stroke statistics—2021 update: A report from the American heart association,'' Circulation, vol. 143, no. 8, pp. e254–e743, Feb. 2021, doi: 10.1161/cir.0000000000000950.

[4]. A. Rajkomar, J. Dean, and I. Kohane, ''Machine learning in medicine,'' New England J. Med., vol. 380, no. 14, pp. 1347–1358, Apr. 2019, doi: 10.1056/nejmra1814259.

[5]. S. Angraal, B. J. Mortazavi, A. Gupta, R. Khera, T. Ahmad, N. R. Desai, D.L.Jacoby,F. A.Masoudi,J.A.Spertus,andH.M.Krumholz,''Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction,'' JACC, Heart Failure, vol. 8, no. 1, pp. 12–21,Jan. 2020, doi: 10.1016/j.jchf.2019.06.013.

[6]. M. Feurer and F. Hutter, ''Hyperparameter optimization,'' in Automated Machine Learning, F. Hutter, L. Kotthoff, and J. Vanschoren, Eds., Cham, Switzerland: Springer, 2019, pp. 3–33, doi: 10.1007/978-3-030-05318-5_1.

[7]. D. Wang, D. Tan, and L. Liu, ''Particle swarm optimization algorithm: An overview,'' Soft Comput., vol. 22, no. 2, pp. 387–408, Jan. 2018, doi: 10.1007/s00500-016-2474-6.

[8]. Y. Xue, J. Jiang, B. Zhao, and T. Ma, ''A self-adaptive artificial bee colony algorithm based on global best for

DOI: 10.17148/IJIREEICE.2026.14365

global optimization,'' Soft Comput., vol. 22, no. 9, pp. 2935–2952, May 2018, doi: 10.1007/s00500-017-2547-1.

[9]. T. Ahmad, A. Munir, S. H. Bhatti, M. Aftab, and M. A. Raza, ''Survival analysis of heart failure patients: A case study,'' PLoS ONE, vol. 12, no. 7, Jul. 2017, Art. no. e0181001, doi: 10.1371/journal.pone.0181001.

[10]. A. Fernandez, S. Garcia, F. Herrera, and N. V. Chawla, ''SMOTE for learning from imbalanced data: Progress and challenges, marking the 15 year anniversary,'' J. Artif. Intell Res., vol. 61, pp. 863–905, Apr. 2018,doi: 10.1613/jair.1.11192.