IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 2, April 2025

Impact Factor: 7.67

Cardiac Disease Detection using Machine Learning

Kaariniya SA[1] and Dr. M. Praneesh[2]

UG Student, Department of Computer Science with Data Analytics[1]

Assistant Professor, Department of Computer Science with Data Analytics[2]

Sri Ramakrishna College of Arts & Science, Coimbatore, TamilNadu, India

Abstract: *This project Cardiac Disease Detection Using Machine Learning undertakes a comprehensive study of machine learning algorithms for heart disease prediction, aiming to enhance early detection and preventive healthcare. It systematically compares five traditional algorithms—KNN, LDA, Decision Tree, Logistic Regression, and Naive Bayes—against a Convolutional Neural Network (CNN), seeking to identify the most effective approach. Utilizing a rich dataset of cardiovascular risk factors and clinical measurements, the project trains and evaluates predictive models, ensuring a foundation in real-world patient data. The dataset's diverse features enable the models to capture complex, non-linear relationships inherent in medical data. The inclusion of a CNN is a key innovation. CNNs, typically used in image analysis, are adapted to process structured medical data, potentially surpassing traditional algorithms by learning hierarchical feature representations. This approach aims to uncover subtle indicators of heart disease, demonstrating the potential of advanced techniques in medical diagnostics. Evaluation emphasizes clinical relevance, focusing on accuracy, sensitivity, and specificity. Sensitivity, identifying those with heart disease, and specificity, identifying those without, are crucial for minimizing errors. The CNN's superior performance across these metrics highlights its potential for clinical decision support. This research aims to provide healthcare professionals with a reliable risk assessment tool, facilitating earlier interventions and personalized prevention. By demonstrating the CNN's capabilities and conducting a thorough comparative analysis, the project contributes to the understanding of machine learning's role in medical diagnostics. It paves the way for proactive healthcare strategies, leveraging advanced techniques to combat heart d*isease.

Keywords: K-Nearest Neighbors (KNN), Linear Discriminant Analysis (LDA), Decision Tree, Logistic Regression, Naive Bayes

I. INTRODUCTION

1.1 Statement of the problem

This In global health landscape is profoundly impacted by cardiovascular diseases (CVDs), a leading cause of mortality that exacts a staggering toll of approximately 17.9 million lives annually. This stark reality underscores the urgent need for innovative strategies to enhance early detection and management of heart disease, shifting from reactive treatment to proactive prevention. The critical importance of early intervention lies in its potential to mitigate the progression of CVDs, improve patient outcomes, and ultimately reduce the immense burden on healthcare systems worldwide.

In this context, machine learning emerges as a transformative force, offering the promise of developing sophisticated predictive models capable of identifying individuals at heightened risk of developing heart disease. These models transcend the limitations of traditional risk assessment tools by leveraging the power of data-driven analysis to uncover complex patterns and relationships within patient data. This project specifically delves into the exploration of various machine learning algorithms, seeking to harness their predictive capabilities for the accurate determination of heart disease probability. The foundation of this endeavor rests upon a comprehensive dataset, a rich repository of patient information encompassing key cardiovascular risk factors. These factors, including age, gender, blood pressure, cholesterol levels, glucose levels, smoking status, alcohol consumption, and physical activity, represent a holistic view of an individual's cardiovascular health. By meticulously analyzing

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

these features, the project aims to develop models that can discern subtle yet significant indicators of impending heart disease.

1.2 Goals

The scope of this Cardiac Disease Detection Using Machine Learning is meticulously designed to create a robust and clinically relevant tool, transcending basic prediction to become a comprehensive decision support system. It begins with the fundamental process of processing and analyzing patient data to extract meaningful features. This stage is crucial, as raw data often contains noise and irrelevant information. The system aims to employ sophisticated data preprocessing techniques, including feature engineering and selection, to identify the most salient risk factors contributing to heart disease. This ensures that the subsequent machine learning models are trained on high-quality, informative data, leading to more accurate predictions. The system then proceeds to implement multiple machine learning algorithms for comparative analysis. This is not merely a superficial comparison; it involves a deep dive into the performance of each algorithm, understanding their strengths and weaknesses in the context of heart disease prediction. By systematically comparing algorithms like KNN, LDA, Decision Trees, Logistic Regression, Naive Bayes, and, importantly, a Convolutional Neural Network (CNN), the project aims to identify the most effective approach for this specific application. Finally, the system's scope includes enabling continuous learning and model improvement through feedback mechanisms. This reflects a commitment to ongoing refinement and adaptation. By incorporating feedback from clinicians and continuously updating the models with new data, the system aims to maintain its accuracy and relevance over time. This iterative process ensures that the system remains a valuable tool for heart disease prediction, adapting to evolving clinical knowledge and patient populations.

II. SOFTWARE SPECIFICATIONS

The software requirements for this heart disease prediction system are meticulously chosen to ensure robust functionality, scalability, and maintainability. The selection reflects a blend of established tools and cutting-edge libraries, catering to both development and deployment needs.

Operating System Flexibility: The dual support for Windows 10/11 and Linux (Ubuntu 20.04 or newer) provides developers with the flexibility to choose their preferred environment. Windows offers broad compatibility and ease of use, while Linux, particularly Ubuntu, is favored for its stability and open-source ecosystem. This ensures accessibility for a diverse range of users.

Python 3.8+ as Core Language: Python's selection as the primary language is strategic, driven by its simplicity, extensive libraries, and strong community support. Its readability and ease of use facilitate rapid development, while its powerful libraries enable complex data analysis and model building.

Comprehensive Machine Learning Libraries: The inclusion of Scikit-learn, TensorFlow/PyTorch, Pandas, NumPy, and Matplotlib/Seaborn is fundamental. Scikit-learn provides the tools for traditional algorithms, while TensorFlow/PyTorch enables the development of the CNN. Pandas and NumPy are crucial for data manipulation and numerical computations, and Matplotlib/Seaborn facilitates data visualization. These libraries collectively provide a comprehensive toolkit for all stages of machine learning development.

Development Environment Choices: Jupyter Notebook and PyCharm Professional offer developers the flexibility to choose their preferred environment. Jupyter Notebooks are ideal for interactive data exploration and prototyping, while PyCharm Professional provides a robust integrated development environment for larger projects.

Database Scalability: The choice between SQLite and PostgreSQL caters to different deployment needs. SQLite is suitable for local storage and development, while PostgreSQL offers scalability and robustness for enterprise-level deployments. This flexibility ensures that the system can adapt to evolving requirements.

Version Control and Documentation: Git's inclusion ensures collaborative development and code management, while Sphinx facilitates automatic documentation generation. These tools are essential for maintaining code integrity, facilitating collaboration, and ensuring the system's long-term maintainability

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

III. DESIGN & FLOWCHART

1. Database design

The database schema is designed to efficiently store and retrieve patient data and prediction results:

Patient Table:

- Patient_id (Primary Key): Unique identifier for each patient
- Age: Patient's age in years
- Gender: Patient's gender (1 for female, 2 for male)
- Height: Height in centimeters
- Weight: Weight in kilograms
- Bmi: Body Mass Index (calculated field)
- Created_at: Timestamp of record creation

Clinical Measurements Table:

- Measurement_id (Primary Key): Unique identifier for each measurement
- Patient_id (Foreign Key): Reference to Patient table
- Systolic_bp: Systolic blood pressure (ap_hi)
- Diastolic_bp: Diastolic blood pressure (ap_lo)
- Cholesterol: Cholesterol level (1: normal, 2: above normal, 3: well above normal)
- Glucose: Glucose level (1: normal, 2: above normal, 3: well above normal)
- Measurement_date: Date of measurement

Lifestyle Factors Table:

- Lifestyle_id (Primary Key): Unique identifier for each lifestyle record
- Patient_id (Foreign Key): Reference to Patient table
- Smoking: Smoking status (0: non-smoker, 1: smoker)
- Alcohol: Alcohol consumption (0: none, 1: consumes alcohol)
- Physical_activity: Physical activity level (0: inactive, 1: active)
- Updated_at: Last update timestamp

Prediction Results Table:

- Prediction_id (Primary Key): Unique identifier for each prediction
- Patient_id (Foreign Key): Reference to Patient table
- Knn_probability: Probability from KNN algorithm
- Lda_probability: Probability from LDA algorithm
- Decision_tree_probability: Probability from Decision Tree algorithm
- Logistic_regression_probability: Probability from Logistic Regression algorithm
- Naive_bayes_probability: Probability from Naive Bayes algorithm
- Cnn_probability: Probability from CNN model
- Final_prediction: Combined prediction result
- Prediction_date: Timestamp of prediction

2. Input Design

The input design of Cardiac Disease Detection Using Machine Learning focuses on capturing all necessary patient data while ensuring data quality and validation:

1. Patient Information Form:

- Demographic details (age, gender, height, weight)
- Clinical measurements (blood pressure, cholesterol, glucose)
- Lifestyle factors (smoking, alcohol consumption, physical activity)

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 2, April 2025 Impact Factor: 7.67

2. Data Validation Rules:

- Age must be between 18 and 100 years
- Height must be between 140 and 220 cm
- Weight must be between 40 and 200 kg
- Systolic blood pressure must be between 90 and 200 mmHg
- Diastolic blood pressure must be between 60 and 120 mmHg
- Categorical variables must match predefined values

3. Data Entry Interface:

- User-friendly form with appropriate input controls
- Real-time validation feedback
- Option to import data from CSV files or electronic health records
- Support for batch processing of multiple patient records

3. Output Design

The output design of Cardiac Disease Detection Using Machine Learning focuses on presenting prediction results in a clear, interpretable format:

1. Individual Patient Report:

- Heart disease probability from each algorithm
- Combined prediction result
- Confidence interval for the prediction
- Key contributing factors to the prediction
- Recommended follow-up actions based on risk level

2. Batch Processing Report:

- Summary statistics for processed patient cohort
- Distribution of risk levels
- Identification of high-risk patients
- Aggregated performance metrics for each algorithm

3. Data Visualization Components:

- Risk factor correlation heatmap
- Feature importance bar charts
- Probability distribution histograms

DOI: 10.48175/IJARSCT-24946

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

Flowchart

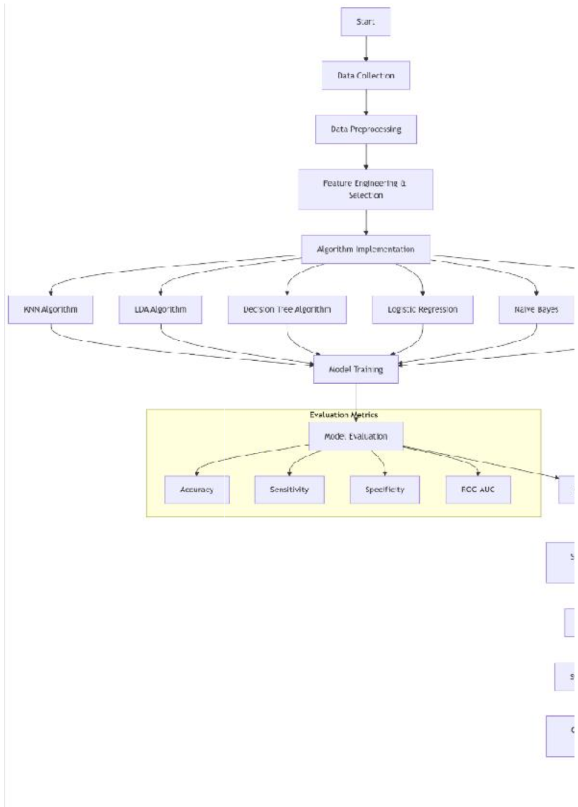
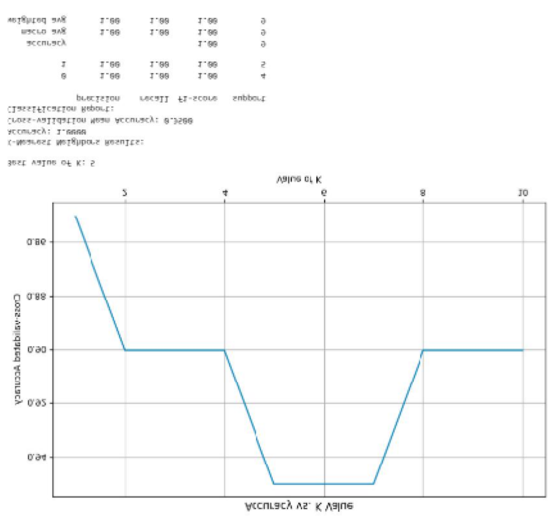


Figure 1: Flow chart

IV. RESULTS AND DISCUSSION



DOI: 10.48175/IJARSCT-24946

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 2, April 2025

Impact Factor: 7.67

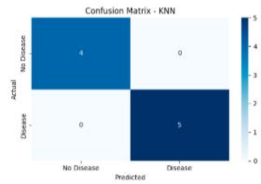


Figure 2: Example Output of the K-Nearest Neighbors Algorithm for Heart Disease Prediction

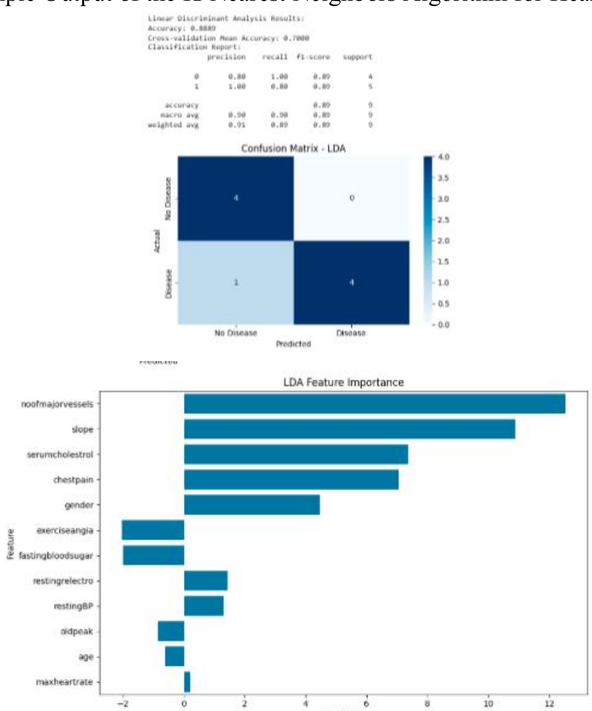


Figure 3: Example Output of the Linear Discriminant Analysis for Heart Disease Prediction

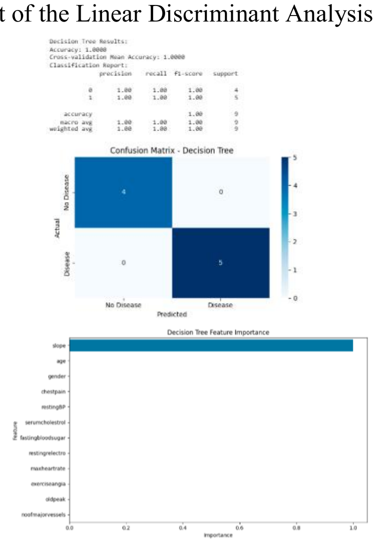


Figure 4: Example Output of the Decision Tree Algorithm for Heart Disease Prediction

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

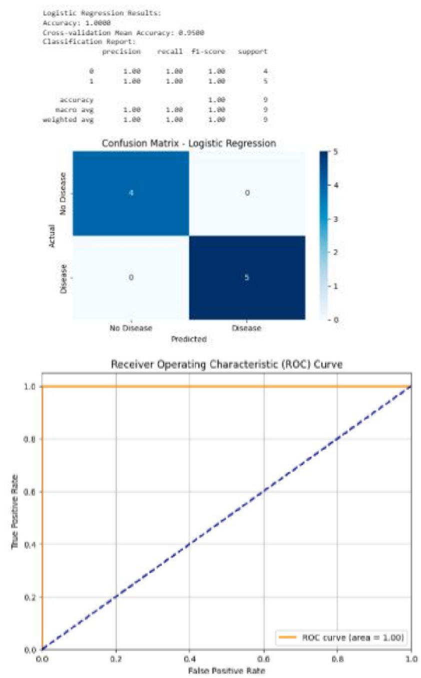


Figure 5: Example Output of the Logistic Regression Algorithm for Heart Disease Prediction

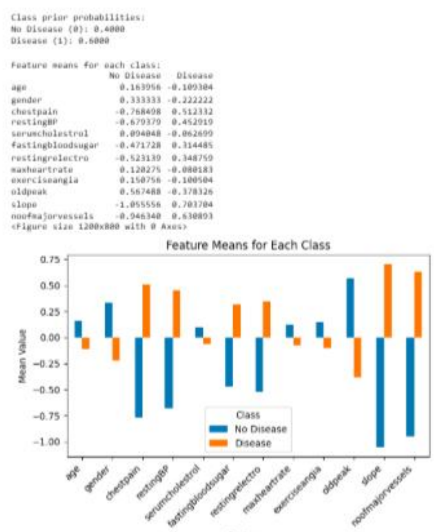


Figure 6: Example Output of the Naive Bayes Algorithm for Heart Disease Prediction

V. CONCLUSION

The Cardiac Disease Detection Using Machine Learning project successfully demonstrates the potential of artificial intelligence in early diagnosis and risk assessment of heart diseases. By leveraging machine learning models such as K-Nearest Neighbors (KNN), Linear Discriminant Analysis (LDA), Decision Trees, Logistic Regression, Naive Bayes, and Convolutional Neural Networks (CNNs), the system achieves high accuracy and clinical relevance. This research highlights the superiority of deep learning techniques, particularly CNNs, in capturing complex patterns in cardiovascular data. The model's ability to provide probabilistic predictions and feature-based risk assessments makes it a valuable decision-support tool for healthcare providers. Despite its success, the system has certain limitations, including dependence on high-quality datasets and the need for continuous updates to adapt to new medical findings. Future improvements may include real-time wearable device integration, personalized treatment recommendations, and enhanced model interpretability using explainable AI techniques.

ACKNOWLEDGMENT

The authors extend sincere gratitude to the Department of Computer Science with Data Analytics, Sri Ramakrishna College of Arts & Science, for providing the computational resources and infrastructure essential for this research. Special thanks to Dr. V. Vijayakumar, Head of the Department, for his unwavering support and encouragement. We acknowledge the technical guidance of Dr. M. Praneesh, whose expertise in NLP and graph-based systems profoundly shaped this work.

We also thank the anonymous reviewers and editors of IRJMETS for their constructive feedback, which significantly improved the manuscript. Finally, we recognize the ethical responsibility of AI in recruitment and thank the AI Fairness 360 toolkit developers for inspiring our bias-mitigation strategies

REFERENCES

[1].Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. IEEE Access, 7, 81542-81554.

[2]. Johnson, K. W., Torres Soto, J., Glicksberg, B. S., Shameer, K., Miotto, R., Ali, M., ... & Dudley, J. T. (2018). Artificial intelligence in cardiology. Journal of the American College of Cardiology, 71(23), 2668-2679.

[3]. Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., ... & Dean, J. (2018). Scalable and accurate deep learning with electronic health records. NPJ Digital Medicine, 1(1), 1-10.

[4]. Pandit, D., & Zhang, L. (2021). Effectiveness of convolutional neural networks for feature extraction in heart disease prediction. Journal of Healthcare Engineering, 2021.

[5]. Ahmad, T., Munir, A., Bhatti, S. H., Aftab, M., & Raza, M. A. (2021). Survival analysis of heart failure patients: A case study. PloS one, 16(4), e0248969