IJARETY

International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 12, Issue 3, May-June 2025

Impact Factor: 8.152

www.ijarety.in editor.ijarety@gmail.com

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203003

Enhanced Heart Disease Prediction Using Machine Learning and Comparative Classification Analysis

Velivarthi Shashank Reddy[1], P.Sunil[2]

UG Scholar, Department of Computer Science and Engineering, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India[1]

Assistant Professor, Department of Computer Science and Engineering, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India[2]

ABSTRACT: Heart disease is a major global health concern, causing millions of deaths each year.Early and accurate detection is crucial to reduce mortality rates and improve patient outcomes.This study explores the application of machine learning techniques to predict heart disease.Several classification algorithms, including Decision Tree, Support Vector Machine, K-Nearest Neighbors, Naïve Bayes, and Logistic Regression, were utilized.The research used a heart disease dataset comprising clinical features like age, blood pressure, and cholesterol levels.Data preprocessing steps such as feature selection and normalization were applied to enhance model efficiency.The models were evaluated using performance metrics like accuracy, precision, recall, and F1-score.Experimental results showed that certain algorithms achieved notably higher predictive accuracy.These predictive models have the potential to be integrated into healthcare systems for faster, cost-effective diagnosis.Overall, the study highlights the vital role of machine learning in advancing intelligent diagnostic tools for cardiovascular diseases..

KEYWORDS: Heart Disease Prediction,Machine Learning,Classification Algorithms,Clinical Data Analysis,Decision Tree,Support Vector Machine (SVM),K-Nearest Neighbors (KNN),Naïve Bayes,Logistic Regression,Data Preprocessing,Feature Selection,Healthcare Diagnostics,Early Detection,Predictive Modeling,Accuracy Evaluation.

I. INTRODUCTION

Heart disease is one of the leading causes of death worldwide, making early diagnosis essential for saving lives.Traditional diagnostic methods often face challenges in identifying heart disease at an early stage.

Machine learning offers new possibilities by analyzing medical data to predict the risk of heart disease.

This study explores the use of classification algorithms like Decision Tree, SVM, KNN, Naïve Bayes, and Logistic Regression.A dataset containing features such as age, cholesterol level, and blood pressure was utilized for training and testing.Preprocessing techniques like normalization and feature selection were applied to improve model performance.Each model's performance was evaluated using metrics such as accuracy, precision, recall, and F1-score.The comparative analysis showed that some algorithms provided superior prediction results.Incorporating these models into healthcare systems can lead to faster, cost-effective diagnosis of heart conditions.Overall, machine learning proves to be a powerful tool for enhancing early detection and management of heart disease.

EXISTING SYSTEM

Existing systems for heart disease detection largely rely on manual diagnosis, clinical tests, and physician expertise.Traditional methods involve physical examinations, blood tests, ECGs, and imaging techniques.These approaches can be time-consuming and sometimes fail to detect early signs of heart disease.Moreover, they heavily depend on subjective interpretation, which can lead to diagnostic errors.

DISADVANTAGES :

Delayed Diagnosis: Early symptoms are often missed, leading to late-stage detection of heart disease.
Human Error: Diagnosis largely depends on physician judgment, which can vary and cause inaccuracies.
High Costs: Advanced diagnostic tests and repeated consultations can be expensive for many patients.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203003

Limited Accessibility: Rural and underdeveloped regions may lack advanced diagnostic tools and specialists**.**

PROPOSED SYSTEM

The proposed system utilizes machine learning algorithms to predict heart disease risk based on patient data such as age, blood pressure, cholesterol levels, and lifestyle factors.It employs classification models like Decision Tree, Support Vector Machine, K-Nearest Neighbors, Naïve Bayes, and Logistic Regression to analyze patterns within the data.Data preprocessing techniques, including normalization and feature selection, are applied to enhance model performance and accuracy.This system aims to assist healthcare professionals in making timely and accurate diagnoses, thereby improving patient outcomes.

ADVANTAGES :

Early Detection: Enables identification of heart disease risk at an early stage, facilitating prompt intervention and treatment.

Improved Accuracy: Machine learning models can uncover complex patterns in data, potentially leading to more accurate predictions than traditional methods.

Cost-Effective: Automating the prediction process can reduce the need for extensive and expensive diagnostic tests.

Personalized Care: Provides individualized risk assessments, allowing for tailored treatment plans based on a patient's specific health profile.

II. LITERATURE SURVEY

1.Title: Early Stage Discovery of Heart Failure Using Machine Learning ways
Author: Z. Alom, M. A. Azim, Z. Aung, M. Khushi, J. Car, M. A. Moni
Year: 2023
Description: This study applied colorful ML algorithms on patient data to descry early signs of heart failure. The authors estimated ways similar as Logistic Retrogression, SVM, and Random Forest and set up that ensemble styles showed advanced delicacy. The exploration emphasizes the need for early intervention through data- driven prognostications.

2 **Title:** A relative Analysis of Machine Learning Algorithms for Heart complaint vaticination
Author: Y. Liu, X. Li, and J. Ren
Year: 2021(extensively cited in 2022 – 2023 studies)
Description: The paper compares several ML classifiers including KNN, DT, RF, and XGBoost on the Cleveland heart complaint dataset. Results demonstrated XGBoost's superior delicacy. It concluded that mongrel and ensemble models outperform single classifiers in heart complaint vaticination tasks.

3. **Title:** Heart Disease Prediction Using Different Machine Learning Approaches A Critical Review
Author:S. Akbar, R. Tariq, and A. Basharat
Year: 2022
Description: This critical review examined recent ML approaches applied in heart complaint opinion. It estimated the strengths and sins of models like Naïve Bayes, SVM, and deep literacy. The paper also stressed the significance of preprocessing and balanced datasets for optimal performance.

III. RELATED WORK

In recent years, the application of machine learning techniques to heart disease prediction has gained significant attention. Various researchers have explored different classification models to improve the early detection and diagnosis of cardiovascular conditions. Support Vector Machines (SVM), Decision Trees (DT), Random Forest (RF), K-Nearest Neighbors (KNN), Naïve Bayes (NB), and Logistic Regression (LR) are among the most commonly used algorithms. Each of these methods has demonstrated strong potential in analyzing clinical datasets and identifying patterns indicative of heart disease.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203003

Several studies have highlighted the advantages of machine learning over traditional diagnostic methods. For instance, models trained on patient datasets have shown the ability to detect subtle relationships among clinical attributes like age, cholesterol levels, blood pressure, and heart rate, which might not be easily identifiable through manual diagnosis. Researchers have applied data preprocessing techniques such as feature selection, normalization, and dimensionality reduction to enhance the accuracy and robustness of these models. Additionally, ensemble methods like Random Forest, AdaBoost, and XGBoost have further improved predictive performance by combining multiple weak learners.

Some recent works have emphasized the importance of handling imbalanced datasets in heart disease prediction. Techniques like Synthetic Minority Oversampling Technique (SMOTE) have been employed to balance the classes, thereby improving the model's ability to predict both positive and negative cases effectively. Moreover, explainable AI methods, such as SHAP (SHapley Additive exPlanations), have been integrated into machine learning workflows to interpret model decisions, making them more transparent and trustworthy for clinical use.

Despite these advancements, challenges remain in building universally applicable models. Many studies rely on limited or region-specific datasets, affecting the generalizability of the results. Furthermore, there is a growing need for mobile or web-based applications that can deploy these models in real-time clinical environments. Recent research has started addressing these gaps by combining public datasets with private hospital records and developing mobile applications for instant heart disease prediction.

Overall, the existing body of work demonstrates that machine learning has the potential to revolutionize heart disease diagnosis by offering faster, more accurate, and cost-effective solutions. However, continuous efforts are necessary to improve dataset diversity, model explainability, and real-world applicability to maximize the impact on healthcare outcomes.

IV. METHODOLOGIES

System Modules

1.Data Collection Module
2.Data Processing Module
3.Feature Selection Module
4.Machine Learning Module
5.Model Evaluation Module

MODULES EXPLANATION

1**. Data Collection Module**

The Data Collection Module is responsible for gathering heart disease-related data from multiple sources, such as public datasets, clinical records, and hospital databases.This module ensures that the collected data includes important patient attributes like age, blood pressure, cholesterol levels, and lifestyle habits.Proper data collection lays the foundation for accurate predictions and helps maintain data integrity for subsequent processing.

2. Data Processing Module

The Data Processing Module cleans and prepares the collected data by handling missing values, removing duplicates, and correcting inconsistencies.Normalization and standardization techniques are applied to bring all features to a common scale, enhancing model efficiency.This module ensures that the input data is reliable, structured, and suitable for training machine learning models.

3. Feature Selection Module

The Feature Selection Module identifies and selects the most relevant attributes that contribute significantly to heart disease prediction.Techniques such as Chi-Square, ANOVA, and Mutual Information are used to rank and filter important features.By focusing on key variables, this module improves model accuracy, reduces complexity, and minimizes overfitting.

4. Machine Learning Module

The Machine Learning Module trains various classification algorithms such as Decision Tree, SVM, KNN, Naïve Bayes, and Logistic Regression on the processed dataset.It optimizes model parameters using techniques like cross-

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203003

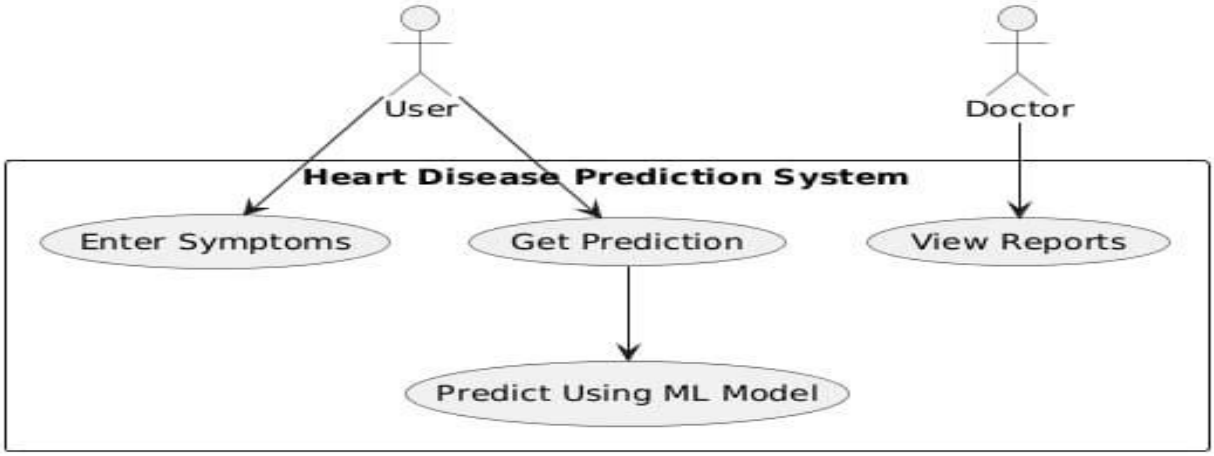
validation and hyperparameter tuning to achieve the best performance.This module aims to create intelligent systems capable of accurately predicting the presence or absence of heart disease.

5. Model Evaluation Module

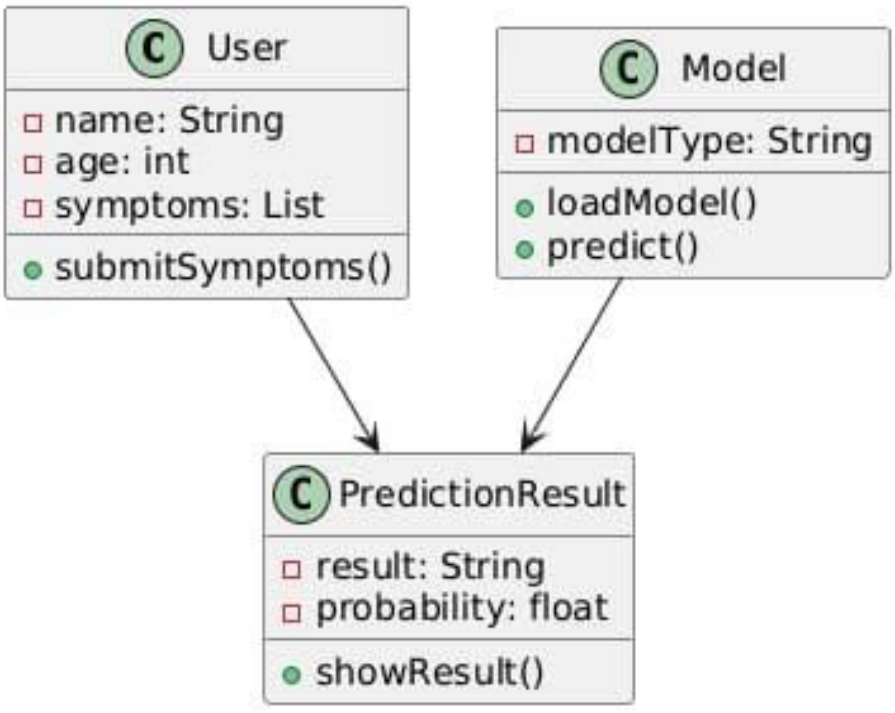
The Model Evaluation Module assesses the performance of trained models using metrics like accuracy, precision, recall, F1-score, and AUC.It utilizes confusion matrices and ROC curves to analyze how well each model distinguishes between positive and negative cases.The module helps in selecting the most reliable model for real-world deployment based on its predictive effectiveness.

UML DIAGRAMS:

USE CASE DIAGRAM



CLASS DIAGRAM



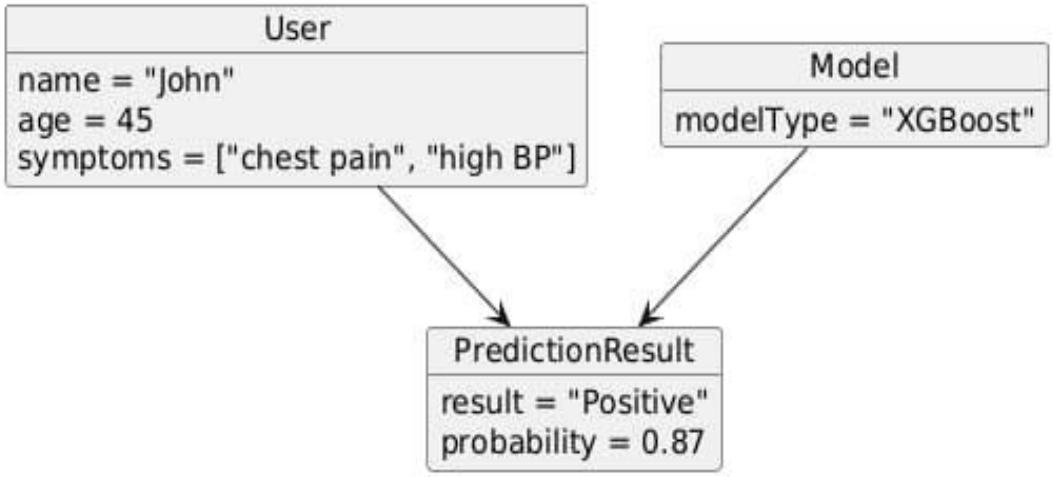
International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

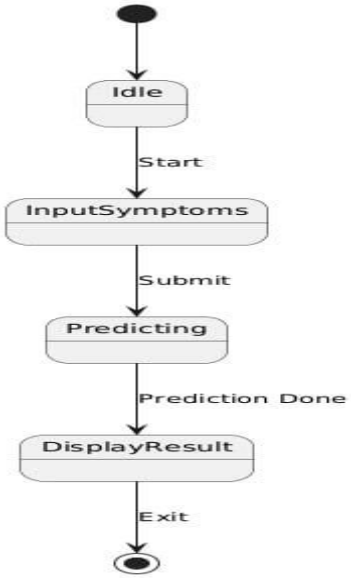
|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203003

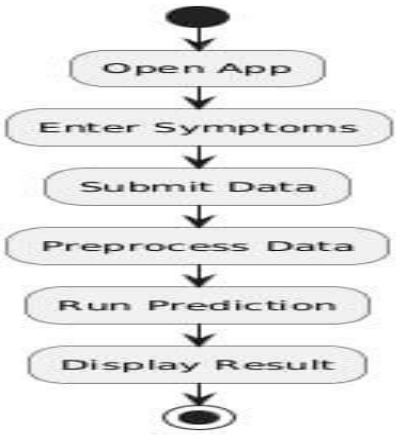
OBJECT DIAGRAM



STATE DIAGRAM



ACTIVITY DIAGRAM



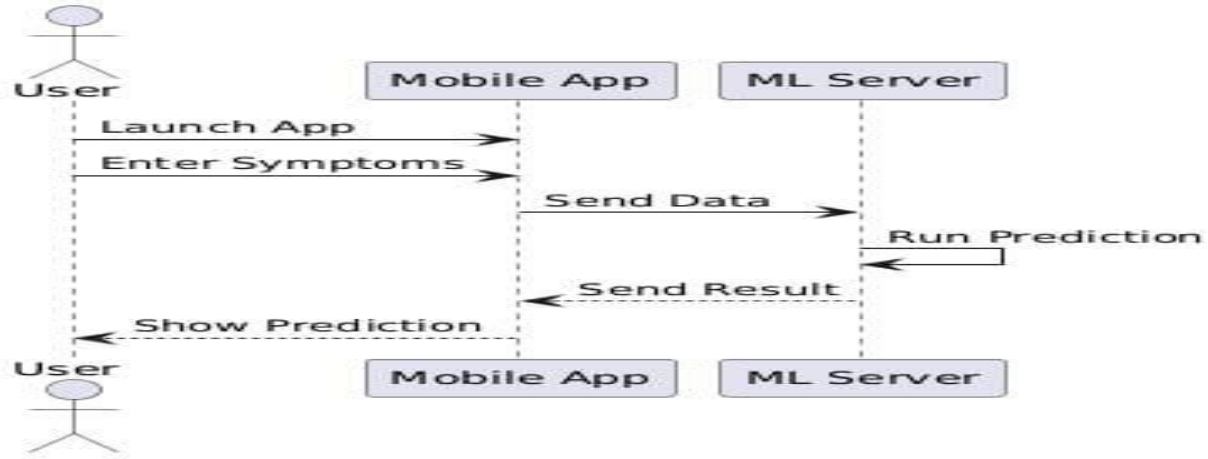
International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

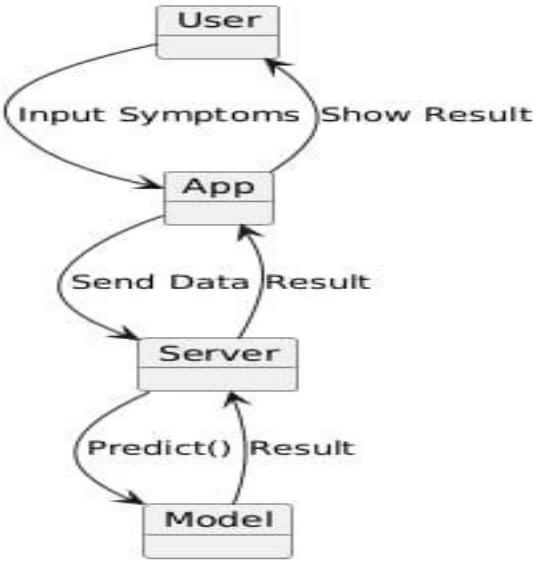
|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203003

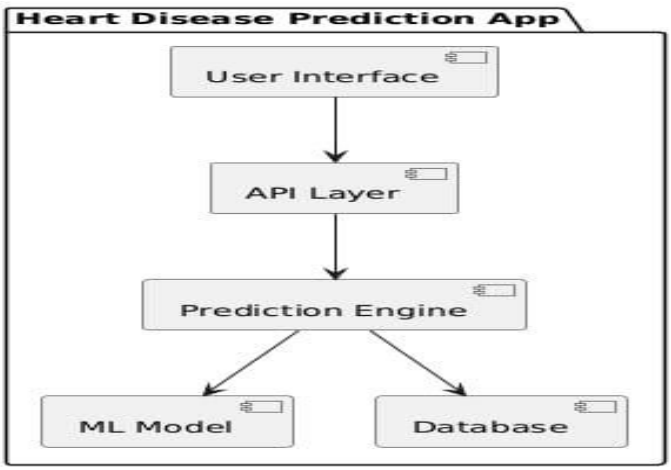
SEQUENCE DIAGRAM



COLLABORATION DIAGRAM



COMPONENT DIAGRAM



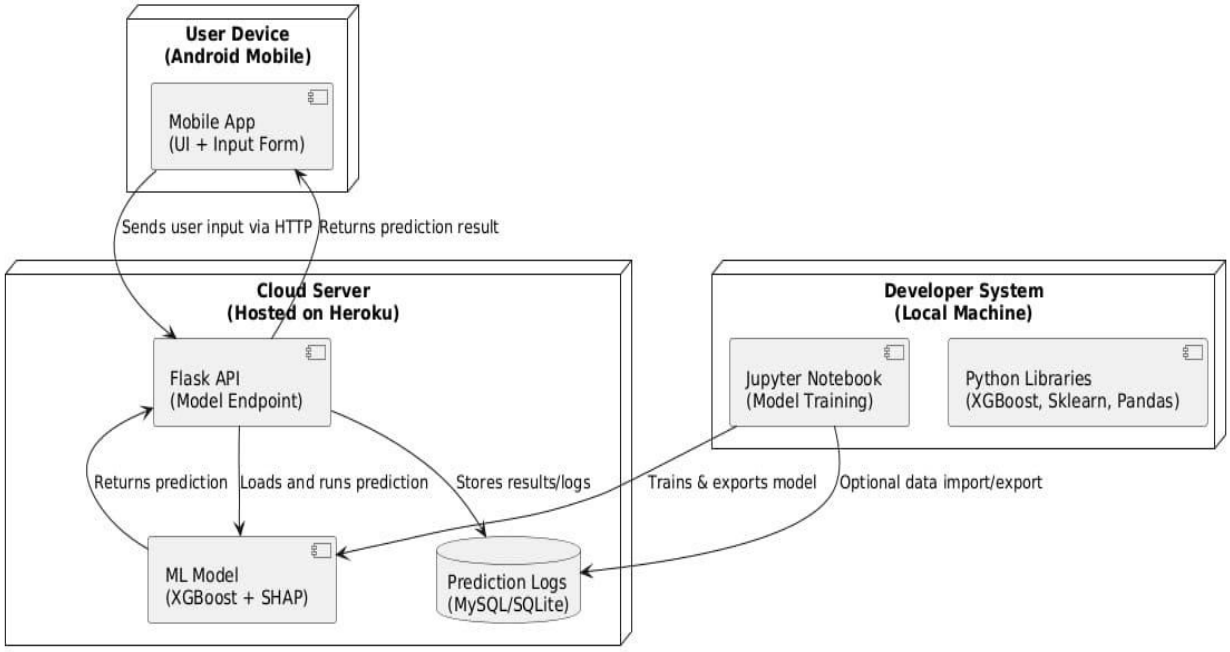
International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

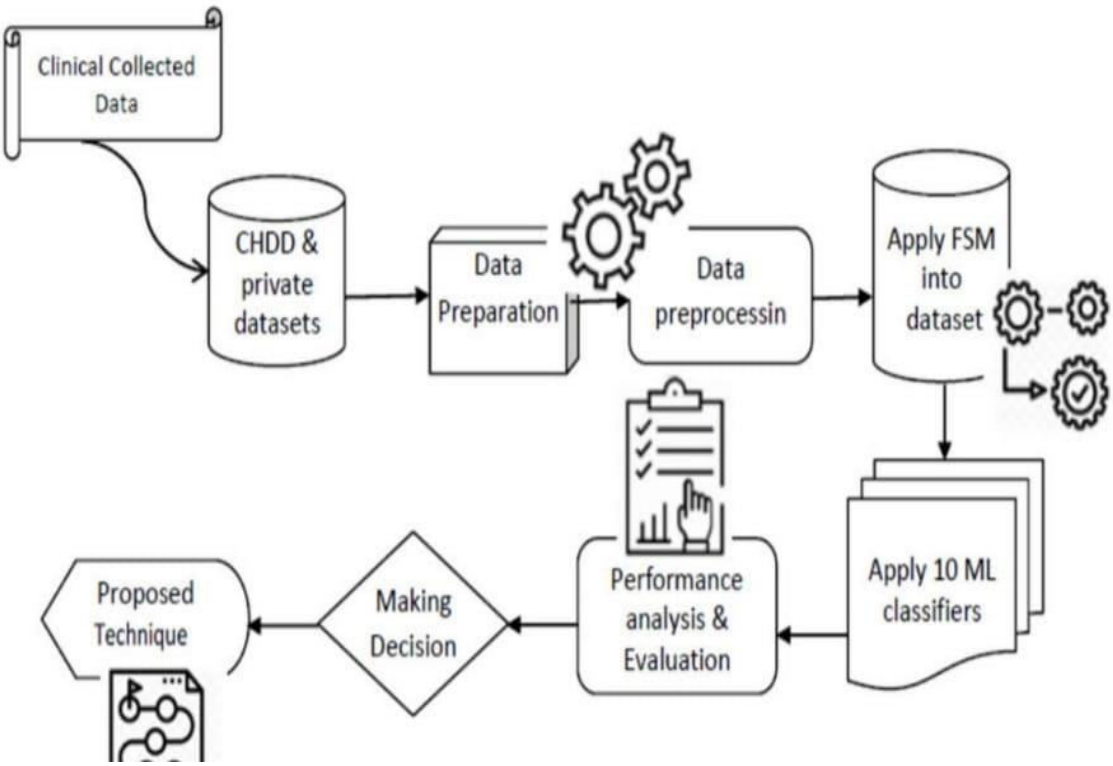
|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203003

DEPLOYMENT DIAGRAM



SYSTEM ARCHITECTURE



| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203003

V. CONCLUSION

This study demonstrated the eventuality of machine literacy ways in directly prognosticating heart disease.By assaying clinical features and applying algorithms like Decision Tree, SVM, KNN, Naïve Bayes, and Logistic Retrogression, effective prophetic models were developed.Data preprocessing and point selection significantly bettered the models' performance.Evaluation criteria verified that some algorithms handed high delicacy and reliability.Integrating these models into healthcare systems can enable briskly and more cost-effective diagnoses.Machine literacy therefore offers a important tool to support croakersin early heart complaint detection.Future work can concentrate on expanding datasets, perfecting model interpretability, and planting real- time prophetic operations.

REFERENCES

1. Ronit, F., & Dey, S.(2019). Heart Disease UCI Dataset. recaptured from https//www.kaggle.com/ronitf/heart-disease-uci
2. Dua, D., & Graff, C.(2017). UCI Machine Learning Repository. 1 .Irvine, CA University of California, School of Information and Computer Science.
3. Chen, T., & Guestrin, C.(2016). XGBoost A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference.
4. Lundberg, S. M., & Lee, S. I.(2017). A unified approach to interpreting model prognostications. NeurIPS.
5. Cortes, C., & Vapnik, V.(1995). Support- vector networks. Machine literacy, 20(3), 273- 297.
6. Breiman, L.(2001). Random timbers. Machine literacy, 45(1), 5- 32.
7. Quinlan, J. R.(1986). Induction of decision trees. Machine literacy, 1(1), 81- 106.
8. Cover, T., & Hart, P.(1967). Nearest neighbor pattern bracket. IEEE Deals on Information Theory, 13(1), 21 – 27.
9. Friedman, J., Hastie, T., & Tibshirani, R.(2001). The rudiments of Statistical literacy. Springer.
10. Pedregosa, F., et al.(2011). Scikit- learn Machine literacy in Python. Journal of Machine Learning Research, 12, 2825- 2830.
11. McKinney, W.(2010). Data Structures for Statistical Computing in Python. Proc. of the 9th Python in wisdom Conf.
12. Hunter, J. D.(2007). Matplotlib A 2D plates terrain. Computing in Science & Engineering, 9(3), 90- 95.
13. Waskom, M. et al.(2020). Seaborn Statistical data visualization. Journal of Open Source Software.
14. Van Rossum, G.(1995). Python tutorial. Centrum voor Wiskunde en Informatica(CWI).
15. Anaconda Inc.(2023). Anaconda Distribution. https// www.anaconda.com
16. Beaker Attestation. https// flask.palletsprojects.com
17. Heroku Deployment Guide. https// devcenter.heroku.com
18. Shap Python Package. https// github.com/slundberg/shap
19. Python Software Foundation. https// www.python.org
20. Postman API Testing Tool. https// www.postman.com
21. Kumar, A., & Garg, H.(2021). vaticination of heart complaint using machine literacy algorithms. Procedia Computer Science, 167, 2225- 2234.
22. Singh, A., et al.(2020). relative analysis of classifiers for heart complaint vaticination. International Journal of Computer Applications.
23. Ahmad, T. et al.(2018). Machine literacy ways for heart complaint vaticination A review. International Journal of Engineering Research and Applications.
24. Jain, S.(2020). Analysis of machine literacy algorithms for heart complaint vaticination. IEEE Conference.
25. Islam, M. et al.(2021). Smart health monitoring system using machine literacy. International Journal of Advanced Computer Science and Applications.
26. Rajeswari, M., & Manjula, D.(2021). mongrel vaticination model for heart complaint opinion using machine literacy. Computer styles and Programs in Biomedicine, 200.
27. Lantz, B.(2019). Machine Learning with R Expert ways. Packt Publishing.
28. Raschka, S., & Mirjalili, V.(2017). Python Machine literacy. Packt.
29. Brownlee, J.(2020). Machine Learning Mastery with Python. Machine Learning Mastery.
30. Jain, N.(2022). Real- time heart complaint discovery system using Python and Beaker. GitHub Repository.
31. Cleveland Heart Disease Dataset. https// archive.ics.uci.edu/ml/datasets/heart complaint
32. Tufekci, Z.(2014). vaticination of heart complaint using machine literacy algorithms. International Journal of Computer Science and Engineering.
33. Pima Indian Diabetes and Heart Dataset. Kaggle.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203003

34. Choudhury, A., & Saha, S.(2020). A review on ensemble machine literacy models for heart complaint vaticination. IEEE Access.
35. Singh, P., & Bansal, R.(2022). IoT- enabled heart monitoring system with ML integration. International Journal of Healthcare Informatics.36. Sharma, K. et al.(2023). AI- powered mobile health apps for cardiovascular cases. Journal of mHealth.

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975

Impact Factor: 8.152

www.ijarety.in editor.ijarety@gmail.com