International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 2, Mar-Apr 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Cardio Guardian Pro - Heart Disease Prediction System

Satish Ramaswamy Valluvar[1], Dr. Manoj Singh[2]
[1](Computer Science, SIES College of Arts, Science and Commerce, Sion (West)
Email: satishvalluvar11@gmail.com)
[2](Head of Computer Science Department, SIES College of Arts, Science and Commerce, Sion (West)
Email: manojks@sies.edu.in)

----------**********----------

Abstract:

Cardiovascular diseases (CVDs) are the leading cause of mortality worldwide, accounting for approximately 17.9 million deaths annually. Early prediction of heart disease is critical for effective prevention and treatment. This research presents Cardio Guardian Pro, a machine learning-based system designed to predict heart disease risk using multiple classification algorithms. The system evaluates models such as Logistic Regression, Random Forest, Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naive Bayes to identify the most accurate predictor.

A comprehensive dataset consisting of clinical and demographic attributes is pre-processed and analysed. The models are evaluated using performance metrics including Accuracy, Precision, Recall, F1-Score, and AUC-ROC. Experimental results indicate that the Support Vector Classifier achieves the highest accuracy of 98.05%, while Random Forest demonstrates superior AUC-ROC performance.

The system is deployed as a web-based application using Flask, enabling users to input health data and receive real-time predictions. This research contributes to predictive healthcare by providing a reliable, interpretable, and accessible decision-support tool for early heart disease detection.

Keywords — **Heart Disease Prediction, Machine Learning, SVM, Random Forest, Healthcare Analytics, Predictive Modelling.**

----------**********----------

INTRODUCTION

Cardiovascular diseases represent a major global health challenge, affecting millions of individuals each year. Conditions such as coronary artery disease, heart failure, and arrhythmias significantly contribute to mortality rates. Early detection plays a crucial role in reducing complications and improving patient outcomes.

Traditional diagnostic approaches rely heavily on clinical expertise and laboratory testing, which may delay early intervention. With the rise of **machine learning**, predictive analytics has emerged as a powerful tool in healthcare, enabling early risk assessment using patient data.

This research aims to develop a **robust and accurate heart disease prediction system** by comparing multiple machines learning algorithms and deploying the best-performing model in a real-world application.

document as a template and simply type your text into it.

Problem Statement

Despite advancements in medical science, early detection of heart disease remains a challenge due to:

- Lack of predictive tools for early diagnosis
- Dependence on manual clinical analysis
- High cost and time-consuming diagnostic procedures
- Limited accessibility in rural and under-resourced areas

This project addresses these challenges by developing a **machine learning-based prediction system** capable of analysing patient data and providing early risk assessment

Objectives

1. To collect and pre-process a comprehensive heart disease dataset
2. To evaluate multiple machine learning algorithms for prediction accuracy
3. To identify the best-performing model
4. To develop a user-friendly web-based prediction system
5. To assist healthcare professionals in decision-making

Literature Review

Previous research in heart disease prediction has explored various machine learning techniques:

- Logistic Regression provides interpretability and probabilistic outputs
- Decision Trees offer transparent decision-making structures
- Random Forest improves accuracy through ensemble learning
- Support Vector Machines handle high-dimensional data effectively
- Neural Networks capture complex patterns in large datasets

However, many existing systems lack **comparative analysis and real-world deployment**, which this study addresses.

Related Work

Heart disease prediction has been widely studied using statistical analysis and machine learning techniques. Researchers have applied various classification algorithms such as Decision Trees, Random Forests, Logistic Regression, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN) to identify patterns in patient data and predict the likelihood of cardiovascular diseases.

Among these, ensemble methods like Random Forest have demonstrated high accuracy due to their ability to reduce overfitting and handle complex datasets. Similarly, Support Vector Machines are effective in high-dimensional spaces and are widely used for medical diagnosis problems.

Traditional statistical approaches, although interpretable, often fail to capture complex nonlinear relationships among risk factors. On the other hand, advanced machine learning models provide improved predictive performance but require large datasets and computational resources.

Recent research also focuses on integrating prediction systems with web-based or clinical decision support systems to assist healthcare professionals. However, many existing systems lack real-time usability, user-friendly interfaces, or comparative evaluation of multiple algorithms.

This work addresses these gaps by performing a **comparative analysis of multiple machines learning algorithms** and deploying the best-performing model in a **user-friendly web-based application** for real-time heart disease prediction.

Methodology

A. System Overview

The proposed system follows a structured pipeline for heart disease prediction:

1. Data Collection
2. Data Pre-processing
3. Feature Selection
4. Model Training
5. Model Evaluation

6. Prediction Generation
7. Web Application Deployment

B. Data Collection

The dataset used in this study is obtained from **Kaggle**, containing medical and demographic information of patients.

Key Attributes Include:

- Age
- Sex
- Chest Pain Type
- Resting Blood Pressure
- Serum Cholesterol
- Fasting Blood Sugar
- ECG Results
- Maximum Heart Rate
- Exercise-Induced Angina
- Old Peak
- Slope
- Number of Major Vessels
- Thalassemia
- Target Variable (0 = No Disease, 1 = Disease)

C. Data Pre-processing

Data pre-processing is a crucial step to ensure data quality and model accuracy.

Steps Performed:

- Handling missing values
- Encoding categorical variables
- Feature scaling using normalization/standardization
- Removing outliers
- Splitting dataset into training and testing sets

D. Performance Evaluation

Multiple machine learning algorithms are implemented and compared:

- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)
- Gaussian Naive Bayes

Among these, **Support Vector Classifier (SVC)** is selected as the final model due to its superior performance.

E. Career / Stream Suggestion

- The performance of each model is evaluated using the following metrics:
- Accuracy
- Precision
- Recall (Sensitivity)
- F1-Score
- AUC-ROC Curve
- Cross-validation score
- These metrics ensure a comprehensive evaluation of model performance.

F. Report Generation

The system generates a structured output including:

- Input patient details
- Predicted risk level
- Probability score
- Health interpretation

G. System Implementation

The system is implemented using:

- **Programming Language:** Python
- **Libraries:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn
- **Framework:** Flask
- **Database:** SQLite

Key Features:

- User-friendly web interface
- Real-time prediction
- Secure data handling
- Responsive design

System Architecture

The architecture of the proposed system consists of:

- **Input Layer:** User health data input
- **Data Processing Module:** Pre-processing and feature engineering
- **Machine Learning Engine:** Model training and prediction
- **Evaluation Module:** Performance analysis

- **Application Layer:** Flask-based web interface
- **Output Layer:** Prediction results display

Experimental Results

The system was tested on student datasets with multiple subjects.

Key Observations:
 SVM achieved the highest accuracy (98.05%)
 Random Forest achieved the best AUC-ROC score
 Decision Tree showed strong generalization ability
 Ensemble methods performed better than single

The models were tested using a heart disease dataset and evaluated based on multiple performance metrics.

Results Summary:

Model	Accuracy	Cross Validation	AUC-ROC
Logistic Regression	86.34%	84.02%	0.9391
Naive Bayes	85.37%	81.83%	0.9311
Random Forest	94.63%	90.24%	0.9924
KNN	87.80%	84.02%	0.9468
Decision Tree	94.63%	93.29%	0.9917
SVM (Best Model)	**98.05%**	92.80%	0.9331

A. Evaluation Discussion
Advantages:
- High prediction accuracy
- Multiple algorithm comparison
- Real-time prediction capability
- User-friendly interface

Limitations:
- Dependent on dataset quality
- Requires periodic model retraining
- Limited real-time clinical validation

Conclusion

This paper presents **Cardio Guardian Pro**, a heart disease prediction system using machine learning techniques. The study compares multiple algorithms and identifies the **Support Vector Classifier as the best-performing model**.
The integration of the model into a web-based application enhances accessibility and usability, making it a valuable tool for early diagnosis and preventive healthcare.
The system demonstrates strong potential to assist healthcare professionals in decision-making and improve patient outcomes.

Future Work

Future enhancements of the system may include:
- Integration of deep learning models
- Mobile application development
- Real-time monitoring using wearable devices
- Cloud-based deployment
- Integration with hospital management systems
- Personalized health recommendations

References

[1] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann, 2012.
[2] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," IEEE, 2010.
[3] Kaggle, "Heart Disease Dataset."
[4] Scikit-learn Documentation, https://scikit-learn.org
[5] Flask Documentation, https://flask.palletsprojects.com

