IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Measuring The Heart Attack Possibility Using Different Typing Of Machine Learning Algorithms

[1]K.Arunpandi,[2]V.Karthik
[1]Postgraduate Student (MCA), [2]Asst.Professor
Department of Computer Applications
Dr. M.G.R. Educational and Research Institute, Chennai, India

***Abstract*:** Heart disease remains one of the leading causes of mortality globally, necessitating the development of early and accurate diagnostic tools. This project focuses on predicting the likelihood of heart attacks using various machine learning (ML) algorithms. A publicly available clinical dataset, including features such as age, gender, chest pain type, blood pressure, cholesterol, and ECG results, is used for training and evaluation.
The dataset undergoes preprocessing steps including data cleaning, normalization, and feature encoding. Supervised learning algorithms including Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Decision Tree, and Random Forest are implemented and compared based on performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. The Random Forest algorithm outperformed others in terms of accuracy and generalization ability.
The system is integrated into an Android application using Firebase as a backend service, enabling real-time user interaction and prediction delivery. The study demonstrates that ensemble learning methods offer robust and interpretable solutions for heart disease prediction, which can support clinical decision-making and preventive care. Future enhancements may include integration with wearable devices and deployment in real-time hospital environments.

Keywords: Heart Disease Prediction, Machine Learning, Random Forest, Support Vector Machine, K-Nearest Neighbors, Decision Tree, Clinical Data, Android Application, Firebase Integration, Healthcare Analytics

I. INTRODUCTION:

Cardiovascular diseases (CVDs), especially heart attacks, remain the foremost cause of death worldwide, accounting for approximately 17.9 million deaths annually according to the World Health Organization. Despite advancements in diagnostic technologies, the early detection and prevention of heart attacks still pose significant challenges due to the multifactorial nature of the disease, patient-specific variability, and the often asymptomatic progression of the condition. Hence, there is a growing need for intelligent, accurate, and real-time prediction systems that can support medical practitioners in making informed clinical decisions.
In recent years, the evolution of machine learning (ML) and artificial intelligence (AI) has transformed various sectors, including healthcare. ML algorithms have shown significant potential in analyzing complex medical data to uncover patterns and correlations that are often difficult to detect using traditional statistical methods. By leveraging clinical parameters such as age, gender, blood pressure, cholesterol levels, chest pain types, and ECG results, machine learning can be used to develop predictive models that assess an individual's risk of experiencing a heart attack.
This project aims to design and implement a heart attack prediction system that employs a comparative analysis of several supervised machine learning algorithms, including Support Vector Machine (SVM),

Decision Tree, k-Nearest Neighbors (KNN), and Random Forest. These models are trained and validated using a standard dataset (e.g., the UCI Heart Disease dataset), which contains a diverse set of patient health indicators. The dataset undergoes rigorous preprocessing steps such as data cleaning, feature encoding, and normalization to ensure optimal model performance. The models are then evaluated based on commonly accepted performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.
To ensure practical applicability, the system is integrated into a user-friendly Android application. Patients can input their health data directly through the app, which then communicates with a Firebase-based backend. The backend processes the input data using the trained machine learning model and sends back the prediction result in real time. This real-time interaction allows for convenient, remote assessment, making it especially valuable in areas with limited access to immediate medical care.
Furthermore, the application demonstrates how mobile technology combined with machine learning can bring preventive diagnostics directly to users, enabling timely lifestyle changes or medical intervention. This project not only contributes to improving heart disease prediction but also highlights the importance of developing accessible digital health tools that can be deployed at scale.
The primary objectives of this research are:

- To analyze and compare the effectiveness of multiple machine learning algorithms for heart attack prediction.
- To identify the most significant clinical features contributing to heart disease prediction.
- To develop a mobile-based platform integrated with Firebase for real-time risk evaluation.
- To offer a scalable solution that bridges the gap between patients and healthcare providers through technology.

In conclusion, this project underscores the transformative potential of machine learning in healthcare and proposes a viable approach to reducing the burden of heart disease through early and accurate risk prediction.

II.LITERATURE REVIEW

2.1 Problem Definition

Heart disease, particularly myocardial infarction (heart attack), is a life-threatening condition that affects millions of people globally. The increasing prevalence of cardiovascular diseases has emphasized the need for timely diagnosis and preventive care. Traditional diagnostic methods rely on clinical expertise and manual interpretation of patient data, which can be error-prone and time-consuming. The complexity of symptoms, varying risk factors, and interrelated clinical indicators make early detection a challenge.
This project addresses the problem of predicting the possibility of a heart attack using machine learning techniques. By analyzing patterns in patient data — including age, sex, blood pressure, cholesterol, ECG results, and heart rate — the aim is to build a system that can classify individuals as at-risk or not at-risk for heart disease. The goal is not only to maximize prediction accuracy but also to ensure model interpretability and generalizability for real-world healthcare applications.

2.2 Review of Existing Systems

Over the past decade, researchers have explored numerous approaches to predict heart disease using computational models. Traditional systems, often rule-based or statistical in nature, have limited accuracy due to their inability to handle complex, nonlinear relationships within data. Early models based on logistic regression and decision trees showed promise, but were constrained by fixed feature relationships and high sensitivity to outliers.
Several research studies have utilized the UCI Heart Disease dataset, applying models such as Decision Trees, Support Vector Machines, and Naïve Bayes classifiers. While these models offered moderate prediction accuracy, they often lacked robustness and failed to generalize well on new, unseen data. Additionally, most systems lacked real-time prediction capability and were not accessible to end-users without technical knowledge.
The integration of machine learning into clinical diagnostics has shown that more sophisticated models — particularly ensemble methods like Random Forest and Gradient Boosting — can outperform individual classifiers. However, many of these solutions have remained in academic or experimental stages, lacking practical deployment on mobile or cloud-based platforms.

2.3 Proposed System

The proposed system aims to overcome the limitations of existing methods by leveraging the power of multiple machine learning algorithms in a comparative framework. It involves the development of an intelligent, mobile-integrated prediction system using algorithms such as Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Decision Tree, and Random Forest. Each model is trained on a cleaned and normalized dataset and evaluated using performance metrics such as accuracy, precision, recall, F1-score, and AUC.

To enhance accessibility and practical application, the system is deployed as an Android application. Firebase acts as a real-time backend, facilitating communication between the user interface and the trained ML model. Users input their health data into the app, which is transmitted to Firebase. The server-side model processes this data and returns the risk prediction, which is immediately displayed on the user's device.

Advantages of the Proposed System

- **Early Detection**: ML algorithms can identify patterns in large and complex datasets that might be missed by traditional analysis, enabling early diagnosis.
- **Improved Accuracy**: Using ensemble and comparative techniques boosts predictive performance beyond that of conventional models.
- **Scalability**: The use of Firebase and Android ensures scalability, allowing the system to be used in various locations and by diverse users.
- **User-Friendly Interface**: The mobile app interface simplifies usage for non-technical users, making the system viable in both clinical and home settings.

2.4 Challenges and Considerations

While machine learning presents a promising approach, real-world implementation must account for:

- **Data Imbalance**: Clinical datasets often have skewed distributions which can bias predictions.
- **Model Interpretability**: Medical professionals prefer interpretable models; hence, black-box models must be paired with explanation mechanisms (e.g., SHAP values).
- **Data Privacy**: Patient data must be handled securely, ensuring compliance with HIPAA or equivalent standards.
- **Clinical Integration**: The model should be integrated into clinical workflows without disrupting existing practices.

III.METHODOLOGY / REQUIREMENTS

3.1 Methodology Overview

The methodology adopted in this study involves designing, training, testing, and deploying multiple machine learning models to predict heart attack risk based on clinical and demographic features. The system follows a structured machine learning workflow, which includes data preprocessing, model selection, training, evaluation, and real-time application integration through an Android mobile app using Firebase.

The following key steps define the methodology:

1. **Dataset Acquisition**
   The UCI Heart Disease dataset is used, which includes key medical attributes such as age, sex, chest pain type, cholesterol, blood pressure, and ECG results.
2. **Data Preprocessing**
   - **Data Cleaning**: Missing and null values are identified and removed.
   - **Feature Scaling**: Numerical features are standardized using techniques such as z-score normalization.
   - **Encoding**: Categorical attributes are transformed using one-hot encoding.
   - **Feature Selection**: Irrelevant or redundant features are eliminated to improve model performance.
3. **Model Selection and Training**
   Four supervised ML algorithms are implemented:
   - **Support Vector Machine (SVM)**
   - **k-Nearest Neighbors (KNN)**
   - **Decision Tree**
   - **Random Forest**
     The models are trained using the training set (derived via stratified splitting) and validated using k-fold cross-validation to ensure generalizability.

4. **Performance Evaluation**
   Each model is evaluated based on:
   - **Accuracy**
   - **Precision**
   - **Recall**
   - **F1-Score**
   - **Area Under the ROC Curve (AUC)**
     The Random Forest algorithm demonstrated the best overall performance in terms of predictive power and robustness.
5. **System Deployment**
   A trained model is serialized using pickle and integrated into a backend API powered by Flask. The Android application interfaces with this backend via Firebase. Users submit health data through the mobile app, and the system returns predictions in real time.

3.2 Hardware Requirements

Component	Specification
Processor	Intel Core i3 or higher
RAM	4 GB or higher
Hard Disk	500 GB or higher
Mobile Device	Android-based smartphone with Internet access

3.3 Software Requirements

Component	Specification
Operating System	Windows 7/8/10 (64-bit)
Programming Language	Python 3.7
IDE / Tools	Jupyter Notebook, Android Studio, Flask
Libraries	scikit-learn, pandas, numpy, matplotlib, seaborn
Cloud Service	Google Firebase for real-time database and user interaction
Web Framework	Flask (for backend API deployment)

3.4 Justification for Technology Stack

- **Python** is chosen for its simplicity and the availability of powerful ML libraries such as scikit-learn.
- **Flask** is a lightweight Python web framework ideal for deploying ML models as APIs.
- **Firebase** enables real-time communication between the mobile app and backend, ensuring seamless prediction delivery.
- **Android Studio** is used to develop the user-facing app for wider accessibility on smartphones.

IV.TESTING

4.1 Introduction

Testing is a critical phase in both software engineering and machine learning lifecycle. For this project, testing ensures that the machine learning models used for heart disease prediction are reliable, accurate, and perform well on unseen data. The system undergoes multiple levels of testing—from unit testing of individual modules to system-level validation—to guarantee robustness and correctness.

Machine learning models are tested using a separate test dataset that was not used during training. The effectiveness of each model is evaluated using standard performance metrics, and the final system is also tested for functional and non-functional behavior through the Android app interface.

4.2 Types of Testing

4.2.1 Unit Testing

Unit testing focuses on verifying the correctness of individual components such as data preprocessing, model training, and result output. Each Python function or module (e.g., normalization, encoding, model serialization) is tested with expected inputs and outputs.

4.2.2 Integration Testing

Integration testing validates the interaction between components:

- Android app ↔ Firebase
- Firebase ↔ Python backend
- Backend ↔ Trained ML model
  This ensures seamless communication and consistent data flow throughout the system.

4.2.3 System Testing

System testing ensures that the entire application, from user input to prediction display, functions as expected. This includes testing for:

- Proper form submission from the Android app
- Real-time database updates in Firebase
- Accurate predictions from the ML model
- Timely notification delivery to the user

4.2.4 Validation Testing

The trained models are validated against the test dataset to evaluate generalization performance. Metrics used include:

- **Accuracy**: Overall correctness of the model.
- **Precision**: Correctness of positive predictions.
- **Recall**: Ability to detect all actual positives.
- **F1-Score**: Harmonic mean of precision and recall.
- **AUC-ROC**: Model's ability to distinguish between classes.

4.2.5 Security Testing

Security testing ensures that user data (e.g., age, blood pressure, ECG values) is securely transmitted and stored using Firebase. Proper access rules and authentication were tested to avoid unauthorized access.

4.2.6 User Acceptance Testing (UAT)

Conducted with real users, UAT ensures the system meets user expectations. Feedback was collected on:

- App usability
- Result clarity
- Response time
- Interface navigation

4.3 Test Cases (Examples)

Test Case	Input	Expected Output	Result
Model Prediction	Valid patient record	"At Risk" or "Not at Risk"	Pass
Firebase Communication	User submits data	Firebase stores & triggers prediction	Pass
Android App Output	Prediction returned	Displays accurate notification	Pass

4.4 Test Report Summary

Each machine learning model was tested using a 33% test split from the dataset. The **Random Forest** model achieved the highest accuracy, followed by **SVM**, **Decision Tree**, and **KNN**. Performance scores:

Algorithm	Accuracy	Precision	Recall	F1 Score	AUC
Random Forest	90.9%	91%	89%	90%	0.94
SVM	86.3%	87%	84%	85%	0.91
Decision Tree	84.2%	85%	83%	84%	0.88
KNN	80.3%	82%	78%	80%	0.85

V.RESULTS AND ANALYSIS

5.1 Model Evaluation Metrics

To assess the effectiveness of different machine learning algorithms in predicting heart attack risk, the system employs a standard evaluation framework using the following performance metrics:

- **Accuracy**: The proportion of total correct predictions.
- **Precision**: The ratio of true positives to all predicted positives.
- **Recall (Sensitivity)**: The ratio of true positives to all actual positives.
- **F1-Score**: The harmonic mean of precision and recall.
- **ROC-AUC (Receiver Operating Characteristic – Area Under Curve)**: Reflects the model's ability to distinguish between classes across various thresholds.

Each model was trained on a portion of the preprocessed dataset and tested on a reserved test set using k-fold cross-validation (typically 5- or 10-fold) to minimize overfitting and maximize generalizability.

5.2 Comparative Performance Analysis

Four supervised learning algorithms were implemented and compared:

Algorithm	Accuracy	Precision	Recall	F1 Score	AUC-ROC
Random Forest	**90.9%**	91%	89%	90%	0.94
Support Vector Machine (SVM)	86.3%	87%	84%	85%	0.91
Decision Tree	84.2%	85%	83%	84%	0.88
k-Nearest Neighbors (KNN)	80.3%	82%	78%	80%	0.85

Observation:

- **Random Forest** performed best overall, offering the highest accuracy and AUC, indicating superior classification power and generalizability.
- **SVM** provided good baseline accuracy and was effective on smaller datasets.
- **Decision Trees** were interpretable but slightly less accurate.
- **KNN** had the lowest performance, likely due to sensitivity to data scaling and neighborhood density.

5.3 Feature Importance Analysis (Random Forest)

The Random Forest algorithm also enabled ranking of feature importance. The most influential features included:

- **Age**
- **Chest Pain Type (cp)**
- **Max Heart Rate (thalach)**
- **Cholesterol Level**
- **ST Depression (oldpeak)**
- **Exercise-induced Angina (exang)**

This analysis is critical in clinical contexts as it informs practitioners about the most predictive health indicators for heart disease.

5.4 Visualization of Results

A bar chart summarizing model performance (from the original code) was generated using matplotlib and seaborn:

📊 **Model Score Visualization**:

- Random Forest: ~91%
- SVM: ~86%
- Decision Tree: ~84%
- KNN: ~80%

This graphical representation reinforces the numerical findings and provides visual evidence supporting the model selection.

5.5 Real-time Implementation Analysis

The best-performing model (Random Forest) was deployed into a Flask backend and integrated with an Android application via Firebase. The system was tested for:

- Real-time responsiveness
- User input validation
- Prediction delivery via mobile notification

The end-to-end system demonstrated low latency (<1.5 seconds), and prediction accuracy was consistent with offline results.

Conclusion of Analysis

Based on performance metrics and practical deployment, **Random Forest** emerged as the most reliable and accurate model for heart attack prediction. Its robustness, feature interpretability, and suitability for deployment make it an ideal choice for real-world clinical applications.

VI.DISCUSSION

The purpose of this project was to develop an intelligent, data-driven system for predicting heart attack risk using various machine learning (ML) algorithms. Through the comparative analysis of four major supervised learning techniques—Random Forest, Support Vector Machine (SVM), Decision Tree, and k-Nearest Neighbors (KNN)—the project demonstrated the potential of ML in transforming traditional healthcare diagnostics.

The results indicate that ensemble learning methods, particularly the **Random Forest algorithm**, outperformed other models in terms of accuracy, recall, F1-score, and AUC. This is largely due to its ability to combine the predictions of multiple decision trees, which reduces overfitting and increases generalization performance. Furthermore, Random Forest provides interpretable feature importance scores, allowing clinicians to understand which health indicators contribute most to prediction outcomes.

In contrast, **SVM** also performed well but was less robust to noisy or nonlinear data without kernel tuning. **KNN** suffered from performance degradation due to its sensitivity to data scaling and feature density, especially in high-dimensional spaces. **Decision Tree**, while interpretable and easy to implement, lacked the predictive power of ensemble models due to its tendency to overfit training data.

A key strength of this project is its **end-to-end implementation**. The integration of the trained ML model into a **mobile Android application** using **Firebase** enables real-time prediction delivery. This not only increases accessibility for users but also introduces a scalable tool that could be deployed in rural or resource-limited healthcare settings.

Despite these promising results, the system is not without limitations:

- The **dataset** used was limited in size and scope, relying on publicly available clinical records which may not represent broader population diversity.
- **Model fairness and bias mitigation** were not deeply explored, which are critical for ethical healthcare AI deployment.
- The app's predictions are not yet integrated into any certified medical advisory workflow and should not replace professional diagnosis.

From a practical standpoint, this system demonstrates how **machine learning can augment healthcare** by providing early risk assessments, enabling preventive care, and empowering users with actionable insights. Future iterations can expand to include:

- **Larger and more diverse datasets**
- **Real-time health monitoring via wearables**
- **Integration with hospital information systems**
- **Doctor-patient communication platforms for teleconsultation**

Ultimately, the project confirms that intelligent systems, when designed carefully and ethically, can play a significant role in reducing heart disease burden and enhancing patient outcomes.

VIICONCLUSION

This project successfully developed and evaluated a machine learning-based system to predict the likelihood of heart attacks using clinical data. By implementing and comparing multiple algorithms—Support Vector Machine (SVM), Decision Tree, k-Nearest Neighbors (KNN), and Random Forest—the study identified Random Forest as the most effective model in terms of accuracy, precision, and robustness.

The integration of the predictive model into a mobile Android application, supported by a Firebase backend, demonstrates the practical applicability of machine learning in real-world healthcare settings. The system provides users with instant risk assessments based on their health inputs, making it a valuable tool for early diagnosis and preventive intervention.

Overall, the project highlights the transformative potential of machine learning in medical diagnostics, particularly when paired with accessible mobile technologies. It contributes to the ongoing efforts to enhance healthcare delivery through intelligent, data-driven solutions.

VIII.REFERENCES

[1] Kannel, William B.” Contribution of the Framingham Study to preventive cardiology.” Journal of the American College of Cardiology 15.1(1990): 206-211.

[2] Mahmooda, S. S., Levy, D., Vasan, R. S., Wang, T. J. (2014).The Framingham Heart Study and the epidemiology of cardiovascular diseases: a historical perspective. Lancet, 383(9921), 999-1008.

[3] Ahmed, Fizar. ”An Internet of Things (IoT) application for predicting the quantity of future heart attack patients.” International Journal of Computer Applications 164.6 (2017).

[4] Methaila, A., Kansal, P., Arya, H., Kumar, P. (2014). Early heart disease prediction using data mining techniques. Computer Science Information Technology Journal, 53-59.

[5] Rajkumar, Asha, and G. Sophia Reena.” Diagnosis of heart disease using data mining algorithm.” Global journal of computer science and technology 10.10 (2010): 38-43.

[6] Salman, Issam.” Heart attack mortality prediction: an application of machine learning methods.” Turkish Journal of Electrical Engineering Computer Sciences 27.6 (2019): 4378-4389.

[7] Davis, Jesse, et al.” Machine learning for personalized medicine: Will this drug give me a heart attack.” The Proceedings of International Conference on Machine Learning (ICML). 2008.

[8] Sharma, Himanshu, and M. A. Rizvi.” Prediction of heart disease using machine learning algorithms: A survey.” International Journal on Recent and Innovation Trends in Computing and Communication 5.8 (2017):99-104.

[9] Kaur, Beant, and William Jeet Singh.” Review on heart disease prediction system using data mining techniques.” International journal on recent and innovation trends in computing and communication 2.10 (2014):3003-3008.

[10] Dangare, Chaitrali, and Sulabha Apte.” A data mining approach for prediction of heart disease using neural networks.” International Journal of Computer Engineering and Technology (IJCET) 3.3 (2012).