IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Heart Disease Prediction Using Machine Learning

Sudhadevi K, Sampath Kumar N , Sree Aravind K , Vinothkumar V

Assistant Professor, Student, Student, Student,

Department of Computer Science and Engineering,

Paavai Engineering College (Autonomous),

Pachal, Namakkal, Tamil Nadu., India.

Abstract: The project presents a **Heart Disease Prediction System** leveraging a **Decision Tree Classifier** to identify the likelihood of heart disease based on medical parameters. The dataset includes key features such as **age, cholesterol levels, blood pressure**, and **exercise-induced metrics**, which are crucial indicators of cardiovascular health. The system is designed to process these inputs, train the model, and deliver predictions with commendable accuracy. By implementing a **Flask-based web application**, the solution ensures user-friendly and real-time access to the predictive tool, making it ideal for both **healthcare professionals** and individuals. This platform can facilitate early detection, enabling proactive measures to prevent severe outcomes. It not only aids in clinical decision-making but also empowers individuals to monitor their heart health. The combination of machine learning and web technologies enhances accessibility, paving the way for efficient, data-driven healthcare solutions. With its robust design, the system contributes to better **heart disease risk assessment** and management.

Index Terms -Heart disease prediction, Decision Tree Classifier, Machine learning, Medical parameters, Cardiovascular health, Risk assessment, Real-time predictions, Flask-based web application, Healthcare professionals, Individual health monitoring, Early detection, Preventive healthcare, Data-driven healthcare solutions, User-friendly interface.

I. INTRODUCTION

Heart disease remains one of the leading causes of death worldwide, underscoring the importance of early detection and proactive management. This project introduces a **Heart Disease Prediction System** that leverages the power of machine learning through a **Decision Tree Classifier**. The model is trained on a comprehensive dataset containing crucial medical parameters such as **age, sex, cholesterol levels, blood pressure,** and **exercise-induced metrics**. By analyzing these features, the system evaluates an individual's risk level and classifies them as either at risk or not, offering an efficient and reliable predictive solution.

To ensure widespread usability, the system is deployed as a **Flask-based web application**, providing a seamless, **user-friendly interface** for real-time predictions. This makes it an invaluable tool for **healthcare professionals**, enabling them to make timely and informed decisions about patient care. Simultaneously, it empowers individuals to assess their cardiovascular health, encouraging proactive monitoring and lifestyle adjustments.

By integrating advanced **machine learning techniques** with healthcare, this project addresses the critical need for **early detection** and **preventive measures** in managing heart disease. Its robust design supports **risk assessment** while promoting **data-driven healthcare solutions**, ultimately contributing to better patient outcomes and reducing the global burden of cardiovascular diseases.

II.OBJECTIVE:

1) **Early Detection:** Facilitate the timely identification of individuals at risk of heart disease to enable early intervention and management.
2) **Risk Assessment:** Develop a reliable machine learning model using a Decision Tree Classifier to classify individuals based on heart disease risk.
3) **Integration of Key Medical Parameters:** Analyze critical factors such as age, sex, cholesterol levels, blood pressure, and exercise metrics to predict cardiovascular risk.
4) **Accessible Technology:** Deploy the system as a Flask-based web application with a user-friendly interface to ensure ease of use for both healthcare professionals and individuals.
5) **Support for Healthcare Professionals:** Provide a decision-support tool to aid clinicians in making informed decisions regarding patient care and treatment.
6) **Empower Individuals:** Enable individuals to monitor their heart health and take preventive measures to reduce the risk of heart disease.
7) **Data-Driven Healthcare Solutions:** Leverage machine learning and real-time predictions to deliver accurate and efficient healthcare outcomes.
8) **Global Impact:** Contribute to reducing the global burden of cardiovascular diseases by promoting preventive healthcare and innovative solutions.

III.EXISTING IDEA:

Heart disease is one of the most significant causes of mortality in the world today. Prediction of cardiovascular disease is a critical challenge in the area of clinical data analysis. Machine learning (ML) has been shown to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry. We have also seen ML techniques being used in recent developments in different areas of the Internet of Things (IoT). Various studies give only a glimpse into predicting heart disease with ML techniques. In this paper, we propose a novel method that aims at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease. The prediction model is introduced with different combinations of features and several known classification techniques. We produce an enhanced performance level with an accuracy level of 88.7% through the prediction model for heart disease with the hybrid random forest with a linear model (HRFLM).

Disadvantages:

1.Dependence on Data Quality: The accuracy of the prediction model heavily relies on the quality, completeness, and relevance of the dataset used for training. Incomplete or noisy data can negatively affect performance.

2. Limited Generalization: The model might perform well on the training dataset but may not generalize effectively to unseen or diverse datasets due to overfitting or insufficient feature diversity.

3. Feature Selection Challenges: While the method emphasizes finding significant features, the process of feature selection may overlook some less obvious but impactful features, leading to a suboptimal model.

4. Model Complexity: The hybrid random forest with a linear model (HRFLM) approach adds complexity to the system, which may increase computational costs and make real-time predictions less efficient.

5. Lack of Interpretability: Hybrid models combining random forests and linear models may reduce interpretability, making it harder for healthcare professionals to understand the rationale behind specific predictions.

6. Resource Intensity: Training a hybrid model requires significant computational resources, which may not be feasible for all organizations or for deployment in resource-limited settings.

IV.RELATED WORKS:

There is significant related work in the field of heart disease prediction and classification, which directly aligns with this study. **Artificial Neural Networks (ANNs)** have been employed to achieve high-accuracy predictions in the medical domain. Specifically, the **backpropagation multilayer perceptron (MLP)** of ANN has been utilized for heart disease prediction, demonstrating improved performance when compared to existing models.

Heart disease patient data collected from the **UCI laboratory** has been analyzed using various algorithms, including **Neural Networks (NN), Decision Trees (DT), Support Vector Machines (SVM),** and **Naive Bayes.** These methods were evaluated based on performance and accuracy, with a proposed hybrid method achieving an **F-measure of 86.8%,** outperforming many existing approaches.

Additionally, **Convolutional Neural Networks (CNNs)** have been applied without segmentation to process **Electrocardiogram (ECG)** signals. This method considers heart cycles starting at different positions during training, enabling CNNs to extract features effectively during testing.

The study highlights that the vast amount of data generated by the medical industry has often been underutilized. The proposed machine learning (ML) and deep learning (DL) techniques reduce costs and enhance the ease and effectiveness of heart disease prediction, demonstrating high accuracy and efficacy.

V.PROPOSED IDEA:

The proposed system utilizes a **Decision Tree Classifier** to predict heart disease risk based on essential medical parameters, including **age, sex, cholesterol levels,** and **exercise data**. The machine learning model is trained on a comprehensive dataset, achieving high accuracy in classifying individuals at risk. To enhance accessibility, a **Flask-based web application** is developed, offering a user-friendly interface for real-time predictions. This enables both **individuals** and **healthcare professionals** to quickly assess heart disease risk, making it a valuable tool for informed decision-making. The system provides a **cost-effective** and **efficient alternative** to traditional diagnostic methods, enabling earlier detection of potential cardiovascular issues. By integrating machine learning with a web application, the system ensures wide accessibility, promoting preventive healthcare and supporting proactive health management. Ultimately, the solution aims to enhance the accuracy, accessibility, and affordability of heart disease risk assessment, contributing to better patient outcomes.

VI.PROPOSED ARCHITECTURE:

1) Data Collection

- **Input Data:** The system uses key medical parameters like **age, sex, cholesterol levels, blood pressure**, and **exercise data**, collected from healthcare records or user inputs through the web application.

2) Preprocessing and Feature Selection

- **Data Cleaning:** Raw data is preprocessed to handle missing values, outliers, and normalization to ensure that the data is suitable for training.
- **Feature Selection:** Significant features are identified from the dataset, such as cholesterol levels, age, and exercise habits, which contribute the most to heart disease prediction.

3) Model Training (Decision Tree Classifier)

- The **Decision Tree Classifier** is trained on the processed dataset. The model uses various medical parameters to classify an individual as either at risk or not at risk for heart disease.
- **Training Environment:** The model is developed using a **Python-based machine learning framework** (e.g., **scikit-learn**) and trained on a large medical dataset.

4) Web Application (Flask)

- A **Flask-based web application** provides a simple and accessible user interface.
- **User Input:** Users (individuals or healthcare professionals) enter their medical details through the web interface.
- **Prediction Output:** Once the user inputs the data, the Flask app calls the trained model to provide a real-time risk prediction for heart disease.

5) Model Inference and Prediction

- The trained Decision Tree model processes the user's data, classifying the individual as **at risk** or **not at risk** of heart disease.
- The result is displayed on the web interface, providing instant feedback to the user.

6) Results and Feedback

- The system offers **real-time predictions** along with health recommendations based on the classification result.
- **User Interface:** The web application provides clear, understandable results for both the user and healthcare professionals.

7) Data Storage and Security

- **Data Security:** Users' personal health data is stored securely, following compliance with healthcare data protection standards (e.g., **HIPAA**).
- **Database:** All predictions and user interactions are logged for future reference, allowing for continuous model improvement.

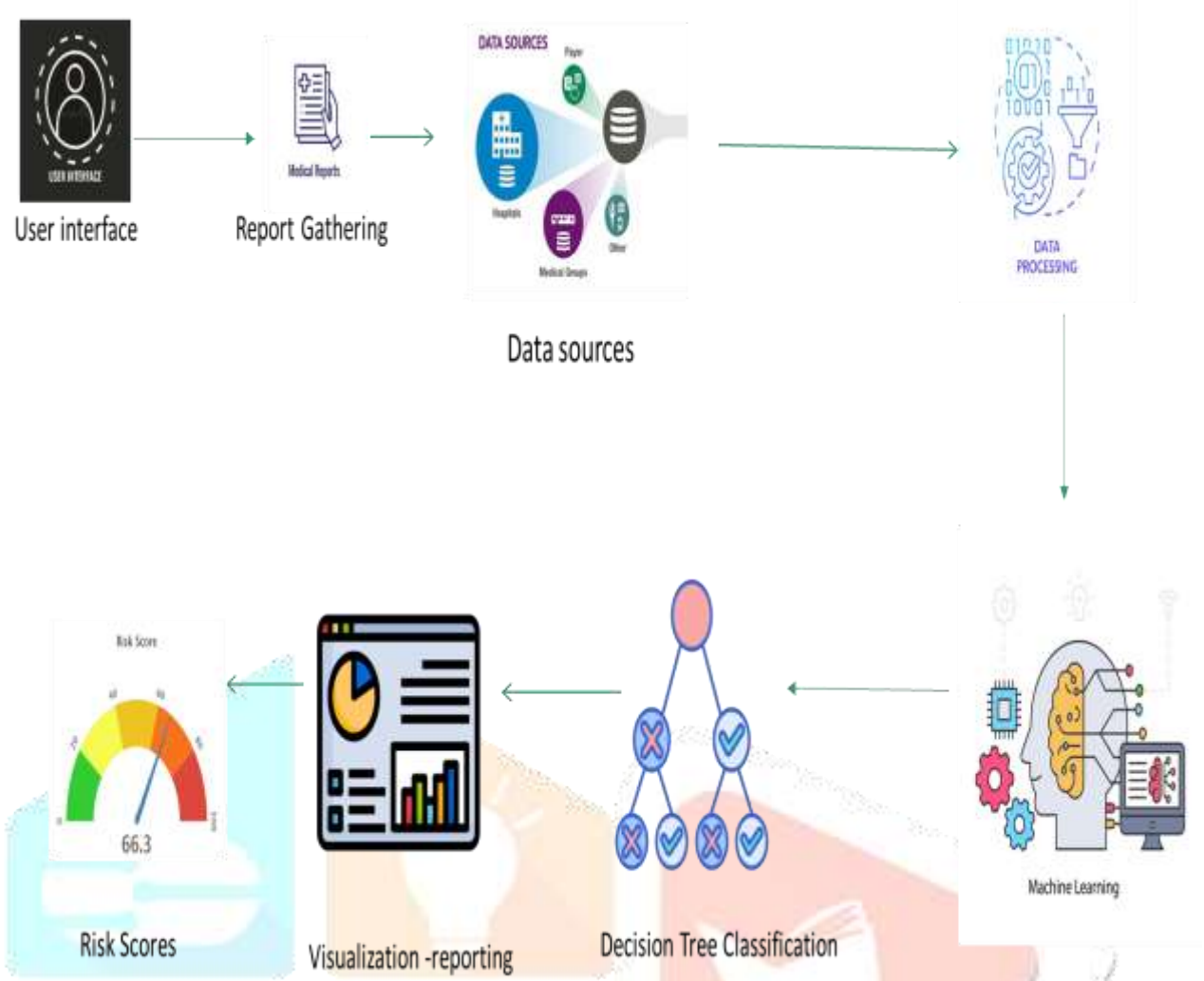

FIG-PROPOSED ARCHITECTURE

VII.CONCLUSION:

In conclusion, the proposed **Heart Disease Prediction System** effectively combines **machine learning** with a **Flask-based web application** to provide an accessible, cost-effective, and accurate solution for predicting heart disease risk. By utilizing a **Decision Tree Classifier** trained on key medical parameters such as **age, sex, cholesterol levels**, and **exercise data**, the system offers high accuracy in classifying individuals as at risk or not at risk. The integration of a user-friendly interface ensures that both **healthcare professionals** and **individuals** can quickly access real-time predictions, facilitating informed decision-making and proactive health management. This system enhances early detection, allowing for timely interventions and reducing the reliance on traditional diagnostic methods. By offering an easily accessible platform, the solution promotes **preventive healthcare** and contributes to better cardiovascular health outcomes. Ultimately, the system supports the transition towards more **data-driven**, **efficient**, and **affordable healthcare** solutions, improving heart disease risk assessment on a broader scale.

VIII.REFERENCES:

1. M. Gupta and S. Sharma, “Heart Disease Prediction using Decision Trees and Logistic Regression”.
2. P. Sharma and R. Gupta, “Using machine using machine learning for early detection of heart disease: a decision tree approach for early detection of heart disease: a decision tree approach”.
3. S. Thomas and A. Rao, “Optimization of Decision Trees for Heart Disease Prediction”.
4. J. Zhang and S. Liu, “Decision Tree Classifiers for Predicting Heart Disease Risk”.
5. X. Zhang and Y. Chen, “Prediction of Heart Disease using Decision Tree in Comparison with KNN to Improve Accuracy”.