PREDICTIVE ANALYTICS IN CARDIOLOGY: TOWARDS EARLY DETECTION OF HEART DISEASES

Prof. Poonam Dharpawar*1, Tanuja Somvanshi*2, Madhuja Patil*3, Minal Raigire*4

*1,2,3,4Department Of Data Science Usha Mittal Institute Of Technology Mumbai, Maharashtra, India.

DOI: https://www.doi.org/10.56726/IRJMETS51397

ABSTRACT
This project aims to develop a heart disease prediction system using Python programming language. Leveraging a comprehensive dataset containing diverse patient attributes including age, gender, blood pressure, cholesterol levels, and other clinical parameters. Through exploratory data analysis, significant patterns and correlations within the dataset are identified to facilitate informed feature selection and model training. The developed prediction system not only aids in the early detection of heart disease but also provides insights into the crucial factors contributing to its onset. Implemented in Python, this project offers a practical and efficient tool for healthcare professionals to assess and mitigate the risk of heart disease, thus contributing to improved patient outcomes and healthcare management strategies. Heart disease prediction is a must for nowadays.

Keywords: Heart Disease Prediction, Predict Heart Disease, Name, Email, Etc

I. INTRODUCTION
Heart disease continues to be a major global health concern, responsible for a significant proportion of morbidity and mortality worldwide. Timely detection and effective management of heart disease risk factors are crucial for reducing its burden on individuals and healthcare systems. With advancements in technology and data analytics, machine learning techniques have emerged as powerful tools for predictive modeling and risk assessment in healthcare. Heart disease is a significant health concern globally, and early prediction is crucial for effective prevention and treatment. In this Python project, we aim to develop a Heart Disease Prediction system that utilizes machine learning techniques to analyze various factors such as gender, resting blood pressure, and maximum heart rate to predict the likelihood of heart disease in individuals. This project will include functionalities for user registration and login to ensure data privacy and security. Users will be able to input their demographic information and health metrics, and the system will utilize a trained machine learning model to predict their risk of developing heart disease.

II. LITERATURE REVIEW
[2] S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction" Using, Hybrid Machine Learning Techniques," in IEEE Access, vol. 7, pp., 2019, doi:/ACCESS. Machine learning models such as SVM, NN, and KNN, were implemented with custom tuning and compared. Finally, a novel method called HRFLM was developed outperforming the existing models. Boukhatem, H. Y. Youssef and A. B. Nassif, "Heart Disease Prediction Using Machine Learning," 2022 Advances in Science and Engineering Technology Inter-national Conferences, (ASET), Dubai, United Arab Emirates, 2022,
[4] Lakshmana Rao et al proposed "Machine Learning Techniques for Heart Disease Prediction" in which the contributing elements for heart disease are more. These literature survey papers offer valuable insights into the current state of research and development in heart disease prediction using machine learning techniques.

They provide a foundation for understanding the methodologies, challenges, and opportunities in the field, which can inform the design and implementation of the Heart diseases prediction using python.
A. Workflow

III. PROPOSED SYSTEM

In proposed system of predictive analytics in cardiology focuses on leveraging advanced data analytics techniques to enable the early detection of heart diseases, with particular emphasis on key physiological parameters including cholesterol levels, resting blood pressure, maximum heart rate, and electrocardiogram (ECG) data. These parameters serve as critical indicators of cardiovascular health and play pivotal roles in assessing the risk of developing various heart conditions.
By integrating cutting-edge data analytics techniques, the system offers a multifaceted approach to cardiovascular risk assessment, incorporating key physiological parameters such as cholesterol levels, resting blood pressure, maximum heart rate, and electrocardiogram (ECG) data. This comprehensive approach enables personalized risk profiling, empowering healthcare professionals to tailor interventions and preventive strategies to individual patient needs. With a focus on transparency and interpretability, the system facilitates informed clinical decision-making, providing clinicians with actionable insights to guide patient management effectively.

These parameters serve as critical indicators of cardiovascular health and play pivotal roles in assessing the risk of developing various heart conditions. By integrating these parameters into the predictive analytics framework, the system can provide accurate risk assessment, enabling timely interventions and personalized preventive strategies. Through the analysis of multidimensional data, the system enhances the ability to detect subtle abnormalities and patterns predictive of heart diseases, ultimately improving patient outcomes and reducing the burden of cardiovascular morbidity and mortality.

IV. DESIGN AND IMPLEMENTATION DETAILS

A. Import Libraries

i) Pandas: Pandas is a powerful data manipulation and analysis library for Python. It provides data structures like DataFrame and Series, which allow for easy handling and manipulation of structured data. Pandas is commonly used for data pre-processing, cleaning, and transformation tasks in data science and machine learning projects.

ii) matplotlib.pyplot: Matplotlib is a popular data visualization library in Python. The pyplot module provides a MATLAB-like interface for creating static, interactive, and publication-quality plots and visualizations. It is widely used for creating charts, histograms, scatter plots, and other types of graphical representations of data.

iii) tkinter: Tkinter is the standard GUI (Graphical User Interface) toolkit for Python. It provides a set of tools and widgets for creating desktop applications with graphical interfaces. Tkinter is easy to use and comes pre-installed with Python, making it a popular choice for developing simple GUI applications.

iv) Scikit-learn model selection: Scikit-learn is a comprehensive machine learning library for Python. The model selection module provides functions for splitting datasets into training and testing sets. The train_test_split function is commonly used to split data into training and testing subsets which is essential for evaluating model performance.

v) Scikit-learn linear model: The linear model module in scikit-learn provides implementations of various linear models for regression and classification tasks. Logistic regression is a classification algorithm used for binary and multiclass classification problems. It is commonly used for predictive modeling tasks in healthcare finance and other domains.

vi) Scikit-learn metrics: The metrics module in scikit-learn provides functions for evaluating the performance of machine learning models. The accuracy score function calculates the accuracy of classification models by comparing predicted labels with true labels. It is a commonly used metric for assessing classification model performance.

B. Data Collection and Preprocessing

1) We loaded the dataset using the pandas library, reading the heartdiseasedataset.csv file into a DataFrame named data. This dataset contains information related to heart disease, with features such as age, sex, cholesterol levels, and other physiological parameters. After loading the dataset, we performed data preprocessing steps to prepare it for modeling. We separated the features (X) from the target variable (y), where the target variable ‘target’ indicates the presence or absence of heart disease.

2) We then split the data into training and testing sets using the train_test_split function from the scikit-learn library. The training set X_train y_train comprises 80% of the data while the testing set X_test y_test contains the remaining 20%. This random splitting ensures that the model is trained on a subset of the data and evaluated on unseen data to assess its generalization performance. Next, we trained a logistic regression model using the Logistic Regression class from scikit-learn. Logistic regression is a commonly used algorithm...
for binary classification tasks, making it suitable for predicting the presence or absence of heart disease based on the input features.

3) The model fit method was called to train the logistic regression model on the training data Xtrain, ytrain where the model learned the underlying patterns and relationships between the features and the target variable. The resulting trained model can be used to predict the presence or absence of heart disease in new, unseen data, based on the learned patterns from the training data by using machine learning algorithm.

C. Development

1) Age is a fundamental demographic factor that significantly influences cardiovascular health. As individuals age, they become more susceptible to various cardiac conditions due to factors such as arterial stiffness, increased plaque accumulation, and diminished cardiac function. Predictive models often incorporate age as a predictive feature, as older age is associated with a higher risk of developing heart diseases, including coronary artery disease, heart failure, and arrhythmias.

2) Sex is another important demographic variable that impacts cardiovascular risk. Men and women may exhibit differences in the prevalence, presentation, and prognosis of heart diseases. For example, men tend to have a higher incidence of coronary artery disease at younger ages, while women may experience atypical symptoms and have a higher risk of adverse outcomes following certain cardiac events. Including sex as a predictive feature allows for gender-specific risk stratification and tailored preventive strategies. Sex is 1-Male, 0 Female.

3) Chest pain type, also known as angina, serves as a clinical indicator of underlying cardiac pathology. The characterization of chest pain according to its type e.g., typical angina, atypical angina, non-anginal chest pain provides valuable diagnostic information regarding the likelihood of coronary artery disease and the urgency of further evaluation. Predictive analytics models often incorporate chest pain type as a symptomatology feature to enhance the accuracy of risk prediction and guide clinical decision-making.

D. Checking Parameters

1) Resting Blood Pressure: Resting blood pressure refers to the blood pressure measurement taken when an individual is in a relaxed, non-stimulated state, typically in a seated or lying position. This measurement provides a baseline assessment of blood pressure without the influence of external factors such as physical activity or stress. In predictive analytics in cardiology, the presence of elevated resting blood pressure serves as a crucial predictive feature for assessing cardiovascular risk. Individuals with hypertension are at increased risk of experiencing adverse cardiovascular events due to the sustained pressure exerted on the arterial walls, which can lead to vascular damage, atherosclerosis, and subsequent complications. Components of blood pressure i) Systolic Blood Pressure: This is the pressure exerted on the arterial walls when the heart contracts and pumps blood out into the arteries. Systolic blood pressure is represented by the higher number in a blood pressure reading, while stage 1 hypertension is defined as systolic blood pressure between 130 mmHg to 139 mmHg. Stage 2 hypertension is systolic blood pressure of 140 mmHg or higher. ii) Diastolic Blood Pressure: This is the pressure exerted on the arterial walls when the heart relaxes between beats and fills with blood. Diastolic blood pressure is represented by the lower number in a blood pressure reading, while stage 1 hypertension is diastolic blood pressure between 90 mmHg to 99 mmHg. Stage 2 hypertension is diastolic blood pressure of 100 mmHg or higher.

2) Cholesterol: Cholesterol is a type of fat found in the blood that plays a crucial role in various bodily functions. When cholesterol levels are greater than 200 mg/dL, it indicates elevated cholesterol levels, which can increase the risk of cardiovascular diseases, including coronary artery disease and stroke. In predictive analytics in cardiology, the presence of elevated cholesterol levels serves as a significant predictive feature for assessing cardiovascular risk. A condition that war-rants further evaluation and management to mitigate cardiovascular risk. Lifestyle modifications, including dietary changes, regular exercise, and medication, may be recommended to lower cholesterol levels and reduce the risk of cardiovascular complications.

Types of Cholesterol

i) LDL Cholesterol: LDL cholesterol is often referred to as “bad” cholesterol because it contributes to the buildup of plaque in the arteries, leading to atherosclerosis and an increased risk of heart disease. The optimal level of LDL cholesterol is generally considered to be less than 100 mg/dL. Levels between 100...
mg/dL and 129 mg/dL are still considered near optimal, but higher levels are associated with increased cardiovascular risk.

ii) HDL Cholesterol: HDL cholesterol is often referred to as "good" cholesterol because it helps remove LDL cholesterol from the bloodstream and transport it to the liver for excretion, thus reducing the risk of plaque buildup and cardiovascular disease. The optimal level of HDL cholesterol is typically considered to be 60 mg/dL or higher. HDL cholesterol levels between 40 mg/dL and 59 mg/dL are generally considered acceptable, but higher levels are associated with greater cardiovascular protection.

3) Resting ECG:

4) Resting electrocardiogram (ECG) is a diagnostic test used to assess the electrical activity of the heart while the patient is at rest. It involves recording the electrical signals generated by the heart as it beats onto a graph paper or digital display. A resting ECG provides valuable information about the heart's rhythm, rate, and overall cardiac health.

i) Resting ECG Category 0 - Rapid Irregular Atrial Activation: This category indicates the presence of rapid atrial activation patterns on the resting ECG. These patterns may suggest atrial fibrillation (AF), a common arrhythmia characterized by irregular and rapid heartbeats originating from the atria.

ii) Resting ECG Category 1 - Further Heart Rate Check But Do Not Check Immediate Risk: This category suggests the need for further evaluation of heart rate abnormalities detected on the resting ECG. While abnormalities may be present, immediate high-risk conditions requiring urgent intervention are unlikely. The ECG findings may indicate variations in heart rate, rhythm disturbances, or other abnormalities that warrant closer monitoring and follow-up assessments.

iii) Resting ECG Category 2 - Monitoring Possible: This category indicates the presence of ECG findings that suggest the need for ongoing monitoring and evaluation. While not necessarily indicative of acute or high-risk conditions, these findings warrant closer attention and periodic assessments. The ECG may reveal abnormalities such as prolonged QT intervals, ST-T wave changes, or other non-specific findings that require monitoring for potential progression or development of clinically significant conditions.

5) Maximum heart rate: When the maximum heart rate reaches 185 beats per minute (bpm), it signifies an elevated heart rate during exertion or stress. A maximum heart rate of 185 bpm or higher typically occurs during periods of intense physical activity, such as vigorous exercise or strenuous physical exertion. During exercise, the heart pumps more blood to deliver oxygen and nutrients to working muscles, resulting in an increase in heart rate to meet the body's increased metabolic demands. While an elevated heart rate during exercise is normal and expected, reaching a maximum heart rate of 185 bpm may indicate a high level of exertion. It's essential to consider individual factors such as age, fitness level, and overall health when interpreting maximum heart rate values. For most adults, the maximum heart rate can be estimated using the formula: Maximum Heart Rate = 220 - Age. However, this formula provides a rough estimate, and actual maximum heart rate values can vary widely among individuals.

i) Low Intensity (40 percentage to 50 percentage MHR): This heart rate zone corresponds to light physical activity or gentle exercise. Heart rate in this zone is approximately 40 percentage to 50 percentage of the maximum heart rate. Activities in this zone may include walking, leisurely cycling, or gentle stretching. Low-intensity exercise is suitable for beginners, those recovering from injury, or individuals looking to maintain general health and fitness.

ii) Moderate Intensity (50 percentage to 60 percentage MHR): The moderate intensity zone represents a comfortable pace that allows for sustained aerobic activity. Heart rate in this zone ranges from 50 Percentage to 60 Percentage of the maximum heart rate. Activities in this zone may include brisk walking, jogging, swimming, or cycling at a moderate pace. Moderate-intensity exercise improves cardiovascular fitness, endurance, and overall health.

iii) High Intensity/Fat Burning Zone (60 percentage to 70 percentage MHR): This heart rate zone is often referred to as the “fat-burning zone” because it encourages the body to burn a higher percentage of calories from fat for fuel. The heart rate in this zone ranges from 60 Percentage to 70 Percentage of the maximum heart rate. Activities in this zone may include running, cycling, or interval training at a challenging pace. High-intensity exercise increases calorie expenditure and promotes weight loss when
combined with a balanced diet.

iv) Max Intensity (85 percentageto 100 percentageMHR): The maximum intensity zone represents near-maximal effort, where heart rate approaches or reaches the individual’s maximum heart rate. Heart rate in this zone ranges from 85 percent to 100 percent of the maximum heart rate. Activities in this zone may include sprinting, high-intensity interval training (HIIT), or competitive sports. Max intensity exercise improves anaerobic capacity, speed, and power but should be approached with caution and tailored to individual fitness levels and goals. Represents near-maximal effort. In this zone, activities such as sprinting, high-intensity interval training (HIIT), or competitive sports are common. Engaging in exercises within this zone can significantly improve anaerobic capacity, speed, and power.

However, it’s crucial to approach maximum intensity exercises with caution. They should be tailored to individual fitness levels and goals to avoid overexertion or potential injury. Monitoring heart rate during exercise can help individuals stay within this zone and ensure they are pushing themselves appropriately while still being mindful of their limits.

E. Prediction Of Heart Diseases

Predicting coronary artery disease (CAD) with 100 percent certainty is a challenging task due to the complex interplay of genetic, lifestyle, and environmental factors influencing its development. However, if we hypothetically consider a scenario where CAD prediction reaches 100 percent accuracy, it would likely involve an advanced predictive analytics system leveraging comprehensive patient data. Coronary artery disease (CAD), also known as coronary artery heart disease or ischemic heart disease, is a common type of cardiovascular disease characterized by the narrowing or blockage of the coronary arteries, which supply oxygen-rich blood to the heart muscle. Here’s a detailed description of coronary artery disease:

Risk Factors: Several factors contribute to the development of CAD, including: Modifiable Risk Factors: These include high blood pressure, high cholesterol levels (particularly LDL cholesterol), smoking, obesity, physical inactivity, and poor diet (high in saturated fats, trans fats, and cholesterol). Non-Modifiable Risk Factors: These include advancing age, male gender (though risk in women increases after menopause), family history of CAD or heart disease, and certain genetic factors. Symptoms: The symptoms of CAD can vary widely and may include: Chest pain or discomfort (angina) - typically felt as pressure, tightness, or squeezing in the chest. Shortness of breath, especially during exertion or emotional stress. Fatigue, weakness, or lightheadedness. Pain, numbness, or tingling in the arms, shoulders, neck, jaw, or back (especially in women). Diagnosis: CAD is often diagnosed based on a combination of medical history, physical examination, imaging tests, and diagnostic procedures, including: Treatment: Treatment for CAD aims to relieve symptoms, reduce the risk of complications, and improve overall cardiovascular health. This may include: Lifestyle modifications (e.g., healthy diet, regular exercise, smoking cessation) Medications to control blood pressure, cholesterol levels, and blood clotting. Procedures such as percutaneous coronary intervention (PCI) with stent placement or coronary artery bypass grafting (CABG) for severe blockages. Cardiac rehabilitation programs to support recovery and improve heart health.

F. Result and Analysis

![Pie chart of Resting Blood Pressure]

Fig. 2. Pie chart of Resting Blood Pressure

www.irjmets.com  @International Research Journal of Modernization in Engineering, Technology and Science [5120]
Fig. 3. Histogram of Cholestrol

Fig. 4. Resting ECG
Max HR >= 185

Fig. 5. Pie chart of Max Heart Rate
In conclusion, the journey towards early detection of heart diseases through predictive analytics in cardiology represents a pivotal advancement in cardiovascular care. By harnessing the power of advanced data analytics techniques and leveraging key physiological parameters, predictive analytics systems offer unprecedented opportunities for personalized risk assessment, timely intervention, and preventive strategies. These systems hold the promise of revolutionizing how we approach cardiovascular health, shifting from reactive treatments to proactive measures aimed at intercepting diseases before they manifest clinically. Through continuous refinement and innovation, predictive analytics in cardiology stands poised to transform the paradigm of cardiovascular care, empowering healthcare providers with actionable insights and improving patient outcomes. As we navigate the complexities of integrating cutting-edge technologies, addressing ethical considerations, and ensuring equitable access to predictive analytics tools, we must remain steadfast in our commitment to advancing cardiovascular health for all. In future scope of predictive analytics in cardiology appears incredibly promising for improving the early detection of heart diseases. With advancing technology, we can expect to see the integration of cutting-edge tools like advanced imaging techniques, genomics, continuous remote monitoring, and artificial intelligence algorithms. These innovations will allow for more thorough assessments of individual risk factors and the development of personalized intervention plans tailored to each patient's unique needs. This could lead to the creation of precision prevention programs aimed at addressing risk factors before they escalate. Collaboration on a global scale and sharing of data will further refine predictive models and expand their applicability across diverse populations. However, as these technologies become more ingrained in healthcare, it's crucial to navigate ethical concerns related to patient privacy, data security, and transparency. By embracing these advancements while addressing associated challenges, predictive analytics in cardiology has the potential to significantly enhance our ability to detect heart diseases early, manage them effectively, and ultimately prevent their onset.
VI. REFERENCES


