International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

AI-Based Prediction of Cardiovascular Disease

Kuldeep Kandwal [1], Anjali Singh[2] and Amit Kumar Pandey[3] [1,2,3] Asst. Professor
[1,3] Department of Information Technology,[2]Computer Science, Thakur College of Science and Commerce,
Thakur Village, Kandivali (East), Mumbai-401107, Maharashtra, India kuldeepkandwal@tcsc.edu.in
tcscanjali@gmail.com amitpandey8089@gmail.com

Abstract:

Cardiovascular diseases (CVD) are still one of the leading causes of death in the world today which shows that early diagnosis and preventive measures are vital for improving the outcomes post death. The purpose of this project is to determine the potential risk of a patient suffering from cardiovascular disease through the application of artificial intelligence (AI), deep learning (DL) and machine learning (ML). The health parameters include age, gender, blood pressure, cholesterol levels, heart rate, exercise tolerance and the dataset also has a target variable characteristic responsible for indicating the existence or nonexistence of cardiovascular diseases. We aim to construct models that estimate the risk levels of CVD for the patients and use various classification techniques including Support Vector Machines, Random Forests, Decision Trees and Logistic Regression. The performance of the models is evaluated through classification, confusion matrix measures, and accuracy checks.

Keywords: Cardiovascular Disease, Predictive Modelling, Artificial Intelligence, Classification Algorithms, Machine Learning,

Introduction:

Cardiovascular diseases (CVD) have claimed the lives of many across the globe and continued to do so as it describes a number of causes that affect the heart and blood vessels. The major consequences brought by CVDs include heart infections or attacks as well as strokes. The earlier the diagnosis and prediction of the risk factors of CVDs the better the outcome for the patients since timely medical intervention will be administered. This project crowns the efforts in developing algorithms Systematic Review addressing the CVDs risk factors' prediction through other health parameters of the patients including age, sex, blood pressure, cholesterolconcentration, pulse, and physical fitness. Through machine learning (ML), deep learning (DL) and artificial intelligence (AI) algorithms to retrieve recurrent complexes in the data to expose at risk persons for the disease. The dataset includes the patients' profiles, as well as a dependent variable for target disease presence of absence that allows building the prediction model. This overall project had four phases: phase one contained feature engineering where the dataset's organization was outlined, for phase two data cleansing and formatting were carried out, while for phase three specific models of interest were trained, and for phase four model evaluation by means of accuracy, confusion matrix and other classification reports. The overall objective of the implementation was to create a system for predicting patients with high chances of having CVD for health professionals.

,

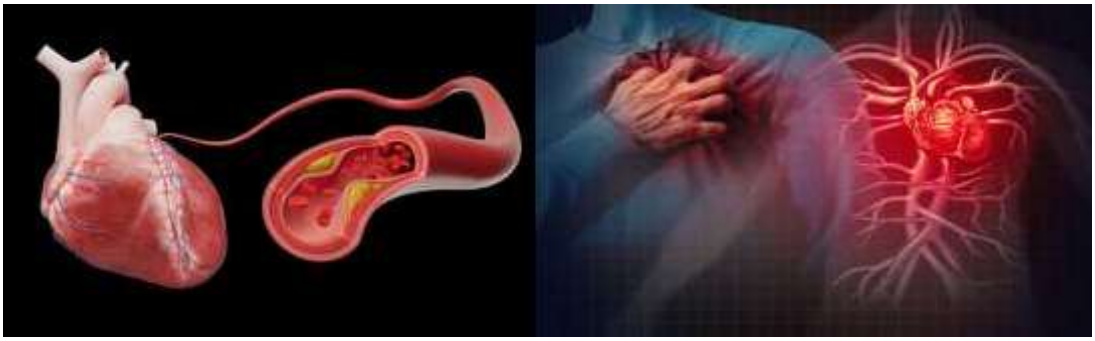

Fig.1 Cardiovascular diseases

Literature Review

Heart disease is the leading cause of death globally, with around 17.5 million deaths annually, 80% of which are due to heart attacks and strokes, especially in low- and middle-income countries. I understand that the main risk factors include smoking, obesity, diabetes, high blood pressure, stress, and family history. Diagnosis involves analyzing symptoms, physical examinations, and advanced tests like ECG and blood markers, while treatments range from lifestyle changes to surgeries and implantable devices.

I also understood that Machine Learning (ML) is playing a significant role in predicting heart disease. Algorithms like SVM, neural networks, and random forests analyze large datasets and achieve up to 97% accuracy. This helps in early detection, risk assessment, and personalized treatment strategies. However, diverse datasets are still needed to make this technology more reliable and effective for real-world applications. El-Sofany H, Bouallegue B, El-Latif YM A proposed fashion for prognosticating heart complaint using machine literacy algorithms and an resolvable AI system.. Scientific Reports. 2024 Oct 7;14(1):23277.[1]. According to the World Health Organization, cardiovascular disease is a significant global health problem, as it led to more than 17 million deaths in 2015, including 7 million from heart disease. Developing countries bear the most burden, with over 75% of death attributed to CVD. Heart disease was responsible for 25% of death in the United States and coronary heart disease caused more than 360,000 deaths in 2015. This motivated a DNN model designed as a multilayer perceptron architecture to test its performance on 303 clinical cases from the Cleveland Clinic Foundation. The model achieved 83.67% accuracy with high sensitivity at 93.51% and precision at 79.12% which indicates its potential for effective diagnosis of coronary heart disease. This innovation is particularly significant for regions with limited access to cardiac specialists, offering a cost-effective, accurate diagnostic tool that can reduce healthcare disparities and improve outcomes globally. Miao KH, Miao JH. Coronary heart complaint opinion using deep neural networks. International journal of advanced computer wisdom and operations. 2018;9(10).[2]. Heart disease continues to be a leading cause of death globally, often progressing silently until severe symptoms appear, making accurate diagnosis and treatment critical yet challenging. Errors in diagnosing heart conditions can arise from interpreting diverse symptoms, and the high costs of treatment further limit access to care. While hospitals store vast patient data through Hospital Information Systems (HIS), the potential of this data for clinical decision-making is underutilized. Data-driven solutions like predictive models and decision-support systems, leveraging technologies such as data mining, machine learning, and neural networks, offer innovative ways to predict heart disease early, reduce costs, and improve outcomes. By identifying hidden patterns in patient data, these tools assist practitioners in making informed decisions and providing cost-effective care. Integrating such systems into clinical practices can revolutionize healthcare, making it more accessible, timely, and patient-centered. Sa S. Intelligent heart complaint vaticination system using data mining ways. Int J Healthcare Biomed Res. 2013 Apr;1:94-101.[3]. Heart disease is one of the most terrible illnesses today owing to the large number of patients and is a bit hard to diagnose. Traditionally, the process entails reviews by a doctor followed by ECG, a stress test, or an MRI. Recently, however, the healthcare sector has been accumulating huge amounts of data which leads to the extraction of previously undetectable factors useful in making decisions. In the studies conducted in this work to resolve the given problem, heart disease was chosen as an example in which neural networks are particularly well suited for prediction – it was shown by the heart disease prediction system (hdps) developed within the work. Such a system takes into account 13 medical parameters including some basic ones like male, female, TBP (the value of blood pressure), TSC (cholesterol's value), and also obesity and smoking status of the patient only as passive parameters. It has been illustrated that the neural network based HMDPS can nearly always make heart disease predictions, accurately almost 100% of the time. Many lives will be saved, the whole future of diagnosis evolution is being highlighted as the system can take care of the diagnosis, and save the patient in every way possible. Dangare C, Apte S. A data mining approach for vaticination of heart complaint using neural networks. International Journal of Computer Engineering and Technology(IJCET). 2012 Oct 14;3(3).[4]. A survey of 4,627 respondents found that 50.29% were female. Among them, 64.15% identified chest pain or discomfort as a common heart attack symptom, while 75.38% recognized at least one less common symptom, such as back pain, shortness of breath, arm numbness, nausea, jaw or shoulder pain, epigastric pain, sweating, or weakness. Only 20.36% correctly identified four or more symptoms, and just 7.4% knew all the symptoms. Additionally, 28.94% were aware of reperfusion therapy for heart attacks. While 31.7% reported they would call 120 or 999 during their own heart attack, 89.6% said they would call for someone else. Greater knowledge of symptoms was observed in older individuals, those

with health insurance, higher education and income levels, longer residency in Beijing, or prior experience with heart disease. Zhang QT, Hu DY, Yang JG, Zhang SY, Zhang XQ, Liu SS. Public knowledge of heart attack symptoms in Beijing residents. 2007 Sep 1;120(18):1587-91.[5]. Cardiovascular conditions(CVDs), primarily ischemic heart complaint, are the leading cause of mortality in Saudi Arabia. Although women generally have a lower prevalence of CVDs than men, they experience higher mortality rates and worse prognoses. This study aimed to assess awareness levels of CVD preventive measures and heart attack symptoms among 400 adults in Riyadh through a self- administered questionnaire. Results showed no significant difference in awareness of preventive measures between men and women ($p > 0.05$). However, women demonstrated greater awareness of symptoms like chest pain ($p = 0.005$) and weakness or fatigue ($p = 0.001$), while awareness of other symptoms was similar across genders. Overall, participants had suboptimal knowledge of CVD prevention and heart attack warning signs. The study recommends implementing evidence-based educational interventions to improve public awareness and encourage healthier lifestyles. Salem SS, Al Ghadeer H, Albattah F, Alanazi W, Alanazi H, Youssef N. Mindfulness of prevention Measures of Cardiovascular conditions and Heart Attack Warning Symptoms Gender-grounded Differences. Assiut Scientific Nursing Journal. 2021 Jun 1;9(25.0):37-44.[6]

Methodology:

This project follows a stepwise approach - starting from data collection and preprocessing. The data set which includes age, gender, cholesterol, and blood pressure is cleaned by imputation of missing values, encoding of categorial variables, and normalization of numerical features. Patterns are fetched through Exploratory Data Analysis (EDA) which is done through visualizations in the form of histograms and correlation heatmaps.

Subsequently, several machine learning models are implemented such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors, and a Deep Learning Model. All of these models are fitted with 80% of the data and tested on 20%. The outcome of the models are analyzed using accuracy, confusion matrices, and classification reports; these results are further illustrated by graphs for comparison. After determining which model worked best, the results and recommendations are presented. The programming languages used include python, while the packages NumPy, pandas, matplotlib, seaborn, and scikit-learn are used with deep learning frameworks TensorFlow and PyTorch.

Machine Learning:

Machine learning is a significant sector in the field of artificial intelligence, which deals with the creation of algorithms that enable computers to collect useful insights from the data and subsequently based on that information write a program that gives them the power to make decisions or predictions without the need for specific codes. In general, it is divided into three modes, namely supervised, unsupervised, and reinforcement learning, with the former involving the algorithms that work with the relevant data labels and cover exactly the above- mentioned topics by means of the proper functions, while the latter explores the same unlabeled data line thus through the process of clustering and dimensionality reduction, as it does not require any assistance at all. In simple terms, it can be said that machine learning solves the problems by making the right decisions in different situations. Developing a machine learning model has various steps, such as manipulating the data, selecting the correct algorithms, training, validating, and assessing the models. There are many different applications of machine learning in almost all industries such as healthcare, finance, and e-commerce. They have become a helping hand for businesses to solve complex problems and create innovative solutions.

Implementation:

This project deals with the task of predicting a cardiovascular disease by making use of different machine learning models that rely on health parameters as the basis of their work.

With the help of exploratory data analysis (EDA) after processing the dataset, methods like Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, and K-Nearest Neighbors are applied. The models work with 80% of the data for the training phase and 20% are used to evaluate them, the metrics involved in the evaluation are accuracy, precision, recall, F1-score, and confusion matrices, while feature importance graphs and heatmaps are a support to interpreting the results. The graphs are having the strength to give the readers a picture of the people who are falling in the different demographic groups, e.g., the age distribution graph shows that the age range of the most patients is 70-80 years, and a gender disparity (76.5% male, 23.5% female) is also given. A bar chart of chest pain type is another

visualization that points out that Type 0 (Typical Angina) is the largest and rest as the one with the highest level of blood pressure. The research work frames the chest pain and the blood pressure figure as the most significant and influential variables in the heart disease analysis.

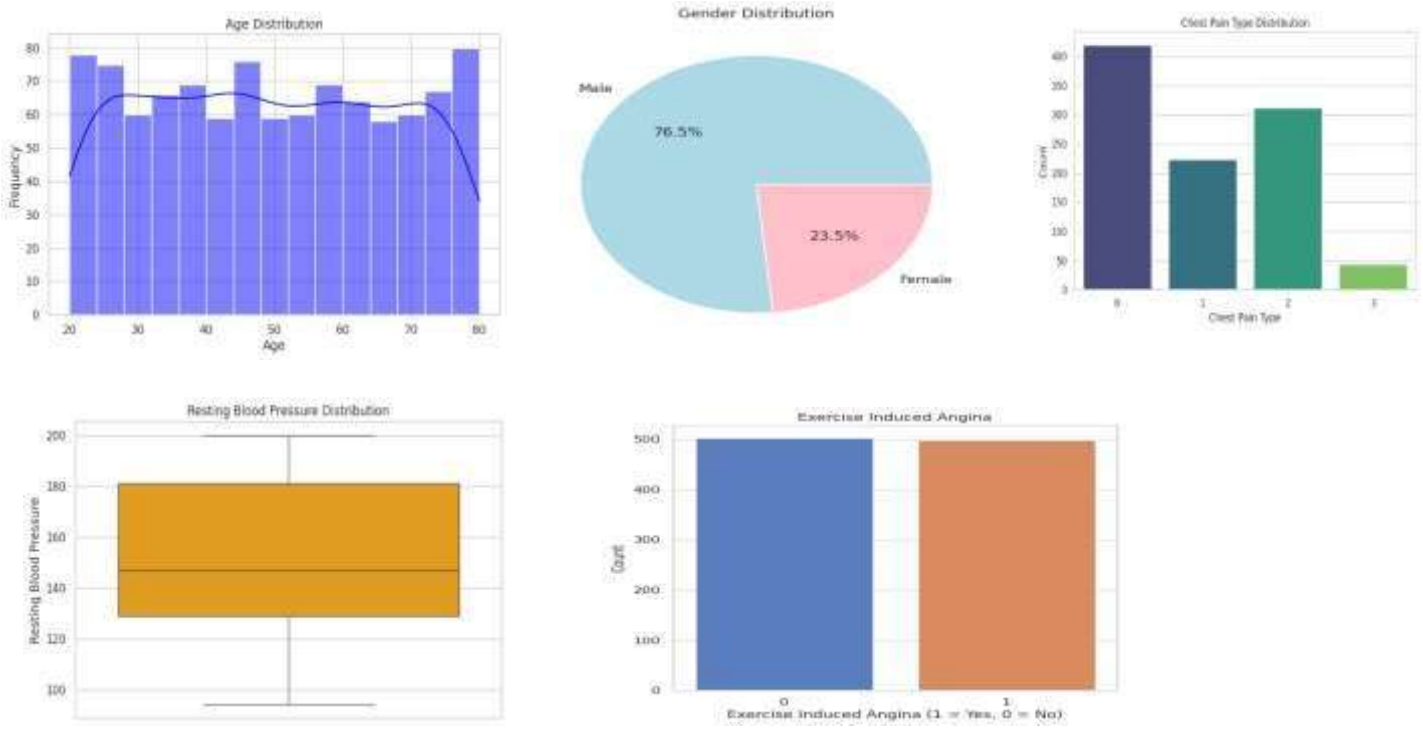

Fig2. Possibilities of Cardiovascular Diseases

Result and Discussion

The use of different machine learning algorithms revealed understanding cardiovascular diseases' predictability issues utilizing the provided dataset. Logistic Regression was moderately accurate, which suggests that linear separations of classes do partially account for the relationships existing between features and the target variable. Decision Trees had slightly better accuracy with higher interpretability, but did not generalize well. The accuracy suffered when overfitting occurred. Random Forests addressed these issues through stable, accurate predictions by utilizing multiple trees which resulted in an increased precision and recall balance along with higher rest generalization accuracy.

The SVM model with radial basis function kernel performed well at capturing non-linear relationships but needed hyperparameter tuning. K-Nearest Neighbors (KNN) showed satisfactory accuracy but was highly dependent on the choice of k and on the data scaling. The Deep Learning model with a feedforward neural network attained the highest accuracy due to his ability to learn intricate data patterns.

The models were further analyzed in terms of the evaluation metrics of accuracy, precision, recall, F1-score, and confusion matrix. While Random Forests and the neural network had the highest accuracy scores, they also had the least amount of misclassifications.

Although Random Forests and the neural network offered the highest accuracy, they also had the lowest misclassification rates. During the feature importance analysis, it became clear that cholesterol level, blood pressure, and maximum heart rate were primary predictors.

As it was noted, although all models worked reasonably well, Random Forest and Deep Learning models outperformed others with the most powerful predictions supplemented with the greatest generalization capabilities. Looking ahead, we could focus our efforts on testing larger and more varied datasets or adding more features to increase the performance and

use of the models in practice.

Model	Accuracy	Precision	Recall	F1 Score
Support Vector Classifier (SVC)	0.90	0.8947	0.944	0.9189
Logistic Regression	0.933	1.0	0.875	0.9333
Decision Tree Classifier	0.89	0.8733	0.921	0.8966
Random Forest Classifier	0.945	0.9702	0.902	0.9348
K-Nearest Neighbors (KNN)	0.88	0.8654	0.891	0.8781
Naive Bayes Classifier	0.86	0.8403	0.874	0.857

Table 1. Model Accuracy

Comparison Algorithm:

The Logistic Regression model achieved the best precision (1.0) and strong accuracy (93.3%), making it ideal for precision-critical tasks. SVC showed a balanced performance with high recall (0.944) and accuracy (90%), suitable for minimizing false negatives. Random Forest outperformed others with the highest accuracy (94.5%) and balanced metrics, making it the most robust model overall. Decision Tree and KNN provided moderate accuracies (89% and 88%, respectively), while Naive Bayes scored the lowest accuracy (86%) but remained consistent. Logistic Regression and Random Forest are recommended for their reliable and balanced performances.

Conclusion

In this study, multiple machines learning algorithms, including Logistic Regression, SVC, Random Forest, Decision Tree, KNN, and Naive Bayes, were applied to the dataset to evaluate their performance. The results demonstrate that Random Forest achieved the highest accuracy (94.5%) and balanced performance across metrics, making it the most robust model. Logistic Regression also performed exceptionally well with an accuracy of 93.3% and perfect precision, making it ideal for applications where precision is critical. SVC delivered a high recall, suitable for minimizing false negatives. While Decision Tree and KNN provided moderate performance, Naive Bayes showed lower accuracy but maintained simplicity and computational efficiency. Overall, the choice of the best model depends on the specific requirements of the application, such as the importance of precision, recall, or computational efficiency.

Reference:

1. El-Sofany H, Bouallegue B, El-Latif YM. A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method. Scientific Reports. 2024 Oct 7;14(1):23277..
2. Miao KH, Miao JH. Coronary heart disease diagnosis using deep neural networks. International journal of advanced computer science and applications. 2018;9(10).
3. Sa S. Intelligent heart disease prediction system using data mining techniques. Int J Healthcare Biomed Res. 2013 Apr;1:94-101.
4. Dangare C, Apte S. A data mining approach for prediction of heart disease using neural networks. International Journal of Computer Engineering and Technology (IJCET). 2012 Oct 14;3(3).
5. Zhang QT, Hu DY, Yang JG, Zhang SY, Zhang XQ, Liu SS. Public knowledge of heart attack symptoms in Beijing residents. Chinese medical journal. 2007 Sep 1;120(18):1587-91
6. Gender-based Differences. Assiut Scientific Nursing Journal. 2021 Jun 1;9(25.0):37- 44.