Koneru Lakshmaiah Education Foundation, Hyderabad(Aziz Nagar,Off Campus)

Heart Disease Detection using Machine Learning

Ashwitha, Nikitha,Sahitya, Rakshitha

Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad-500075, Telangana, India.

Abstract

In various fields around the world, machine learning is used. There are no exceptions in the healthcare sector. Diagnosis and prediction of heart related diseases requires more precision, perfection, and correctness because a little mistake can cause fatigue problem or death of the person, there are numerous death cases related to heart and their counting is increasing exponentially day by day.

To deal with the problem there is an essential need for a prediction system for awareness about diseases. Machine learning is the branch of Artificial Intelligence(AI), it provides prestigious support in predicting any kind of event which take training from natural events We are attempting to use machine learning algorithms to predict potential heart conditions in humans. In this project, we compare the performance of various classifiers, Including Extra Three classifier, SVC, KNN and Random Forest Classifier. We also propose an ensemble classifier that performs hybrid classification by combining the best features of both strong and weak classifiers because it can use a large number of training and testing samples. This work presents several machine learning approaches for predicting heart diseases, using data of major health factors from patients.

Introduction

Human's most vital organ, the heart helps to create and circulate blood throughout the body and requires special attention. A person can fight off any sickness but not with the heart problem, it takes several days months to recover. The essential pumping organ of the human body requires proper care. Cardiovascular disease (CVD) remains a leading global health concern, emphasizing the need for advanced predictive tools to aid in early detection and prevention. In response to this challenge, machine learning has emerged as a promising avenue for enhancing risk assessment models.

The major challenge faced in the world of medical sciences today is the provision of quality service and efficient and accurate prediction. The later problem can be solved by automation with the help of Data Mining and Machine Learning. Data mining is defined as a process used to extract usable data from a large set of raw data. It implies analyzing patterns in large batches of data by making use of various software. It also involves effective data collection and warehousing coupled with computer processing. Machine Learning (ML) which is subfield of data mining that deals with large scale well-formatted dataset efficiently. In the medical field, machine learning can be used for diagnosis, detection and prediction of various diseases.

Therefore, detection of cardiac abnormalities at the early stage and tools for the prediction of heart diseases can save a lot of life and help doctors to design an effective treatment plan which ultimately reduces the mortality rate due to cardiovascular diseases. Due to the development of advance healthcare systems, lots of patient data are nowadays available which can be used for designing predictive models for Cardiovascular diseases. This integration of machine learning into heart disease prediction is not only transformative in terms of accuracy but also offers the potential for early intervention and preventive strategies.

Researchers are using machine learning algorithms to accurately diagnose heart diseases based on electronic health data. One such approach is the use of supervised learning algorithms such as support vector machine(SVM), Random Forest(RF) to build prediction models.By combining the power of data science and healthcare, we can revolutionize heart disease diagnosis and improve patient outcomes.

Methods and Materials

Loading Dataset: This dataset consists of 11 features and a target variable. It has 6 nominal and 5 numeric variables.
These features include age, gender, systolic blood pressure, and diastolic blood pressure. The target class, "cardio," indicates whether a patient has cardiovascular disease (represented as 1) or is healthy (represented as 0).

Data Cleaning and Exploration:

1. Handle missing data: Address any missing values in your dataset.
2. Explore the data: Understand the distribution of features, identify outliers, and perform statistical analysis

Data Preprocessing:

1. Encode categorical variables: Convert categorical variables into a numerical format (e.g., one-hot encodi
2. Scale features: Normalize numerical features to ensure they have similar scales.

Feature Selection:

1. Identify important features. Used feature of correlation analysis
2. Consider features with higher absolute correlation values as potentially more important.

Split the Data:

1. Among the 1179 patient records. The training is 937 and 235 is for testing set.

Train and Evaluate the Model:

1. Use the training data to teach the model to recognize patterns.
2. Use the testing data to assess the model's accuracy, precision, recall, F1 score, and other relevant metri

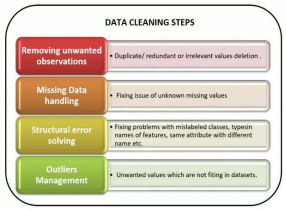


Figure 1. Datacleaning steps.

Results

	Model	Accuracy	Precision	Sensitivity	Specificity	F1 Score	ROC	Log_Loss	mathew_corrcoef
0	Random Forest	0.893617	0.865672	0.943089	0.839286	0.902724	0.891188	3.834431	0.789339

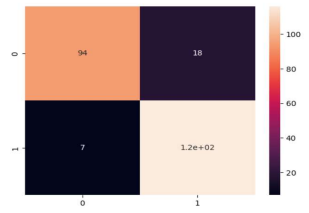


Figure 1: Model accuracy

	Model	Accuracy	Precision	Sensitivity	Specificity
0	KNN	0.804255	0.781022	0.869919	0.732143
1	SVC	0.812766	0.784173	0.886179	0.732143
2	EXtra tree classifier	0.902128	0.884615	0.934959	0.866071

	F1 Score	ROC	Log_Loss	mathew_corrcoef
0	0.823077	0.801031	7.055353	0.609859
1	0.832061	0.809161	6.748599	0.628251
2	0.909091	0.900515	3.527677	0.804719

By Doing Feature Selection and Also Outliers Detection, the dataset is normalized and the feature selection is done and also the outliers are. The correlation comparison can be seen in Figure . The accuracy of Extra tree classifier got the better accuracy 90% than other classifier algorithms.

Figure 1. ROC AUC curve.

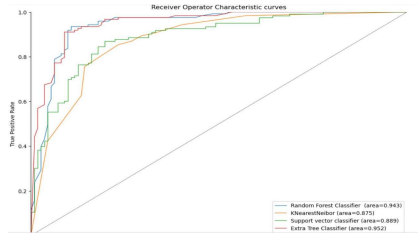
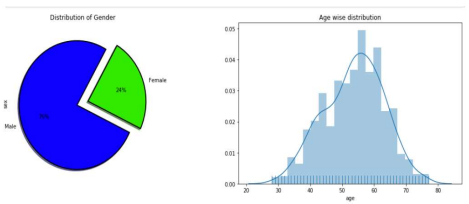


Chart 1: Checking Gender and Age-wise distribution.



Discussion

The article presents a machine learning approach for predicting heart disease using a dataset of health factors from patients.
In this project, we aimed to develop a predictive model for cardiovascular disease using a dataset consisting of 11 features. The age group between 50-60 are mostly affected by heart disease.
The importance of considering the broader context of the problem, interpreting evaluation metrics, and identifying avenues for future improvements.
It performs data preprocessing and feature selection steps before building the models. The models are evaluated based on the accuracy, precision, recall, and F1-score.
•The reports that the Extra tree classifier model performed best with 90.23% accuracy. It also compares the results with other existing systems
This project serves as a foundation for further research and applications in the domain of medical diagnostics using machine learning.

Conclusions

In conclusion, the project successfully developed a machine learning model for predicting cardiovascular disease based on a set of features. The Extra Trees Classifier, in conjunction with feature selection techniques, demonstrated strong predictive capabilities..We have also interpreted second best performing algo i.e., random forest algorithm

The top 5 most contribution features are:

Max heart Rate achieved, **Cholestrol**, **st_depression**, **Age**

exercise_induced_angina

Understanding these influential features is essential for interpreting the model's predictions and gaining insights into the factors that contribute to the risk of cardiovascular disease in the dataset you analyzed

.

Future Directions

1.Continuous refinement and validation of the model with additional data and expert input will enhance its clinical applicability.
2.Ethical considerations, including privacy safeguards and bias mitigation, should be prioritized for responsible deployment in healthcare settings.
3.Collaborations between data scientists and healthcare professionals are essential to ensure the model's accuracy and relevance in real-world medical scenarios.

Contact Information

Ashwitha Reddy (2010040043)
Sai Nikitha (2010040082)
Sahitya Gouri (2010040094)
Rakshitha (2010040123)

References

1. R. Katarya and P. Srinivas, "Predicting Heart Disease at Early Stages using Machine Learning: A Survey," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 302-305, doi: 10.1109/ICESC48915.2020.9155586
2. A Kumari and A. K. Mehta, "A Novel Approach for Prediction of Heart Disease using Machine Learning Algorithms," 2021 Asian Conference on Innovation in Technology (ASIANCON), PUNE, India, 2021, pp. 1-5, doi: 10.1109/ASIANCON51346.2021.9544544.
3. S. Dhar, K. Roy, T. Dey, P. Datta and A. Biswas, "A Hybrid Machine Learning Approach for Prediction of Heart Diseases," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018, pp. 1-6, doi: 10.1109/CCAA.2018.8777531
4. R. Buettner and M. Schunter, "Efficient machine learning based detection of heart disease," 2019 IEEE International Conference on E-health Networking, Application & Services (HealthCom), Bogota, Colombia, 2019, pp. 1-6, doi: 10.1109/HealthCom46333.2019.9009429
5. R. Aggarwal and S. Kumar, "MLPPCA: Heart Disease Detection using Machine learning," 2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC), Solan, Himachal Pradesh, India, 2022, pp. 457-461, doi: 10.1109/PDGC56933.2022.10053266.
6. A.Ed-Daoudy and K. Maalmi, "Real-time machine learning for early detection of heart disease using big data approach," 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), Fez, Morocco, 2019, pp. 1-5, doi: 10.1109/WITS.2019.8723830

Acknowledgements

We pay our grateful acknowledgment & sincere thanks to our Head of the Department **Dr. M. Goutham,** Associate Professor & Head of the Department, Department of ECE, K L University Hyderabad.for his exemplary guidance, monitoring and constant encouragementthroughout the course of the project. We pay our sincere thanks to our Project Coordinator **Dr. Ravi Boda** who has been constantly inculcating logistic support and has given constant encouragement during our project. And I thankful to my group members for completing this project.