International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

EARLY DETECTION OF CARDIOVASCULARDISEASE USING ARTIFICIAL INTELLIGENCE-A SURVEY PAPER

1st Riddhi Desai
Information Technology
Goa College of Engineering
riddhidesai1234567@gmail.com

2nd Dr. Aisha Fernandes
Associate Professor
Information Technology
Goa College of Engineering

***Abstract*—The early detection of cardiovascular disease (CVD) is crucial for effective prevention and management. In recent years, advancements in machine learning have provided new opportunities for the development of AI-based solutions for early cardiovascular disease detection.. Results from various studies suggest that machine learning algorithms have the potential to outperform traditional methods for early cardiovascular disease detection. However, several challenges remain, including the need for large and diverse datasets, the development of standardized evaluation metrics, and the integration of AI into clinical practice. Despite these challenges, machine learning holds great promise for the early detection of cardiovascular disease and is expected to play a significant role in the prevention and management of this disease in the future.**

***Index Terms*—cardiovascular disease detection, evaluation metrics, AI**

I. Introduction

Heart failure(HF) is the failure of heart to pump sufficient amount of blood to meet the conditions of the body. Narrowing or blockage of the coronary arteries is considered to be the main cause of heart failure. They are responsible for carrying blood to the heart itself. The common symptoms include briefness of breath, blown feet and weakness of the body. There are different trouble factors that lead to heart complaint that have been reported. These trouble factors are divided into two groups. The first group includes case's family history, sex and age. These trouble factors can't be changed. Still, the alternate group includes trouble factors that are related to life style of the case. Hence, these factors can be changed e.g., high cholesterol level, smoking, physical inactivity and high blood pressure Due to lack of respectable individual tools and medical experts especially in developing countries, the opinion of heart complaint is worse. likewise, conventional styles for opinion of HF complaint are predicated on different tests recommended by croakers , analysis of applicable symptoms and assessing cases ' medical history. Among the tests, angiography is considered a pivotal tool for determination of heart failure. It's a type of decision used to confirm heart complaint cases and is regarded as the best system for the determination of heart failure. Still, it also has some limitations like the high cost and side effects associated with it. Also, it requires high expertise of specialized know- how. Hence, to resolve these issues, expert systems needs to be designed

Over the years, researchers find out that machine learning algorithms perform truly well in assaying medical data sets. These data sets will be directly given to the algorithms, and these algorithms will perform according to their nature, and will give some outputs. There are some common attributes which are used to predict the heart conditions are- Gender- Age. - Resting blood pressure. - Types of chest pain. - Serum cholesterol in mg/ dl. - Fasting blood sugar. - ECG results. - Heart rate. - Thalassemia. -Old peak. The ECG signals give information about the electrical exertion of the heart, which can be used to identify cardiovascular conditions analogous as arrhythmias, myocardial infarction, and heart failure. The algorithms can be used to dissect ECG signals and extract features that are reflective of cardiovascular complaint. These features can be used to develop prophetic models that can identify individualities at high trouble of developing the complaint. Echocardiogram videos give visual information about the structure and function of the heart. These videos can be used to estimate parameters analogous as left ventricular ejection bit, which is an important indicator of cardiac function. ML algorithms can be used to anatomize echocardiogram videos and extract features that are reflective of cardiovascular complaint. These features can be used to develop prophetic models that can identify individualities at high trouble of developing the complaint.

II. Literature Review

A. In 2017, S. Xu, Z. Zhang et al. proposed a method for predicting cardiovascular risk using a combination of feature selection and machine learning. The authors proposed a framework that uses the Correlation-based Feature Selection (CFS) method to evaluate the importance of features and a Random Forest classification algorithm to make predictions[1].

B. In 2018 Ambekar et al. evaluates their method on a dataset of patients with cardiovascular disease and healthy individuals. The results show that the CNN-based model outperforms traditional machine learning algorithms in terms of accuracy, sensitivity, and specificity. The authors also perform

a sensitivity analysis to show that their method is robust to different parameter settings.

In conclusion, the authors demonstrate that their proposed CNN-based model is a promising method for predicting cardiovascular disease risk. The use of a CNN provides a powerful tool for extracting important information from large and complex datasets and making accurate predictions. The results of this study highlight the potential of using deep learning algorithms for the early detection and prevention of cardiovascular disease[2].

C. In 2020,Katarya et al. The authors discuss the strengths and weaknesses of each algorithm and provide insights into the challenges and opportunities for future research in this area. They also provide a comprehensive analysis of the most important features used for heart disease prediction, including demographic, lifestyle, and clinical features.

In conclusion, the authors provide a comprehensive overview of the current state of the field in using machine learning for early detection of heart disease. The authors highlight the potential of machine learning for improving the accuracy and speed of heart disease prediction and suggest areas for future research in this area. The results of this study provide valuable insights into the use of machine learning for improving the early detection and prevention of heart disease[4].

D. In 2020, Motarwar et al. proposed a cognitive approach to heart disease prediction that leverages machine learning algorithms to analyze demographic, lifestyle, and clinical data.

The authors evaluate the performance of their method using a large dataset of patients with heart disease and healthy individuals. The results show that their cognitive approach outperforms traditional machine learning algorithms in terms of accuracy, sensitivity, and specificity. The authors also perform a sensitivity analysis to show that their method is robust to different parameter settings.

In conclusion, the authors demonstrate the potential of their cognitive approach for improving the early detection and prevention of heart disease. The use of machine learning algorithms provides a powerful tool for analyzing large and complex datasets and making accurate predictions. The results of this study highlight the potential of using cognitive approaches for improving the accuracy and speed of heart disease prediction[5].

III. MACHINE LEARNING ALGORITHMS

3.MACHINE LEARNING ALGORITHMS There are several approaches that can be used for cardiovascular complaint discovery using machine learning on the Kaggle cardiovascular ailment dataset of 70k analyses. Some of the most common approaches are Logistic Retrogression :Logistic retrogression is a popular algorithm used for binary classification(prognosticating one of two classes) or multiclass classification(prognosticating among further than two classes). It models the probability of the target variable belonging to a particular class grounded on the input features. Logistic retrogression assumes a direct relationship between the features and uses a logistic function to collude the prognostications into a probability range between 0 and 1.

K-Nearest Neighbors(KNN): KNN is anon-parametric algorithm used for both classification and regression tasks. It classifies data points by chancing the k- nearest neighbors grounded on a distance metric(e.g., Euclidean distance) and assigning the class that appears most constantly among the neighbors. KNN makes prognostications grounded on the supposition that analogous data points tend to have similar labels.

Gaussian Naive Bayes(Gaussian NB): Naive Bayes is a probabilistic algorithm grounded on Bayes' theorem. Gaussian NB assumes that the features follow a Gaussian distribution. It calculates the probability of a data point belonging to a particular class by assuming independence among the features and using the probability density function of the Gaussian distribution.

Random Forest :Random Forest is an ensemble algorithm that combines multiple decision trees to make predictions. It generates a large number of decision trees by bootstrapping the data and opting a arbitrary subset of features for each tree. Random Forest summations the prognostications of individual trees to produce the final vaticination, reducing the threat of overfitting and perfecting accuracy.

Decision Tree :Decision trees are adaptable algorithms that produce a tree- suchlike model of decisions and their possible consequences. Each node in the tree represents a point, and the branches represent the possible values of that point. Decision trees make prognostications by covering the tree from the root to a leaf node grounded on the point values, and assigning the most frequent class among the training samples in that leaf node as the prediction.

Extra Trees: Extra Trees, also known as Extremely Randomized Trees, is another ensemble algorithm analogous to Random Forest. It generates a large number of decision trees, like Random Forest, but rather of opting the best split at each node, Extra Trees selects splits aimlessly. This randomness can lead to advanced diversity among the trees and potentially ameliorate generalization.

AdaBoost Classifier: AdaBoost(Adaptive Boosting) is an ensemble algorithm that combines multiple weak classifiers into a strong classifier. It iteratively trains the weak classifiers by giving further weight to misclassified samples in each replication. The final vaticination is made by combining the predictions of all weak classifiers, ladened by their performance during training. grade Boosting Classifier grade Boosting is a boosting algorithm that successionally builds a series of weak predictive models(generally decision trees) to produce a strong prophetic model. Each subsequent model is trained to correct the crimes made by the former models. grade Boosting uses grade descent to minimize a loss function and ameliorate the model's performance. These algorithms offer different approaches to working colorful machine literacy problems, and the choice depends on the specific conditions of the task, the characteristics of the data, and the asked outgrowth.

IV. EXPERIMENTS

1) *Implementation*

: Data collection and preparation: The first step is to obtain the cardiovascular disease dataset from Kaggle and clean the data to make it ready for analysis. This includes handling missing values, converting categorical variables into numerical values, and normalizing the data. Feature selection: After cleaning the data, the next step is to select the features that will be used to build the model. Features can be selected based on their correlation with the target variable, or by using feature selection techniques such as Recursive Feature Elimination (RFE) or Feature Importance. Model building: Once the features have been selected, the next step is to build a machine learning model. There are several algorithms that can be used for this task, including Logistic Regression, Random Forest, Extra trees, Xgboost, Adaboost classifier, gradient boosting. Model evaluation: After building the model, the next step is to evaluate its performance. This can be done by using metrics such as accuracy, precision, recall, and F1- score. Hyperparameter tuning: If the model's performance is not satisfactory, hyperparameter tuning can be performed to improve the model's performance. This involves adjusting the parameters of the model to optimize its performance.

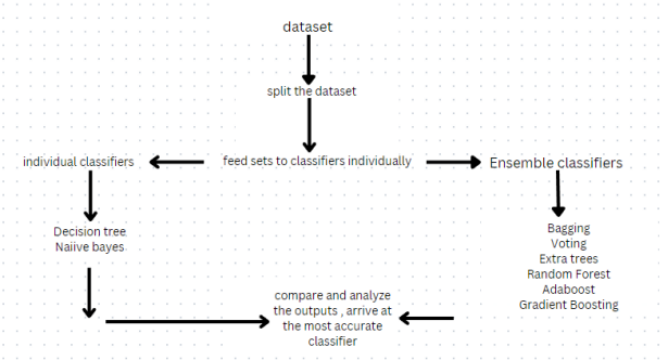

Fig-1:individual and ensemble classifiers

	Model	Training Accuracy %	Testing Accuracy %
0	XGBoost	93.27	84.64
1	LogisticRegression	87.12	85.00
2	K-nearest neighbors	90.40	85.83
3	GaussianNB	88.85	87.35
4	Random Forest Classifier	89.45	87.75
5	Decision Tree	89.66	87.72
6	ExtraTrees	90.54	87.75
7	AdaBoostClassifier	90.26	87.94
8	GradientBoostingClassifier	90.37	88.00

Fig -2:different model with their accuracies

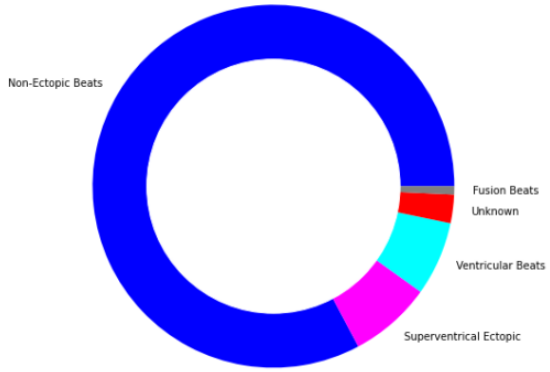

Fig-3:distribution of beats

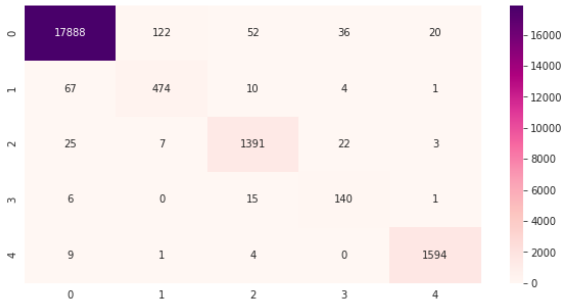

Fig -4:correlation matrix

V. CONCLUSION

Heart ailment is a veritably critical issue in the present growing world. So, there is a need for an automated system to predict heart disease at earlier stages, so that it will be useful for the physician to diagnose the patients efficiently, and it will be useful to the people also because they can track their health issues by using this automated system.Feature selection and prediction, these two are essential for every automated system. By choosing features efficiently, we can achieve better performance. In future, we can add further conditions and prognosticate the threat which patient suffers from specific disease.

VI. FUTURE SCOPE

Real-time data from different hospitals may be collected for detecting heart disease patients and compute the effectiveness of classifiers for more consistent diagnosis of heart disease patients.

VII. REFERENCES

1. S. Xu, Z. Zhang, D. Wang, J. Hu, X. Duan and T. Zhu, "Cardiovascular risk prediction method based on CFS subset evaluation and random forest classification framework," 2017.

2. S. Ambekar and R. Phalnikar, ”Disease Risk Prediction by Using Convolutional Neural Network,” 2018.
3. L. Ali et al., ”An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure,” in IEEE Access, vol. 7, pp. 54007-54014, 2019.
4. 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), 2020.
5. P. Motarwar, A. Duraphe, G. Suganya and M. Premalatha, ”Cognitive Approach for Heart Disease Prediction using Machine Learning,” 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020.
6. Y. Liu, M. Zhang, Z. Fan and Y. Chen, ”Heart disease prediction based on random forest and LSTM,” 2020 2nd International Conference on Information Technology and Computer Application (ITCA), 2020.
7. L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan, “An Automated Diagnostic System for Heart Disease Prediction Based on 2 Statistical Model and Optimally Configured Deep Neural Network,” IEEE Access, vol. 7, pp. 34938–34945, 2019.
8. Purushottam, K. Saxena, and R. Sharma, “Efficient Heart Disease Prediction System,” Procedia Comput. Sci., vol. 85, pp. 962–969, 2016.
9. L. Ali et al., “An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure,” IEEE Access, vol. 7, pp. 54007–54014, 2019.
10. K. Deepika and S. Seema, “Predictive analytics to prevent and control chronic diseases,” Proc. 2016 2nd Int. Conf. Appl. Theor. Comput. Commun. Technol. iCATccT 2016, no. January 2016, pp. 381–386, 2017.
11. Md. Ashraful Alam,M.Shakir ,“Early detection of coronary artery blockage using image processing: segmentation, quantification, identification of degree of blockage and risk factors of heart attack”,2017