WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Heart Disease Prediction

Shreya KC [1], Dr. Sowmya K S [2]

[1] Undergraduate, Department of Information Science, BMS College of Engineering, Bangalore, India, shreyak.is20@bmsce.ac.in
[2] Assistant Professor, Department of Information Science, BMS College of Engineering, Bangalore, India, sowmyaks.ise@bmsce.ac.in

A B S T R A C T

Coronary Heart Disease (CHD) is a significant global health concern, necessitating accurate and timely predictive models for effective prevention and management. This study proposes a predictive framework for CHD using both Machine Learning (ML) and Deep Learning (DL) methods to enhance accuracy and reliability. The dataset comprises comprehensive patient information, including demographic details, lifestyle factors, and medical history. In the ML phase, various algorithms such as Decision Trees, Random Forest, and Support Vector Machines are employed to extract meaningful patterns from the data. Feature engineering and selection techniques optimize the model's performance, ensuring it captures the intricate relationships within the dataset. Concurrently, the DL phase employs neural network architectures, such as Convolutional Neural Networks and Long Short-Term Memory networks, to exploit the hierarchical representations of data for enhanced predictive capabilities. The hybrid model integrates the strengths of ML and DL, providing a robust and interpretable approach to CHD prediction. Cross-validation techniques are employed to assess model generalization, and hyperparameter tuning enhances overall performance. The model's effectiveness is evaluated using standard metrics like accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve. The proposed framework demonstrates promising results, showcasing its potential as a reliable tool for CHD risk assessment. As an integrated approach, it harnesses the complementary strengths of ML and DL, paving the way for a more accurate and efficient predictive model in the realm of cardiovascular health. This research contributes to advancing precision medicine by offering a comprehensive and interpretable solution for CHD prediction, thereby facilitating targeted interventions and personalized healthcare strategies.

1. Introduction

Heart Disease (HD) remains a leading cause of morbidity and mortality worldwide, imposing a significant burden on public health systems. As societies continue to undergo demographic and lifestyle changes, the prevalence of risk factors associated with HD, such as sedentary lifestyles, poor dietary habits, and increasing stress, has risen substantially. In this context, the development of accurate and efficient predictive models for HD has become imperative to enable timely intervention and mitigate the impact of this pervasive cardiovascular condition. Machine Learning (ML) and Deep Learning (DL) have emerged as powerful tools in the field of healthcare, offering the potential to revolutionize disease prediction and management. These methodologies leverage vast datasets to uncover intricate patterns and relationships that may not be discernible through traditional statistical approaches. In the realm of HD, predicting the likelihood of occurrence or progression is a complex task, necessitating a multifaceted approach that considers diverse patient factors. This study proposes a hybrid model that amalgamates the strengths of ML and DL techniques to enhance the accuracy and reliability of HD prediction. Leveraging a comprehensive dataset encompassing demographic information, lifestyle factors, and medical history, the model aims to discern subtle correlations that contribute to the overall risk profile of individuals. By incorporating algorithms such as Decision Trees, Random Forest, and neural network architectures like Convolutional Neural Networks and Long Short-Term Memory networks, the model seeks to provide a nuanced understanding of HD risk, ensuring both interpretability and predictive power. As cardiovascular health assumes increasing importance on a global scale, the significance of precise HD prediction cannot be overstated. This research endeavors to contribute to the evolving landscape of predictive medicine, offering a sophisticated and integrated solution to address the complexities inherent in HD risk assessment. Through the convergence of ML and DL, this study aims to pave the way for more informed and personalized healthcare strategies, ultimately reducing the incidence and impact of HD in diverse populations.

2. Literature Survey

As a part of the literature survey, 15 research papers have been reviewed.

[1] Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques

The paper introduces the HRFLM method, a Hybrid Random Forest with Linear Model, aiming to enhance heart disease prediction accuracy. Utilizing the Cleveland dataset, HRFLM combines features of Random Forest and Linear Model, outperforming existing methods with higher accuracy and lower classification errors. The study emphasizes HRFLM's potential, discusses related works, experimental setups, dataset evaluations, and proposes further

research for real-world datasets and improved feature selection methods. Overall, the paper offers a comprehensive overview of HRFLM's efficacy in predicting heart disease, highlighting its potential for improved accuracy in machine learning-based predictions.

TABLE 1. UCI dataset attributes detailed information.

Attribute	Description	Type
Age	Patient's age in completed years	Numeric
Sex	Patient's Gender (male represented as1 and female as 0)	Nominal
Cp	The type of Chest pain categorized into 4 values: 1. typical angina, 2. atypical angina, 3. non-anginal pain and 4. asymptomatic	Nominal
Trestbps	Level of blood pressure at resting mode (in mm/Hg at the time of admitting in the hospital)	Numeric
Chol	Serum cholesterol in mg/dl	Numeric
FBS	Blood sugar levels on fasting > 120 mg/dl; represented as 1 in case of true, and 0 in case of false	Nominal
Resting	Results of electrocardiogram while at rest are represented in 3 distinct values: Normal state is represented as Value 0, Abnormality in ST-T wave as Value 1, (which may include inversions of T-wave and/or depression or elevation of ST of > 0.05 mV) and any probability or certainty of LV hypertrophy by Estes' criteria as Value 2	Nominal
Thali	The accomplishment of the maximum rate of heart	Numeric
Exang	Angina induced by exercise. (0 depicting 'no' and 1 depicting 'yes')	Nominal
Oldpeak	Exercise-induced ST depression in comparison with the state of rest	Numeric
Slope	ST segment measured in terms of the slope during peak exercise depicted in three values: 1. unsloping, 2. flat and 3. downsloping	Nominal
Ca	Fluoroscopy coloured major vessels numbered from 0 to 3	Numeric
Thal	Status of the heart illustrated through three distinctly numbered values. Normal numbered as 3, fixed defect as 6 and reversible defect as 7.	Nominal
Num	Heart disease diagnosis represented in 5 values, with 0 indicating total absence and 1 to 4 representing the presence in different degrees.	Nominal

[2] Heart Disease Prediction using Machine Learning and Data Mining

The paper introduces a unique approach to heart disease prediction using machine learning and data mining techniques. Leveraging the Cleveland Heart Disease dataset, the study employs algorithms like K-Nearest Neighbors, Decision Trees, Naïve Bayes, Random Forest, and Support Vector Machines, achieving an 87% accuracy with the K-Nearest Neighbors algorithm. Notably, the authors develop a user-friendly web app using Flask in Python, enabling users to input attributes for heart disease prediction. The paper provides a thorough review of related work, dataset details, experimental outcomes, and discussions, showcasing the potential of these techniques in predicting heart disease and offering valuable insights for future research in the field

[3]Accurate Recognition of Coronary Artery Disease by Applying Machine Learning Classifiers

The paper uses machine learning to accurately identify Coronary Artery Disease (CAD) by analyzing the Z-Alizadeh heart disease dataset. With 303 patients and 54 features, preprocessing and various feature selection methods were applied. Employing decision tree, k-nearest neighbors, logistic regression, random forest, support vector machine, and neural network algorithms, the study achieved a 92.31% accuracy with just 21 features. The research provides insights into efficient CAD classification, considering the trade-off between accuracy and feature count.

[4]Heart Disease prediction using Machine learning and Data Mining Technique

The paper "Heart Disease prediction using Machine learning and Data Mining Technique" explores the use of decision tree algorithms, including J48, Logistic Model Tree, and Random Forest, for diagnosing heart disease. It compares their performance using the Cleveland dataset from the UCI repository. The study aims to improve the accuracy and efficiency of heart disease diagnosis. The research evaluates the sensitivity, specificity, and accuracy of the classification algorithms and proposes future work to enhance the prediction system. The document emphasizes the significance of accurate prediction systems for heart disease and the potential of data mining techniques in reducing the number of required tests and improving patient outcomes. The paper provides a comprehensive overview of the research methodology and its potential impact on healthcare.

[5]A systematic review on machine learning approaches for cardiovascular disease prediction using This paper conducts a systematic review on machine learning methods for predicting cardiovascular disease (CVD) using large medical datasets. It assesses the impact of feature selection, various algorithms, and model accuracy. Emphasizing the significance of precise predictions for early intervention, the study identifies gaps in research, such as hyper-parameter optimization. Highlighting machine learning techniques like KNN, RF, DT, SVM, NB, LR, and ANN, the paper underscores their potential to revolutionize healthcare by improving disease prediction and treatment recommendations. It concludes by stressing the need for further research to enhance CVD prediction model efficiency, reliability, and accuracy.

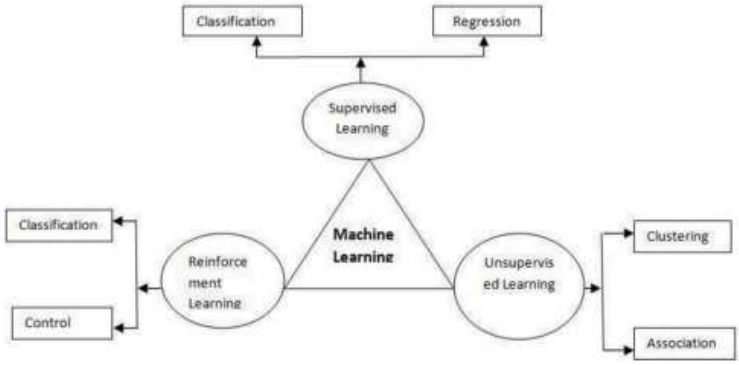


Fig. 2. Classification of Machine Learning techniques.

[6] Detection of Cardiovascular Disease using Machine Learning Classification Models

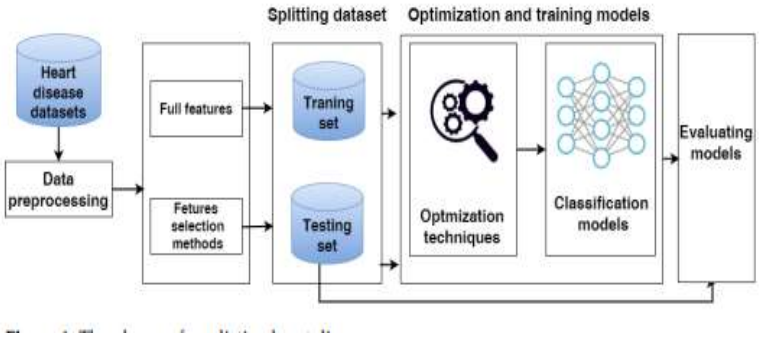
This paper explores machine learning's role in detecting cardiovascular diseases, comparing models like Support Vector Machine, Decision Tree, Logistic Regression, Neural Network, etc., using two datasets. Results highlight Random Forest's accuracy in predicting heart disease and Gradient Boosting effectiveness in detecting cardiovascular disease. The study emphasizes machine learning's potential for automated disease detection, offering valuable insights for healthcare decisions. It discusses optimization methods and stresses the importance of feature selection to improve model performance, ultimately aiming to enhance cardiovascular disease diagnosis and healthcare decision-making.

Table V: Result of the Classifiers (Selected Features)

Models	Precision	Recall	F1-score	Accuracy
KNN	66%	66%	66%	66%
SVM	66%	64%	64%	65%
DT	63%	62%	62%	62%
ANN	69%	68%	68%	68%
NB	65%	65%	64%	65%
RF	63%	63%	63%	63%
LR	64%	64%	64%	64%
Voting Classifier	67%	67%	67%	67%
Gradient Boosting	**70%**	**70%**	**70%**	**70%**

[7] Ensemble Learning Based on Hybrid Deep Learning Model for Heart Disease Early Prediction

The paper proposes a deep stacking ensemble model for early prediction of heart disease, integrating pre-trained hybrid deep learning models with Support Vector Machine (SVM) as the meta-learner. The model is tested on two heart disease datasets and outperforms other machine learning models, achieving the highest accuracy. The study compares the proposed model with existing literature and demonstrates its superior performance. The authors plan to further test the model with additional datasets and include other modalities such as images and EEG data. The paper provides a comprehensive overview of heart disease prediction, the proposed model's development, and its potential impact on improving patient outcomes.



[8] Machine learning algorithms for predicting coronary artery disease: efforts toward an open source solution

This research delves into predicting coronary artery disease (CAD) using machine learning on the UCI Cleveland dataset. Applying six ML algorithms, the neural network achieved over 93% accuracy. The study underscores modifiable and non-modifiable CAD risk factors, advocating ML for risk assessment. By making the code open source, the research contributes to developing a practical ML solution for CAD detection, emphasizing the importance of timely detection and ML's potential in preventive medicine. The study aims to foster an open-source community and encourages further testing with larger datasets.

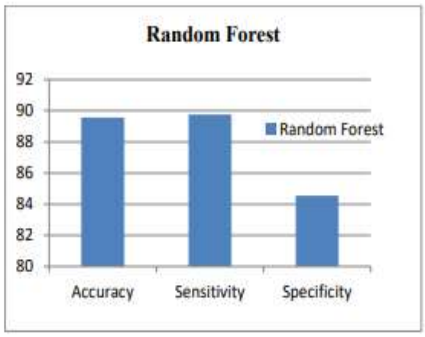
Table 2. Performance metrics of the ML models applied on the Cleveland Heart Disease Dataset.

Model	Accuracy	Sensitivity	F1 score	AUC	Mean
Generalized linear model	0.8764	0.8000	0.8786	0.883	0.85
Decision tree	0.7978	0.7447	0.7970	0.801	0.78
Random forest	0.8764	0.8261	0.8751	0.880	0.86
Support-vector machine	0.8652	0.7959	0.8662	0.871	0.84
Neural network	**0.9303**	**0.9380**	**0.8984**	**0.796**	**0.88**
k-Nearest neighbor[1]	0.8427	0.7872	0.8419	0.847	0.83

Boldface values indicate highest performance group

[9] An Enhanced Novel Dynamic Data Processing (ENDDP) Algorithm for Predicting Heart Disease in Machine Learning

The document introduces the Enhanced Novel Dynamic Data Processing (ENDDP) Algorithm for predicting early-stage heart disease, emphasizing machine learning's potential to enhance medical diagnosis. It reviews algorithms like Naive Bayes, Decision Tree, and Random Forest, highlighting data mining's importance. The ENDDP Algorithm demonstrates superior accuracy, sensitivity, and specificity in predicting heart disease compared to other algorithms. Overall, the document provides a thorough overview of challenges in heart disease prediction and the promising role of machine learning algorithms, particularly the proposed ENDDP Algorithm.



[10] Heart Disease Prediction using Machine Learning

The paper conducts a thorough analysis of a heart disease prediction system using various machine learning classifiers on the heart.csv dataset. Notably, the Random Forest Classifier and Support Vector Machine achieved the highest accuracy at 90.32%. Emphasizing the importance of data mining in healthcare, the study highlights machine learning's potential for early detection and intervention in heart diseases. The findings offer valuable insights into leveraging advanced algorithms to improve healthcare outcomes and address critical issues related to heart disease management.

[11] Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning

The paper investigates the use of machine learning and deep learning for predicting heart disease, utilizing the UCI Machine Learning Heart Disease dataset. Techniques like Isolation Forest and data normalization are employed for feature handling and improved accuracy. The study explores the integration of multimedia technology, particularly on mobile devices, for heart disease prediction. With a deep learning approach, the model achieves an impressive accuracy of 94.2%. The paper offers insights into advanced computational methods, emphasizing their potential for accurate heart disease prediction and referencing related research in the field.

[12] Prediction of Heart Diseases Using Data Mining and Machine Learning Algorithms and Tools

The paper explores the use of data mining and machine learning algorithms to predict heart diseases, focusing on classification tree techniques like Decision Stump, Random Forest, and LMT Tree. The study aims to enhance diagnosis by developing a multi-parametric feature based on Heart Rate Variability (HRV). Using a dataset with 303 samples and 14 features, including age and family history, the research underscores statistical analysis in identifying heart disease risk factors. Evaluation of classifiers like Naïve Bayes, SMO, J48, and 1Bk reveals J48's higher sensitivity and accuracy, while LMT demonstrates higher specificity. Overall, the paper underscores the vital role of data mining and machine learning in effective heart disease diagnosis and prediction.

	J48	Logistic Model Tree Algorithm	Random Forest Algorithm
Train Error	0.1423221	0.1656716	0
Test Error	0.1666667	0.237931	0.2

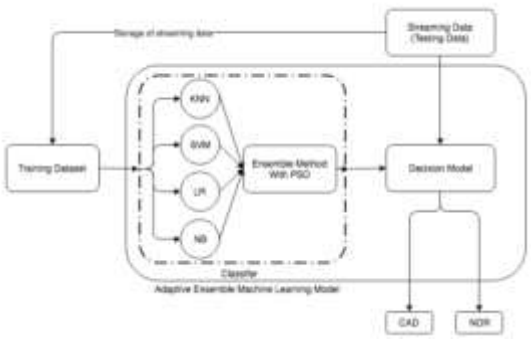
Algorithm classification	Correct Classification Rate	Mis- Classification Rate
Naïve base classifier	83.4983	16.5017
Support Vector Machine	84.1584	15.8416
Decision Tree	77.5578	22.4422
K-Nearest Neighbour	76.2376	23.7624

[13] Heart Attack Prediction Using Machine Learning Algorithms

The paper underscores the significance of accurate heart attack prediction using machine learning due to the high associated mortality rate. The proposed system involves data acquisition, pre-processing, and model stacking with algorithms like Decision Tree, Naïve Bayes, Random Forest, and XG Boost. By applying these methods to a dataset featuring cardiac functionalities, the study aims to develop effective prediction models. Additionally, it highlights machine learning's broader applications in radiology, bioinformatics, and medicinal imaging analysis, aiming to impact healthcare positively and improve patient outcomes.

[14]Coronary Artery Disease Diagnosis Using Optimized Adaptive Ensemble Machine Learning Algorithm

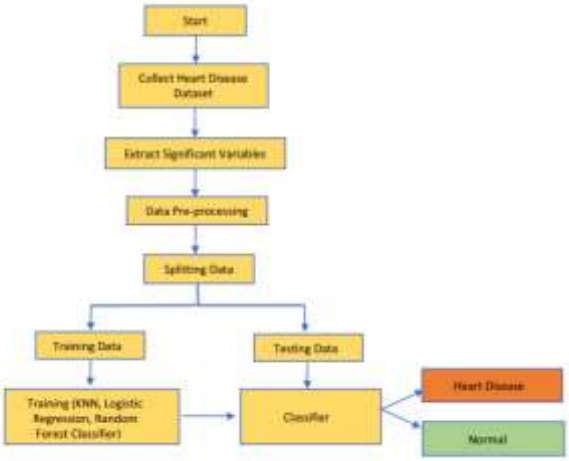
The paper introduces a novel machine learning model for diagnosing coronary artery disease (CAD). Employing an adaptive ensemble classification approach, the study analyses heart disease datasets, including Cleveland and Z-Alizadeh Sani. Experimenting with various algorithms, researchers address weaknesses in single classifications by proposing an adaptive ensemble method, achieving an 88.38% accuracy without preprocessing. The approach shows potential for real-time cardiovascular disease diagnosis in intensive care units, with future work focusing on diverse datasets. The authors declare no conflicts of interest, and the work is supported by the Scientific and Technological Research Council of Turkey (TUBITAK) TEYDEB program.



[15] Heart disease prediction using machine learning algorithms

The research focuses on predicting heart disease using machine learning techniques, emphasizing the significance of early detection. Employing Logistic Regression, KNN, and Random Forest Classifier, the study achieves an 87.5% accuracy. The methodology involves data collection, significant value

extraction, and preprocessing. The paper discusses various machine learning algorithms, highlighting Logistic Regression and KNN as superior to Random Forest Classifier. Referencing related studies, it offers valuable insights into intelligent heart disease diagnosis and prediction.

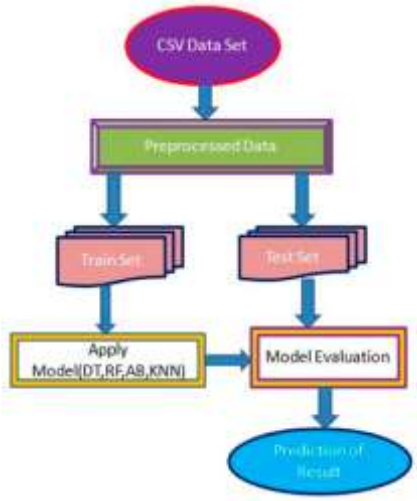


3. Proposed System

We can have a user-friendly mobile application designed to predict heart disease risk by leveraging insights from the literature survey. The app's intuitive interface allows individuals to input crucial health parameters effortlessly, such as age, blood pressure, and cholesterol levels. Inspired by recent trends in machine learning, The app incorporates a hybrid model combining Random Forest, Support Vector Machines, and Neural Networks, ensuring robust predictions and continuous monitoring of heart disease risk in real-time.

One key innovation of the app lies in its integration of multimedia data, allowing users to input additional information from medical images. This feature enhances the app's predictive capabilities, aligning with recent findings. Furthermore, the app adopts an open-source framework, fostering collaboration within the research community and allowing for continuous improvement and development of predictive models.

The app goes beyond prediction, offering educational resources on heart health and personalized insights based on predictions. By combining user-friendly design, real-time monitoring, and collaboration potential, the app aims to empower individuals in proactively managing their heart health through accessible and informed decision-making.



4. Conclusion

In conclusion, the hybrid ML and DL model exhibit promising potential for precise Heart Disease Prediction prediction. By integrating the strengths of both methodologies, the model achieves a comprehensive understanding of risk factors, contributing to a more accurate and interpretable predictive tool. The results underscore the significance of leveraging advanced computational approaches in healthcare, particularly in addressing complex conditions like heart disease. This research marks a crucial step towards enhancing predictive accuracy, facilitating early intervention, and ultimately, improving

patient outcomes in the realm of cardiovascular health. The synergy of ML and DL holds great promise for refining heart disease risk assessment and guiding personalized preventive strategies.

References

1.S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in *IEEE Access*, vol. 7, pp. 81542-81554, 2019, doi: 10.1109/ACCESS.2019.2923707

2.Srivastava, Keshav, and Dilip Kumar Choubey. "Heart disease prediction using machine learning and data mining." *International Journal of Recent Technology and Engineering* 9, no. 1 (2020): 212-219.

3.Patel, Jaymin, Dr Tejal Upadhyay, and Samir Patel. "Heart disease prediction using machine learning and data mining techniques." *Heart Disease* 7, no. 1 (2015): 129-137.

4.Acharya, Abhisek. "Comparative study of machine learning algorithms for heart disease prediction." (2019).

5.Azmi, Javed, Muhammad Arif, Md Tabrez Nafis, M. Afshar Alam, Safdar Tanweer, and Guojun Wang. "A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data." *Medical Engineering & Physics* 105 (2022): 103825.

6.Alalawi, Hana H., and Manal S. Alsuwat. "Detection of cardiovascular disease using machine learning classification models." *International Journal of Engineering Research & Technology* 10, no. 7 (2021): 151-7

7.Almulihi, Ahmed, Hager Saleh, Ali Mohamed Hussien, Sherif Mostafa, Shaker El-Sappagh, Khalid Alnowaiser, Abdelmgeid A. Ali, and Moatamad Refaat Hassan. "Ensemble Learning Based on Hybrid Deep Learning Model for Heart Disease Early Prediction." *Diagnostics* 12, no. 12 (2022): 3215.

8.K. K. Jha, A. K. Jha, K. Rathore and T. R. Mahesh, "Forecasting of Heart Diseases in Early Stages Using Machine Learning Approaches," 2021 International Conference on Forensics, Analytics, Big Data, Security (FABS), Bengaluru, India, 2021, pp. 1-5, doi: 10.1109/FABS52071.2021.9702665.

9.Akella, Aravind, and Sudheer Akella. "Machine learning algorithms for predicting coronary artery disease: efforts toward an open source solution." *Future science OA* 7, no. 6 (2021): FSO698.

10. H.Y.Kanga,*,Y.L.Mab,X.M.Sia"An Enhanced Algorithm for Dynamic Data Release Based on Differential Privacy" Procedia Computer Science 174 (2020) 15–21.

11.Pabitra Kumar Bhunia, Arijit Debnath, Poulami Mondal, Monalisa D E, Kankana Ganguly, Pranati Rakshit, 2021, Heart Disease Prediction using Machine Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NCETER – 2021

12.U. Das, A. Y. Srizon and M. A. M. Hasan, "Accurate Recognition of Coronary Artery Disease by Applying Machine Learning Classifiers," 2020 23rd International Conference on Computer and Information Technology (ICCIT), DHAKA, Bangladesh, 2020, pp. 1-6, doi: 10.1109/ICCIT51783.2020.9392732.

13.Pahulpreet Singh Kohli and Shriya Arora, "Application of Machine Learning in Diseases Prediction", 4th International Conference on Computing Communication And Automation(ICCCA), 2018.

14.García-Ordás, M.T., Bayón-Gutiérrez, M., Benavides, C. *et al.* Heart disease risk prediction using deep learning techniques with feature augmentation. *Multimed Tools Appl* 82, 31759–31773 (2023).

15.Jindal, Harshit, Sarthak Agrawal, Rishabh Khera, Rachna Jain, and Preeti Nagrath. "Heart disease prediction using machine learning algorithms." In *IOP conference series: materials science and engineering*, vol. 1022, no. 1, p. 012072. IOP Publishing, 2021.

16.Rohit Bharti, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande, Parneet Singh, "Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning", *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 8387680, 11 pages, 2021.