JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Review: Machine Learning Approach for Heart Disease Prediction

[1] Prof. Ashish Digambar Sawant, [2] Dr M.S. Ali

[1,] Assitant Professor at Prof Ram Meghe College of Engineering and Management Badnera .
[2] Research Guide at Prof Ram Meghe College of Engineering and Management Badnera.

Abstract: Machine learning, powered by vast healthcare data, helps diagnose heart diseases early, saving lives. This research investigates the use of machine learning approaches for accurate prediction of heart diseases. Heart disease remains the leading cause of death, such that nearly one-third of all deaths worldwide are estimated to be caused by heart-related conditions.[1] Coronary heart disease (CHD) /Acute coronary syndrome (ACS).[2] The mortality rate is expanding due to obesity, cholesterol, high blood pressure and usage of tobacco among the people.[3] An accurate prediction of heart disease is necessary for the early stage of treatment and overcoming the mortality rate.[4] Machine learning (ML) can be used to assist clinical decision-making. We developed a ML model for the prediction of 1-year mortality after heart transplantation (HT) in adults with congenital heart disease.[5] This study proposes a machine learning approach to Prediction of heart disease more accurately. Moreover, Accuracy, F1-Measure, Precision, and Recall are used to measure the performance of machine learning models.[4]Different types of supervised machine learning model can be applied on data set.[6] We wanted to design and develop machine learning model for Prediction of heart disease. In addition, a user-friendly web app and a user-friendly mobile app are built based on the most accurate model.[4]

IndexTerms -. Coronary heart disease (CHD), Machine learning (ML), supervised machine learning model, Accuracy, F1-Measure, Precision, and Recall etc.

I. INTRODUCTION

Heart is one of the most important organs in Human's body. In life, some changes may happen that may bring various diseases like, blood pressure, sugar, etc. Similarly, heart failure is also a dreadful disease. Heart failure is a serious condition.[7] Heart diseases are a significant global health concern, and early detection is crucial for effective treatment and prevention. Nowadays, one of the most important illnesses is a heart disease which causes most patients dead. The medical diagnosis of heart disease is quite difficult. This diagnosis is a challenging process that requires accuracy and efficiency. The chance of death will be decreased with early heart disease detection.[8] Hypertension describes elevated blood pressure, which significantly impacts cardiovascular diseases.[9] [10] Predicting heart disease has become one of the most

difficult medical jobs in recent years. Researchers looked at a variety of closely related traits to discover the most reliable predictors of these conditions.[8] Various supervised and unsupervised learning methods are proposed to better detect heart related diseases.[11] Artificial intelligence (AI) is a part of computer science that has the task of making computers more intelligent. Since the most basic requirement of intelligence is learning, hence came the subfield of AI that is called machine learning (ML). ML is one of the most rapidly evolving fields of AI which is used in many areas of life, primarily in the healthcare field. ML has a great value in the healthcare field since it is an intelligent tool to analyze data, and the medical field is rich with data.[12] Machine learning models have shown great promise in predicting heart diseases by analyzing various medical data. In this study, Machine Learning (ML) techniques are used to identify the presence of cardiac abnormalities. The proposed method predicts the chances of heart disease and classifies patient's risk level by using different ML algorithm techniques.[8], [13] Machine learning is closely related to computational statistics, which focus on using mathematical optimization to deliver methods, theory and application domains to solve medical, industry, social and business problems in the real world. It can be divided into two broad categories: supervised learning and unsupervised learning. In supervised learning, the algorithm builds a mathematical model from a set of data that contains both the inputs and the desired outputs. While in unsupervised learning, the algorithm builds a mathematical model from a set of data that contains only inputs and no desired output labels. In this study, the Classification and Regression Tree (CART) algorithm, a supervised machine learning method, has been employed to predict heart disease [1], [14], [15] Machine Learning is a branch of AI research [2] and has become a very popular aspect of data science. We use principal component analysis (PCA) to transform the higher dimensional feature space into a lower dimension subspace, and understand the relative importance of each input attributes.[16]The Machine Learning algorithms are designed to perform a large number of tasks such as prediction, classification, decision making etc. To learn the ML algorithms, training data is required. After the learning phase, a model is produced which is considered as an output of ML algorithm. This model is then tested and validated on a set of unseen real time test dataset. The final accuracy of the model is then compared with the actual value, which justify the overall correctness of predicted result.[17] [17]Hybrid model is created by combining support vector machine (SVM) approach and modified particle swarm optimization model.[12] In this study, we propose machine learning (ML) for risk factors analysis and survival prediction of Heart Failure (HF) patients using a survival dataset.[6] In this synopsis we discuss the rapid growth of machine learning approaches and intelligent tools for Prediction of heart disease using machine learning model. This synopsis provides an overview of a research project focused on developing a machine learning-based prediction model for heart diseases. The mortality rate and the number of affected people by this disease are expected to grow with the population of the world. But this mortality rate can prevent by early treatment and early prediction.[4] The dataset is pre-processed in the first phase. The pre-processed dataset is then input into several machine learning algorithms in the second phase. The output of the models is then examined using various metrics in the third phase.[4]This research aims at analyzing the machine learning algorithms based performance metrics such as Accuracy, Precision, F1 Score and Recall. Using confusion matrix we can easily find out the evaluation metrics. Eleven classifiers and different machine learning techniques including oversampling, hyperparameter tuning, and crossvalidation are employed in this research work to

reach the best result.[4] It is analyzed using the specific variables in the dataset by using the python programming as well as different supervised machine learning algorithms which include, Decision Tree, Logistic Regression, KNN and Random Forest. Decision Tree (DT), Decision Tree Regressor (DTR), Random Forest (RF), XGBoost, and Gradient Boosting (GB) algorithms.[20] Anaconda jupyter notebook is used for implementing python scripting.[21] These models may have the potential to be used in initial screening tests to identify high-risk individuals before undergoing costly investigation.[22], [23]

II. RESEARCH METHODOLOGY

Research Methodology:

- Problem Definition
- Research Objectives
- Data Collection
- Data Pre-processing
- Model Selection
- Model Evaluation
- Reporting and Documentation
- Continuous Improvement
- Dissemination and Knowledge Transfer

III. LITERATURE REVIEW

Table1. Summary of Analyzed Journal Research Articles and identified research gaps

Sr.no	Author	Journal	Year of Publicati -on	Title	Research Gaps identified along with viability/ feasibility
1	Mert Ozcan, Serhat Peker	Healthcare Analytics_el sevier	2022	A classification and regression tree algorithm for heart disease modeling and prediction	Used both classification and regression, we will focus more on classification
2	Nitish Biswas, Khandaker Mohammad Mohi Uddin , Sarreha Tasmin	Healthcare Analytics_el sevier	2022	A comparative analysis of machine learning classifiers for stroke prediction: A predictive analytics approach	Need to explore Performance measures including accuracy, recall, precision, and f1-score. i.e. confusion matrix for individual classifier and

	Rikta a, Samrat Kumar Dey				accuracy.
3	Mandakini Priyadarshani Behera, Archana Sarangi*, Debahuti Mishra , Shubhendu Kumar Sarangi	International Conference on Machine Learning and Data Engineering	2023	A Hybrid Machine Learning algorithm for Heart and Liver Disease Prediction Using Modified Particle Swarm Optimization with Support Vector Machine	Optimization of parameters on larger data sets i.e feature engineering
4	Nidhi Bhatla Kiran Jyoti	International Journal of Engineering Research & Technology (IJERT)	2012	An Analysis of Heart Disease Prediction using Different Data Mining Techniques	Focus on Data Balancing rather Data Mining Techniques
5	Ibomoiye Domor Mienye, Yanxia Sun, Professor, Zenghui Wang, Professor	Healthcare Analytics_elsevier	2020	An improved ensemble learning approach for the prediction of heart disease risk	A comparative study ensemble learning approach
6	Muhammad Salman Pathan, Avishek Nag, Muhammad Mohisn Pathan, Soumyabrata Dev	Healthcare Analytics_elsevier	2022	Analyzing the impact of feature selection on the accuracy of heart disease prediction	Correlation matrix and confusion Matrix
7	Tanmay	Biomedical	2022	Classifier	Deep Learning using

	Sinha Roy, Joyanta Kumar Roy, Nirupama Mandal	Engineering Advances_elsevier		identification using deep learning and machine learning algorithms for the detection of valvular heart diseases	ANN
8	Juan-Jose Beunzaa, Enrique Puertas	Journal of Biomedical Informatics _elsevier	2019	Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease)	Need to explore Evolution metrics
9	Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin	Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) IEEE Xplore	2019	Design And Implementing Heart Disease Prediction Using Naives Bayesian	We will explore another classifier algorithms
10	Mr.VALLE HARSHA VARDHA, Mr.UPPALA RAJESH KUMAR,	Journal of Engineering Sciences	2023	HEART DISEASE PREDICTION USING MACHINE LEARNING	Will use Data visualization
11	Vijeta Sharma	2nd International Conference on Advances in Computing,	2020	Heart Disease Prediction using Machine Learning Techniques	Will explore Data preprocessing , EDA

		Communication Control and Networking (ICACCCN)			
12	V.V. Ramalingam, Ayantan Dandapath, M Karthik Raja	International Journal of Engineering & Technology,	2018	Heart disease prediction using machine learning techniques : a survey	Principal component analysis, Web App development using Streamlit

3.1 Significance: This research project addresses a critical health issue by leveraging machine learning to improve early detection and prevention of heart diseases. The developed model and tool have the potential to enhance patient care and reduce the burden of heart-related illnesses.

- Early Intervention and Prevention
- Reduced Healthcare Costs
- Reduction in Unnecessary Testing
- Data-Driven Insights

In summary, the significance of predicting heart disease using machine learning lies in its potential to save lives, improve patient outcomes, reduce healthcare costs, and advance our understanding of cardiovascular health. By harnessing the power of data and predictive analytics, we can make more informed decisions in healthcare, leading to a healthier population and a more efficient healthcare system.

3.2 Research Gaps in prior work:

Predicting heart disease using machine learning is a critical area of research. However, there are several research gaps and challenges that still need to be addressed to improve the accuracy, of heart disease prediction models. Some of the key research gaps in this field include:

- Appropriate Data Set selection
- Data Imbalance and Bias: Many publicly available datasets for heart disease prediction are imbalanced, with a majority of samples belonging to one class (usually the non-disease class). This can lead to biased models. Addressing data imbalance and mitigating bias is a critical research challenge.
- Feature Selection and Engineering: While a variety of features, have been used for heart disease prediction, there is still room for research in feature selection and engineering to identify the most informative and relevant features. This can help reduce dimensionality and improve model interpretability.
- Dimensionality Reduction

Real-time Prediction: Developing models that can provide real-time or near-real-time predictions of heart disease risk is valuable for early intervention. Research on efficient and scalable algorithms for real-time prediction is needed.

3.3 Machine Learning Approach:

Machine Learning is the basic practice of using algorithms to make predictions by parsing data and learning from it. These models have the capability of learning by itself from prior experience or from historical data. These algorithms can figure out extract the important tasks to be performed by generalizing from examples provided to them as training sets.

Different types of ML algorithms have evolved. These are grouped by either learning style (i.e. supervised learning, unsupervised learning, and semi-supervised learning) or by similarity or by their functioning (i.e. classification, regression, decision tree, clustering, deep learning, etc.). The basic three types of ML Techniques are: Supervised Learning, Unsupervised Learning and Reinforcement Learning. The selection of the algorithm and the learning type can be made by different approaches like depending on the task accomplished, (or) the amount of data involved (or) the different types of data that are available. This exhibits a dynamic role that plays out in applications of medical diagnostics as it involves creating self-learning algorithms. Heart diseases prediction involves supervised learning technique as labelled data is required for training the model.

In case of prediction of heart diseases, regular diagnosis is very important in the initial stages of treatment which in turn reduces the risks associated with it. The most common and vital diagnostic tests include echocardiography (echo), cardiac magnetic resonance imaging (MRI), and computed tomography (CT). High quality cardiac images are produced by MRI and CT scans, which are not preferred for predictions as they have prolonged acquisition time, limited availability and involve the use of radiations.

Electrocardiogram (ECG) is a graphical representation that is produced by repolarization and depolarization of the ventricles and atria. Though there are several advancements made in the field of prediction and diagnostics, the accurate way of avoiding heart attack is not known as there are no proper symptoms associated with it. To discover a disease that forms the main causes of death such as HIV, Cancer, heart diseases, machine learning can be used. It is a great consequence to research.

Heart diseases are caused due to the deposits of fat (cholesterol) in the inner walls of arteries which narrows down or blocks coronary arteries. An efficient heart disease prediction system can be a beneficial way to exactly predict the diseases and save the patient's life. The system model presented in can interpret human patterns and accurately determine trends in the patients' records.[3]

IV. Organization of Work Elements:

- **Project Initiation and Planning:** Define project goals, objectives, and scope along with Develop a project timeline and milestones.
- **Data Acquisition**: Identify and access relevant healthcare datasets. Data Preprocessing i.e. Exploratory Data Analysis (EDA): Clean, preprocess, and format the data. Address missing values, outliers, and data quality issues. Conduct exploratory data analysis to gain insights into the dataset. Visualize data distributions, correlations, and patterns.

- **Feature Engineering**: Select relevant features or transformations to enhance model performance based on domain knowledge and analysis.
- **Data Splitting**: Split the dataset into train and test data sets.
- **Model Selection**: Choose machine learning algorithms and architectures. Implement and train multiple models with different techniques.[28]
- **Model Evaluation**: Assess model performance using appropriate evaluation metrics.
- **Reporting and Documentation:** Prepare a comprehensive research report detailing methodology, findings, and conclusions. Create visualizations and charts to communicate results effectively. Draft scientific papers for publication or conference presentations.
- **Continuous Improvement:** Implement mechanisms for ongoing model monitoring and updates with new data. Maintain and enhance the predictive model as necessary.
- **Dissemination and Knowledge Transfer:** Share research findings through peer-reviewed journals, conferences, and presentations. Engage with healthcare practitioners, policymakers, and stakeholders to discuss implications and applications

V. CONCLUSION

In conclusion, the machine learning approach for heart disease prediction holds immense promise in revolutionizing cardiovascular healthcare. Through the utilization of advanced algorithms and extensive healthcare data, this approach offers a proactive and personalized solution for identifying individuals at risk of heart disease. The machine learning approach for heart disease prediction represents a significant step towards personalized, preventive healthcare. Furthermore, the interdisciplinary nature of this research area fosters collaboration between healthcare professionals, data scientists, and researchers, leading to innovative solutions and advancements in predictive analytics and healthcare delivery.

REFERENCES

[1]M. Ozcan and S. Peker, "A classification and regression tree algorithm for heart disease modeling and prediction," *Healthc. Anal.*, vol. 3, p. 100130, Nov. 2023, doi: 10.1016/j.health.2022.100130.

[2]J. Emakhu *et al.*, "A hybrid machine learning and natural language processing model for early detection of acute coronary syndrome," *Healthc. Anal.*, p. 100249, Sep. 2023, doi: 10.1016/j.health.2023.100249.

[3]M. Swathy and K. Saruladha, "A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning techniques," *ICT Express*, vol. 8, no. 1, pp. 109–116, Mar. 2022, doi: 10.1016/j.icte.2021.08.021.

[4]N. Biswas, K. M. M. Uddin, S. T. Rikta, and S. K. Dey, "A comparative analysis of machine learning classifiers for stroke prediction: A predictive analytics approach," *Healthc. Anal.*, vol. 2, p. 100116, Nov. 2022, doi: 10.1016/j.health.2022.100116.

[5]M. Emfietzoglou *et al.*, "A MACHINE LEARNING MODEL FOR THE PREDICTION OF 1-YEAR MORTALITY AFTER HEART TRANSPLANTATION IN ADULTS WITH CONGENITAL HEART

DISEASE," *J. Am. Coll. Cardiol.*, vol. 79, no. 9, Supplement, p. 507, Mar. 2022, doi: 10.1016/S0735-1097(22)01498-X.

[6]Md. M. Ali *et al.*, "A machine learning approach for risk factors analysis and survival prediction of Heart Failure patients," *Healthc. Anal.*, vol. 3, p. 100182, Nov. 2023, doi: 10.1016/j.health.2023.100182.

[7]R. R. Sanni and H. S. Guruprasad, "Analysis of performance metrics of heart failured patients using Python and machine learning algorithms," *Glob. Transit. Proc.*, vol. 2, no. 2, pp. 233–237, Nov. 2021, doi: 10.1016/j.gltp.2021.08.028.

[8]K. M. Mohi Uddin, R. Ripa, N. Yeasmin, N. Biswas, and S. K. Dey, "Machine learning-based approach to the diagnosis of cardiovascular vascular disease using a combined dataset," *Intell.-Based Med.*, vol. 7, p. 100100, Jan. 2023, doi: 10.1016/j.ibmed.2023.100100.

[9]A. Shrivastava, M. Chakkaravarthy, and M. A. Shah, "A new machine learning method for predicting systolic and diastolic blood pressure using clinical characteristics," *Healthc. Anal.*, vol. 4, p. 100219, Dec. 2023, doi: 10.1016/j.health.2023.100219.

[10]A. N. Repaka, S. D. Ravikanti, and R. G. Franklin, "Design and implementing heart disease prediction using naives Bayesian," in *2019 3rd International conference on trends in electronics and informatics (ICOEI)*, IEEE, 2019, pp. 292–297.

[11]A. Ghasemieh, A. Lloyed, P. Bahrami, P. Vajar, and R. Kashef, "A novel machine learning model with Stacking Ensemble Learner for predicting emergency readmission of heart-disease patients," *Decis. Anal. J.*, vol. 7, p. 100242, Jun. 2023, doi: 10.1016/j.dajour.2023.100242.

[12]M. Aljanabi, M. H. Qutqut, and M. Hijjawi, "Machine learning classification techniques for heart disease prediction: a review," *Int. J. Eng. Technol.*, vol. 7, no. 4, pp. 5373–5379, 2018.

[13]B. Linse, M. Ohlsson, J. Stehlik, L. H. Lund, B. Andersson, and J. Nilsson, "A machine learning model for prediction of 30-day primary graft failure after heart transplantation," *Heliyon*, vol. 9, no. 3, p. e14282, Mar. 2023, doi: 10.1016/j.heliyon.2023.e14282.

[14]M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. Quinn, and M. A. Moni, "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison," *Comput. Biol. Med.*, vol. 136, p. 104672, 2021.

[15]M. Ozcan and S. Peker, "A classification and regression tree algorithm for heart disease modeling and prediction," *Healthc. Anal.*, vol. 3, p. 100130, Nov. 2023, doi: 10.1016/j.health.2022.100130.

[16]S. Dev, H. Wang, C. S. Nwosu, N. Jain, B. Veeravalli, and D. John, "A predictive analytics approach for stroke prediction using machine learning and neural networks," *Healthc. Anal.*, vol. 2, p. 100032, Nov. 2022, doi: 10.1016/j.health.2022.100032.

[17]V. Sharma, S. Yadav, and M. Gupta, "Heart disease prediction using machine learning techniques," in *2020 2nd international conference on advances in computing, communication control and networking (ICACCCN)*, IEEE, 2020, pp. 177–181.

[18]M. P. Behera, A. Sarangi, D. Mishra, and S. K. Sarangi, "A Hybrid Machine Learning algorithm for Heart and Liver Disease Prediction Using Modified Particle Swarm Optimization with Support Vector Machine," *Procedia Comput. Sci.*, vol. 218, pp. 818–827, 2023, doi: 10.1016/j.procs.2023.01.062.

[19] M. P. Behera, A. Sarangi, D. Mishra, and S. K. Sarangi, "A Hybrid Machine Learning algorithm for Heart and Liver Disease Prediction Using Modified Particle Swarm Optimization with Support Vector Machine," *Procedia Comput. Sci.*, vol. 218, pp. 818–827, Jan. 2023, doi: 10.1016/j.procs.2023.01.062.

[20]Md. M. Ali *et al.*, "A machine learning approach for risk factors analysis and survival prediction of Heart Failure patients," *Healthc. Anal.*, vol. 3, p. 100182, Nov. 2023, doi: 10.1016/j.health.2023.100182.

[21]J.-J. Beunza *et al.*, "Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease)," *J. Biomed. Inform.*, vol. 97, p. 103257, Sep. 2019, doi: 10.1016/j.jbi.2019.103257.

[22]H. Wang *et al.*, "Using machine learning to predict cardiovascular risk using self-reported questionnaires: Findings from the 45 and Up Study," *Int. J. Cardiol.*, vol. 386, pp. 149–156, Sep. 2023, doi: 10.1016/j.ijcard.2023.05.030.

[23]V. V. Ramalingam, A. Dandapath, and M. K. Raja, "Heart disease prediction using machine learning techniques: a survey," *Int. J. Eng. Technol.*, vol. 7, no. 2.8, pp. 684–687, 2018.

[24]M. S. Pathan, A. Nag, M. M. Pathan, and S. Dev, "Analyzing the impact of feature selection on the accuracy of heart disease prediction," *Healthc. Anal.*, vol. 2, p. 100060, Nov. 2022, doi: 10.1016/j.health.2022.100060.

[25]V. Chang, M. A. Ganatra, K. Hall, L. Golightly, and Q. A. Xu, "An assessment of machine learning models and algorithms for early prediction and diagnosis of diabetes using health indicators," *Healthc. Anal.*, vol. 2, p. 100118, Nov. 2022, doi: 10.1016/j.health.2022.100118.

[26]I. D. Mienye, Y. Sun, and Z. Wang, "An improved ensemble learning approach for the prediction of heart disease risk," *Inform. Med. Unlocked*, vol. 20, p. 100402, 2020, doi: 10.1016/j.imu.2020.100402.

[27]A. Rajdhan, A. Agarwal, M. Sai, D. Ravi, and P. Ghuli, "Heart disease prediction using machine learning," *Int. J. Eng. Technol. IJERT*, vol. 9, no. O4, 2020.

[28]A. A. Altae and A. Ehsani Rad, "Diagnosing heart disease by a novel hybrid method: Effective learning approach," *Inform. Med. Unlocked*, vol. 40, p. 101275, Jan. 2023, doi: 10.1016/j.imu.2023.101275.

[29]C.-L. Lin, C. Tseng, C.-J. Wang, S.-H. Chao, Y.-S. Hwang, and L.-J. Kau, "Development of smart cardiovascular measurement system using feature selection and machine learning models for prediction of sleep deprivation, cold hands and feet, and Shanghuo syndrome," *Measurement*, vol. 221, p. 113441, Nov. 2023, doi: 10.1016/j.measurement.2023.113441.