Heart Disease Prediction using Machine Learning Method – A Review Article

Sreekumari S [1], Rajni Bhalla [1] and Geetha Ganesan [2]

[1] *Lovely Professional University, Phagwara, Punjab, India*
[2] *Jain (Deemed-to-be) University, Bengaluru, India*

Abstract

This abstract focuses on the significance of heart disease as the leading cause of disability in the human body, primarily stemming from the reduction or blockage of coronary arteries. Coronary Artery Disease (CAD) emerges as the most prevalent form of heart disease, often remaining asymptomatic until symptoms of a heart attack or heart failure arise. This paper aims to explore various machine learning techniques utilized in heart disease prediction, supporting the medical field in expedient diagnosis. The availability of vast amounts of patient information online presents an opportunity for effective utilization. Accurate diagnosis of heart disease is of utmost importance, as incorrect predictions can have severe consequences on individuals' lives. Numerous regression and classification Machine Learning Algorithms are available, including K Nearest Neighbor Algorithm (KNN), Support Vector Machine (SVM), Neural Network, Logistic Regression (LR), Decision Tree (DT), and Random Forest (RF) Classification. Previous research indicates that SVM and Random Forest methods often exhibit higher levels of accuracy compared to other approaches. This paper provides a comprehensive discussion of commonly employed machine learning algorithms in the field of heart disease prediction.

Keywords

Coronary Artery Disease, Early Prediction, Regression and Classification, Machine Learning.

1. Introduction

Congestive heart failure is the main root of dysfunction in anatomy. It is affected by the reduction/blockage of coronary arteries. Coronary Artery Disease (CAD) is the most common type of Congestive heart failure. Common signs of Heart attack are Chest pain, body discomfort, and breath shortness. Common symptoms of Arrhythmia are fluttering feelings in the chest or in common we can quote it as “Palpitation.” Whereas common symptom of Heart failure is Breath shortness, Fatigue, or swelling on the Leg, feet, ankles, neck veins, or even in the abdomen. Other than these there are other medical conditions and life cycles that can make people at high risk for heart disease like, Diabetes, Obesity, an Unbalanced diet, Physical incapacity or inactivity, and excessive usageof alcohol.

The current method of using an angiogram to identify heart diseases has few risks to be overruled. They include allergic reactions to the dye used to see the coronary arteries, bleeding at the sitewhere the catheter is placed, and the usage of dye can even cause damage to the kidney. So, we recommend the usage of non-invasive methods. Predicting heart disease is extremely critical. Becauseone error in prediction can lead to the death of the person. If a patient has diseases such as kidney failure or diabetes, then they must experience a higher risk after an angiogram. Some of such risks are:

- May be allergic to anesthesia and contrast dye.

CVMLH 2023: Workshop on Computer Vision and Machine Learning for Healthcare, May 20 – 21, 2023, Chennai, India.
EMAIL: sreekumari.sreelesh@gmail.com (Sreekumari S)
ORCID: 0000-0002-0207-4738 (Sreekumari S)

- Bleeding at the insertion point of the catheter.
- Blood clots or injury to an artery or vein.

These kinds of people must take contrast medicine before 24 hours to check whether they are allergic or not. All human beings do not possess higher pain tolerance. Those who are having low pain tolerance cannot undergo an angiogram because of the high-risk factors. In such cases, we must use non-invasive methods to predict heart diseases. Heart diseases are of diverse types. They are:

- Coronary Artery disease (CAD)
- Heart Arrhythmias
- Heart failure
- Heart valve disease
- Stroke etc.

To predict heart disease with the best accuracy prediction algorithm should be the right one. Otherwise, the prediction will go wrong which leads to the death of a person. Here, we have summarized recent activities in which we can find distinct predictive analytics.

Since heart disease datasets are easily available online, data collection becomes remarkably simple. There is a lack of knowledge of data. That will be achieved after processing the data. This should be performed by collecting the proper data after avoiding inappropriate data. Then find out the missing data and remove all the duplicates. This step is called data preprocessing. After this split the dataset into two such as the training set and testing set. Then apply different classification and regression algorithms to predict whether a person has a chance of any heart disease. From the accuracy, choose the best predictive method. Fig-1 depicts the predictive method.

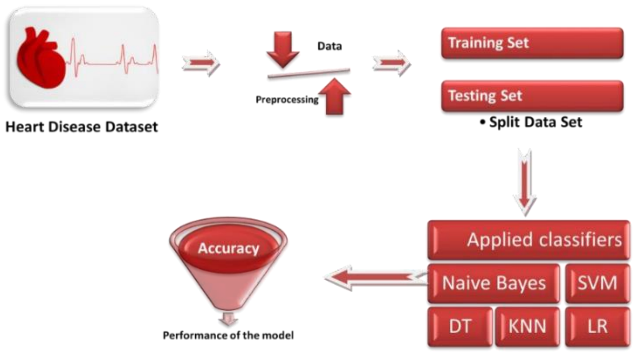

Figure 1: Predictive Model

2. Supervised Machine Learning Algorithms

In supervised machine learning, training data is provided to the machine which works as a supervisor that teaches the machines to predict the output accurately. In this method, labeled data is provided to train the model and the model learns about distinct types of data. After the training process, the model is evaluated based on test data which is a subset of the training set and then it starts predicting the output. Supervised learning can again be divided into two. They are Regression and Classification. (Fig-2). The following are different machine learning algorithms.

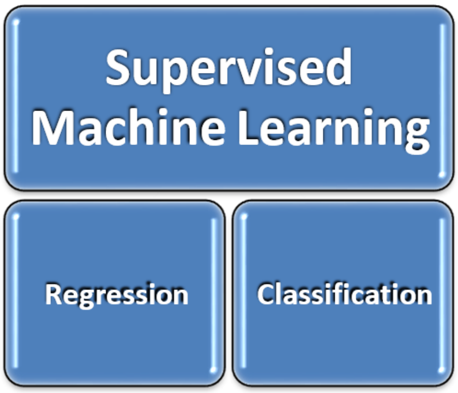

Figure 2: Categories of Supervised Learning

2.1. Logistic Regression

It is one of the most significant machine learning algorithms because it can provide probabilities and classify new data by using discrete and continuous datasets. The first step is to extract the dependent and independent variables. Then splitting of the dataset and feature scaling is performed. The outcome must be of a categorical or discrete value.

2.2. K- nearest neighbor algorithm

This supervised learning algorithm stores all the given data and based on the similarity it classifies a new data point. It is also known as the "Lazy- Learner Algorithm." Because it acts on the dataset during classification instead of learning from the training dataset. It can be used for both Regression and Classification. But typically used for Classification problems. It will be more effective if the training data is large.

2.3. Support Vector Machine

This is a well-liked supervised learning algorithm. In this plot, each data is a point in n- dimension space with each value of feature being a particular coordinate value. Then classification is performed after finding the hyper-plane which is also known as the best decision boundary.

2.4. Artificial Neural Network

It is a special type of machine-learning algorithm that is used to model the human brain. ANN is capable of learning from the data, and it provides results as classification or prediction such as neurons in our nervous system learning from past data. ANNs are non-linear statistical models that show complex relationships between inputs and outputs. The three important layers are the input layer, hidden layers, and output layer. The hidden layer may be single or multiple. In this model, different parameters affect the performance of the model. The output is dependent on this data. The input layer communicates with the hidden layer and the activated neurons will continue passing until it reaches the output layer.

2.5. Decision Trees

It is a type of supervised machine learning algorithm, in which the data set is always split based on a defined constraint. Decision nodes and leaves are the two results that can be derived from the decision tree.

2.6. Random Forest Classification (RFC):

RFC is based on the concept of combo learning, which means a process of solving a complex problem to an improved and more accurate level of performance of the model by combining the multiple classifiers.

Since this review paper is about machine learning techniques, we have gone through different machine learning algorithms. There are ample methods used in heart disease prediction other than machine learning. They are deep learning, Data mining, IoT tools, etc.

3. Literature Review

E. K. Oikonomou et al. [1], they have given a clear idea about Artificial intelligence, Machine learning, big data, and different prediction methods like supervised and unsupervised methods. CT (Cardiac Computed Tomography) imaging is the best way to predict Cardio Artery Disease (CAD), which will mitigate the complexity of doctors to predict CAD by using images, and it will be helpful for the patients to identify the disease faster. Predictions are performed using Regression analysis and Classification methods. Regression analysis is performed on the image using a deep neural network (DNN) algorithm or Support vector machines or decision trees. They have given a good explanation of Radiomics. For image processing, detection, and segmentation they have used CCTA (Coronary CT Angiography) scans and Machine Learning (ML) Algorithms. They stated that it can predict many risks which are not possible with humans. They concluded by explaining the limitations and stating that this work has offered a lot to the patients as well as doctors.

Li Bin et al. [2] tried to establish a disease pre-alert model within the 3 Year gap for hypertension patients. To achieve public data analysis through AI technologies they have used spark data analysis for hypertension patients. Using data mining, unbalanced data is converted to usable form using the Z-score standard. Stroke, Heart failure, and Renal Failure symptoms viz. 3 group predictions ascertain the pre-alert risk warning using a data mining algorithm. AI technologies used in this model are Logistic Regression (LR), Support Vector Machine (SVM), and Naive Bayes Algorithm (NB). For feature selection, they have used the Chi-Square test. In this, they have used three types of datasets. They are Stroke, Heart failure, and Renal Failure. But they perform classification on two datasets viz. Stroke and Heart failure. The algorithms used were Naïve Bayes, SVM, and Logistic regression. The stroke data set shows an accuracy of more than .75, while classification results on the heart failure dataset show an accuracy of more than .85. Logistic regression is superior to Naïve Bayes and SVM. They have shown the Receiver Operating Characteristic curve as well (ROC). They have used SVM-RFE for feature selection, which is not suitable for choosing the most suitable feature subset. For getting better results need to use a different algorithm for comparing and analyzing in the next research because SVM-RFE cannot evaluate various feature selection algorithms.

S Manish et al. [3] they have given a brief description of the technologies used for the study. They have stated that this is the first paper that uses Optimally Time-Frequency Concentrated (OTFC) even-length biorthogonal wavelet filter bank (BWFB) for automatically detecting CAD. The Fuzzy Entropy (FE) and log Energy (LogE) features were extracted from various subbands of the filter. They used Raw ECG data from twenty males and twenty females and CAD ECG from Six females and one male. Then signals were classified using Gaussian Support Machine Classifier (GSVM). They have stated that this study shows high accuracy compared with other methods and while others are using thirty features for classification, they use only twelve features.

A Moloud et al. [4] proposed a data mining method for CAD analysis. Data manipulation and normalization were performed first. Then they selected features using GA and PSO algorithms. 10-fold cross-validation was performed on subsets. For classification purposes in the prelims test, they used ten algorithms. Among them, the best three SVM classifiers were selected for the final. They are nu-SVC, nu-SVM, and Linear SVM. To optimize data, they used two optimization methods. Then they generated a confusion matrix for evaluating results. They have selected fitness functions. It has shown an accuracy of 93.08% based on the confusion matrix. They concluded by saying that instead of normalization other preprocessing approaches should be used in the future.

R Silvia et al. [5] The objective of this review article is to describe the contemporary state of artificial intelligence in clinical practice. In this paper, they have explained very well about Artificial Intelligence and different subfields of Artificial intelligence like Machine learning, and deep learning and their subfields. They have given proper descriptions of new emerging communication and information technologies like Mobile health and IOT. These are subfields of E-health. And explained applications in cardiovascular imaging as well. Echocardiography, Magnetic resonance Imaging, Cardiac computed tomography, and Electrocardiography are different applications in Cardiovascular Imaging.

Johnson et al. [6] have done a study about how AI and ML relate to statistics and why cardiology needs AI. They have explained in detail how to choose the best algorithm for feature selection. Through this paper, we can easily understand different supervised Machine Learning algorithms used in Cardiovascular disease prediction like Logistic Regression, Regularized Regression, Tree-Based methods, Bootstrap, and Support Vector Machine. For explaining these they have given proper examples. In Unsupervised Machine learning, they have given a brief description of Neural networks and Deep Learning methods. The two most common forms of DL are CNN and RNN and explained the major disadvantage of deep learning. They have concluded by stating to use this study as a decision tool for medical practitioners.

Acharya et al. [7] explained invasive and non-invasive techniques used in CAD detection. They have used Gaussian Mixture Model (GMM) Classifier for this purpose. Non-linear parameters are extracted from echocardiography images. They have taken four hundred normal images of thirty normal subjects and Four hundred CAD images of thirty affected subjects. They have noticed the ischemic changes. Twelve lead electrocardiogram images were checked for wall motion abnormality. The training set and test set are there. An expert physician can mark the image collected as belonging to a normal or CAD-affected person. Then different classifiers are used to extract the features. A total of 559 features are extracted from each image. T-test was used for statistical analysis. If p<.01 or.05, the feature is measured as very perceptive. A three-fold stratified cross-validation technique is used for developing and testing the classifiers. In this, they have used six classifiers. They are DT, FS, GMM, RBPNN, KNN, and NBC. Then the heart index is calculated which is a combination of nine distinctive features. The heat index for normal is 2.52± 0.07 and for CAD 2.86± 0.11. They have stated that this can be transformed into software, can be implemented on any device, and concluded that this shows 100% accuracy.

S E Golovenkin et al. [8] tried to improve prediction by using the voting method of three competing systems and by eliminating sparse data columns. In this study, the prediction was performed using Matlab2016 System, the Neural Network Toolbox module. In Artificial Neural Network, there was one input layer and two hidden layers. All layers are interconnected, and the output layer predicts the results. The voting method is used in this. The “Sliding- Window” method was used to increase the efficiency of the learning process. Standard indicators are Accuracy, Sensitivity, and specificity. They have eliminated sparse set input data columns. They have also stated that the use of 3 Neural Networks increases accuracy by 1.2%.

Babaoglu et al. [9] used Principle Component Analysis (PCA) which is a method of identifying patterns in data and expressing them by highlighting differences and similarities. They have taken a dataset of 480 patients after excluding patients with some problems. They have used a Support Vector Machine for classification and regression tasks. The assessments are implemented in Matlab 7.0 application using the LIBSVM package. The SVM method for twenty-three features shows less accuracy. So, using the PCA method, features were reduced to eighteen. Then it shows an accuracy of 79.17%. They have concluded by stating that, PCA reduced dataset using the SVM method increases the diagnostic accuracy rate, decreases the sum of the training and test time, and training error in the assessment of EST (Exercise Stress test) in the determination of CAD with SVM method.

Zong Chen et al. [10] tried to provide the recent adaptive image-based classification techniques. They have explained diverse types of heart diseases. They have compared the prediction methods used in different studies. The first step is image registration which can be performed on a raw image database and de-noising processing should be done. The feature is extracted from the image. They have used two different training models. They are Naive- Bayes Classification and SVM. They have concluded by stating that SVM is more accurate than Naïve Bayes Classification. The database needs to be updated with more descriptions of patients.

M M Ali et al. [11] used the Kaggle dataset. They have done a comparison study and developed their experimental model. They have applied six classification algorithms to compare the accuracy of the best performer and statistical variables by using a ten-fold cross-validation method. The six classifiers were KNN, Random Forest, Decision Tree, AdaboostM1(ABM1), Logistic Regression, and Multilayer Perception (MLP). They have listed the five most crucial features based on feature importance and correlation value. In this study compared to other classifiers RF shows 100% accuracy. KNN and MLP failed to generate feature importance scores or coefficient values. But this method has one limitation, this data model is not enough to address all the issues.

Yazdani A et al. [12] used WARM (Weighted Associative Rule Mining) to predict heart disease. They have compared the efficiency of earlier models that used the WARM method to predict heart disease. In this model, weights are assigned to identify the best features mined. Feature values are evaluated. For Eg: Feature values for sex are Male and Female. After that calculate the total weight for the feature. Then applied WARM using the Apriori algorithm. Of thirteen features, eight significant features were identified. Finally, the confidence score is generated. This model shows a 98% confidence score which is better than the earlier studies.

Sumit S et al. [13] proposed this study to create a deep learning neural network heart disease prediction model by using a new optimization technique called Talos Hyperparameter. They have used UCI Heart Disease Dataset. They applied different learning algorithms like Logistic regression, KNN, SVM, Naïve Bayes, and Hyperparameter optimization (Talos). Talos follows POD (Prepare, Optimize, and Deploy) process workflow. Reporting and evaluation are the last process after POD. Talos method shows an accuracy of 90.78% which is better compared to other learning methods.

Bharati R et al. [14] conducted this study to compare the performance of both Machine learning as well as deep learning methods. They have considered a public dataset. In this, they compared both machine learning as well as deep learning methods. In pre-processing the data, checked the distribution of data, skewness of data, stats of normal distribution data, feature selection, and checking for duplicates. after that proposed machine learning classifiers as well as deep learning classifiers. Then they evaluated the process and concluded by stating that dataset size needs to be increased and more optimization and normalization techniques need to be used.

Ishaq A et al. [15] conducted this study for comparing different machine learning techniques to select the most suitable method for heart disease survival prediction. The dataset used was UCI Repository. Then applied RF to employ feature ranking. Classified the data into training and testing sets. They applied one oversampling method called SMOTE (Synthetic Minority Oversampling Technique) to deal with imbalanced data in medicine. Performance evaluation matrices were Accuracy, Precision, Recall, and F-Score. Among 9 classifiers, they got more accuracy for RF, and it was .8889. With SMOTE, it shows 10% more accuracy. Extra Tree Classifier (ETC) shows an accuracy of .9262. Here RF produces Constant approximation, ETC produces multi-linear approximation.

Rajni Bhalla et al. [16] presents a detailed examination of the effectiveness and performance of both proposed and existing methodologies for classification task. The study aims to compare and evaluate the advantages and limitations of different methodologies for collecting and analyzing data for structured data. By conducting a comparative analysis, the paper provides valuable insights for researchers and practitioners seeking to optimize their methodologies for classification tasks and for data that is in a structured format.

Abdel-Basset M et al. [17] proposed this study to develop an IoT model to detect and monitor heart failure-infected patients. Achieved from various wearable sensors. They used WBAN or medical sensor nodes. The data extracted from WBAN convey are sent to the mobile application through Bluetooth or ZigBee and stored in the cloud server to store, process, and broadcast data. Then N- MCDM model is used to determine the percentage of heart failure disease. If the patient has severe heart failure, then an ambulance will be sent to the patient and treatments will be started. This will provide the ratio of heart failure so that the clinicians can easily decide the type of treatment.

Rajni Bhalla et al. [18] presents a novel ensemble-based machine learning model that combines multiple categorical datasets to enhance prediction accuracy. The proposed approach consists of three main components: dataset alignment, model training, and prediction aggregation. The dataset alignment phase focuses on mapping categorical variables across different datasets, ensuring consistency and compatibility.

Rajni Bhalla et al. [19] concludes by identifying the current research trends and challenges in sentiment classification using hybrid ensemble-based approaches. It highlights the need for further investigation into the combination of diverse classifiers and the integration of deep learning techniques in ensemble frameworks. The hybrid ensemble approach can be applied to classification dataset like heart disease detection to achieve better accuracy.

Sivagowry S et al. [20] created a review of the application of data mining in heart disease prediction. In this, they have compared different data mining algorithms and concluded like heart disease prediction with data mining will become most successful with a smaller number of attributes, and text mining the medical data needs to be extended in predicting the health care data.

4. Methodology

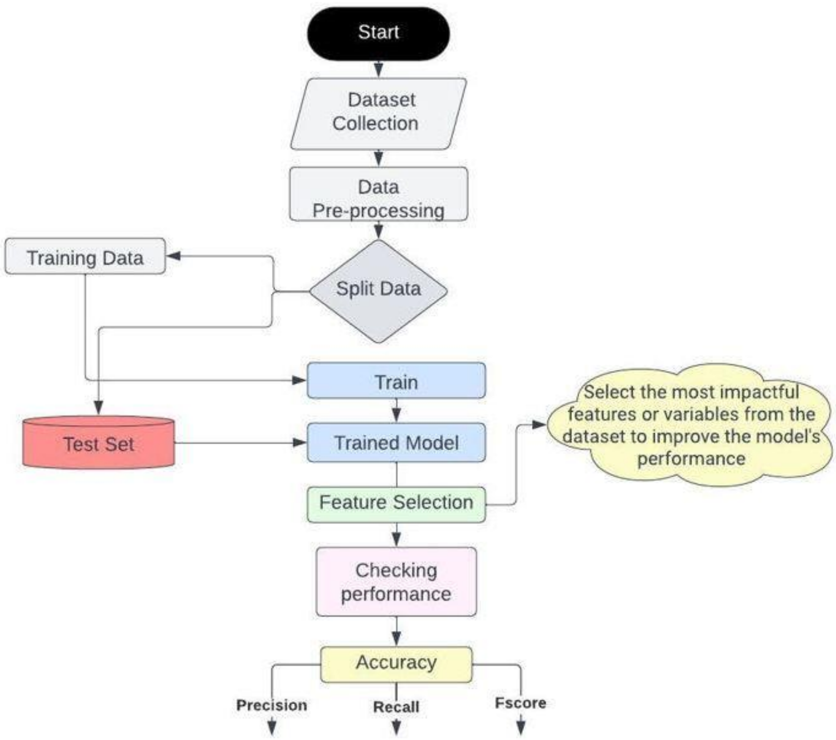

Figure 3: Methodology

Data collection is the first step. Data can be collected in two ways such as Online data and real-time data. Then we have to remove noise, missing values, useless values and all. This step is called data preprocessing. Most of the online datasets are already preprocessed. So, we can directly use it without preprocessing. But for real-time data, preprocessing can be done in different ways such as by using machine learning, deep learning or data mining methods. After this step we will get the data in usable format. Now, need to split the dataset into Training and Testing set. We can split the dataset in such a way that 70% for training the model and 30% for testing. First train the model with train set. Then select the most significant feature for improving the prediction accuracy. Now, give test set to the model. It will predict accordingly. There are different evaluation metrics for checking the efficiency of the model such as Precision, Recall, F1-score, confusion matrix and ROC curves etc. (Fig-3 depicts the same).

5. Results

There are different regression and classification algorithms used to predict heart disease. Different methods are used in different papers, and they have shown the accuracy that they obtained while using different methods. Table 1 shows the comparison of different works. In most of the papers, the different algorithms used are Naive Bayes, Logistic Regression, ANN, SVM, DT, RF, etc. Among them, ANN and RF show more accuracy.

Table 1: Comparison of previous works

Reference	Methodology	Dataset	Accuracy
[1]	Support Vector Machine	42 CCTA	Sensitivity- 93% Specificity-95%
[2]	Naive Bayes SVM (Linear) Logistic Regression	Stroke Dataset 5257 Heart failure Dataset 714	On Stroke Dataset Naive Bayes- 0.7662(Accuracy) SVM (Linear)-0.7827 Logistic Regression-0.7820 On Heart Failure Dataset Naive Bayes- 0.8571 SVM(Linear)-0.8595 Logistic Regression-0.8529
[3]	Signals were segmented into two seconds (Set A) and five seconds (Set B)duration. Then applied GSVM.	1. Signals used are i)Normal ECG database. ii)CAD ECG signals. Raw ECG data from twenty males and twenty females and CAD ECG from 6 females and one male.	GSVM -99.53%
[4]	Cross-validation – stratified 10-fold. Machine learning- KNN, reglog, GaussNB, LDA, QDA, RandomForest,MLP, SVM (3 types, including C-SVC, nu-SVC, and Linear SVM.	303 patients annotated with fifty-four factors. (CAD and Normal, with216 CAD and 87 Normal (non-CAD) patients).	nuSVM method with GA and ACC fitness function testing set shows an accuracy of 93.08%
[7]	Heart Index	400 normal images of 30 subjects and 400 CAD images	The heart index for normal is 2.52± 0.07 and for CAD.

		of 30 affected subjects using the Vivid 7 Dimension modality forGE Health care. The normal group age range was 20- 45, and the CAD group age range was 35-65.	2.86± 0.11
[8]	Voting method Eliminated the sparse set input datacolumns.	1700 patients with MyocardialInfarction. There were 1065 (62.6%) Men and 635(37.4%) Women.	Neural network -95% accuracy for 3 out of 12 diagnoses with up to 98% specificity.
[11]	Supervised machine learning algorithms 1) KNN 2) Random Forest 3) Decision Tree 4) Adaboost M1 5) Logistic Regression 6) Multilayer Perception	Kaggle Dataset	Random Forest -100% accuracy.
[12]	Applied WARM Apriori Algorithm	Cleveland dataset from UCI Machine Learning Repository.	Confidence Score-98%
[13]	Learning Algorithms a. Logistic Regression b. KNN c. SVM d. Naïve Bayes e. Hyper-parameter optimization (Talos) f. Random Forest Classifier	UCI Heart Disease Dataset.	Talos-90.78%
[14]	Machine learning classifiers Deep learning pseudocode	Public health dataset. 4 Databases 1. Cleveland 2. Hungary 3. Switzerland and 4. Long Beach V	Machine learning K neighbors – 84.8% Deep Learning 94.2%
[15]	9 Classifiers Synthetic Minority Over Sampling technique (SMOTE)	UCI	RF-.8889 Extra Tree Classifier with SMOTE- .9262
[17]	IoT model WBAN sensors N-MCDM model	Wearable sensors	Returns Heart failure ratio.

6. Conclusion and Future Work

In conclusion, the diagnosis of heart disease remains a paramount concern within our medical society. Manual predictions, though rare, can still yield incorrect results due to the need for precise image assessment. Non-invasive methods provide an alternative for individuals concerned about traditional diagnostic approaches, enabling them to take proactive preventive measures. This study highlights various approaches such as machine learning, deep learning, data mining, and IoT tools for predicting heart disease. However, the accuracy of these methods poses challenges, given the high stakes involved in predicting human life. Even models achieving 100% accuracy during testing may encounter failures when applied to real-time data. In this software-driven era, it is crucial to develop robust and cost-effective solutions that deliver more accurate results. Further comprehensive studies should be conducted to advance our understanding and achieve better outcomes in this field.

7. References

[1] Oikonomou E K, Siddique M & Antoniades C, Artificial Intelligence in Medical Imaging: a radiomic guide to precision phenotyping of Cardiovascular disease, Cardiovascular Research, 116(13), 2040-2054. https:// doi:10.1093/cvr/cvaa021.

[2] Li, B, Ding S, Song G, Li J, Zhang Q. 2019. Computer-Aided Diagnosis and Clinical Trials of Cardiovascular Diseases Based on Artificial Intelligence Technologies for Risk-Early Warning Model. J Med Syst 43, 228. https://doi.org/10.1007/s10916-019- 1346-x.

[3] Manish Sharma, U. Rajendra Acharya, 2019. A New Method to Identify Coronary Artery Disease with ECG Signals and Time-Frequency Concentrated Antisymmetric Biorthogonal Wavelet Filter Bank, Pattern Recognition Letters. S0167-8655(19)30121- 7. https://doi.org/10.1016/j.patrec.2019.04.014.

[4] Abdar M, Ksiazek W, Rajendra Acharya U, Tan R S, Makarenkov V, Plaviak P. 2019.A New Machine Learning Technique for an Accurate Diagnosis of Coronary Artery Disease, Computer Methods, and Programs in Biomedicine. S0169-2607(18)31458-5. https://doi.org/10.1016/j.cmpb.2019.104992.

[5] Romiti S, Vinciguerra M, Saade W et al.,2020. "Artificial Intelligence (AI) and Cardiovascular Diseases: An Unexpected Alliance", Cardiology Research and Practice, vol. 2020, Article ID 4972346, 8 pages. https://doi.org/10.1155/2020/4972346

[6] Johnson, K. W., Torres Soto, J., Glicksberg, B. S., Shameer, K., Miotto, R., Ali, M., ... & Dudley, J. T. (2018). Artificial intelligence in cardiology. Journal of the American College of Cardiology, 71(23), 2668-2679. https://www.jacc.org/doi/abs/10.1016/j.jacc.2018.03.521.

[7] Acharya U R, Sree S V, Krishnan M M R, Krishnananda N, Ranjan S, Umesh P & Suri J S, (2013). Automated classification of patients with coronary artery disease using grayscale features from left ventricle echocardiographic images, Computer Methods and Programs in Biomedicine, Volume 112, Issue 3,2013, Pages 624-632. https://doi.org/10.1016/j.cmpb.2013.07.012.

[8] Golovenkin S E, Dorrer M G, Nikulina S Y, Orlova Y V, & Pelipeckaya E Y, (2020), Evaluation of the effectiveness of using artificial intelligence to predict the response of the human body to cardiovascular diseases. Journal of Physics: Conference Series, Volume 1679. No. 4, https://iopscience.iop.org/article/10.1088/1742- 6596/1679/4/042017/meta.

[9] Babaoglu I, Findik O & Bayrak M, (2010). Effects of principle component analysis on assessment of coronary artery diseases using support vector machine, Expert Systems with Applications, Volume 37, Issue 3, 2010, Pages 2182-2185. https://doi.org/10.1016/j.eswa.2009.07.055.

[10] Chen J I Z, & Hengjinda P. (2021). Early Prediction of Coronary Artery Disease (CAD) by Machine Learning Method- A Comparative Study, Journal of Artificial Intelligence and Capsule Networks, Vol.03/ No.01, Pages: 17-33. https://doi.org/10.36548/jaicn.2021.1.002.

[11] M M Ali, B K Paul, Kawsar Ahmed, Francis M B, Julian M W, Q M A Moni, (2021), Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison, Computers in Biology and Medicine, Volume 136, https://doi.org/10.1016/j.compbiomed.2021.104672.

[12] Yazdani A, Varathan K D, Chiam Y K, Malik A W, & Wan Ahmad, W A (2021). A novel approach for heart disease prediction using strength scores with significant predictors. BMC Medical Informatics and Decision Making 21(1). https://doi.org/10.1186/s12911-021-01527-5.

[13] Sumit Sharma, Mahesh Parmar (2020), Heart Diseases Prediction using Deep Learning Neural Network Model, International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-9, Issue-3. https://doi.org/10.35940/ijitee.C9009.019320.

[14] Bharti R, Khamparia A, Shabaz M, Dhiman G, Pande S, Singh P. Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning. Comput Intell Neurosci. 2021 Jul 1; 2021:8387680. doi- 10.1155/2021/8387680. PMID: 34306056; PMCID: PMC8266441.
[15] Abid Ishaq, Saima Sadiq, Muhammad Umer, Saleem Ullah, Seyedali Mirjalili, Vaibhav Rupapara, and Michele Nappi, Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques, IEEE Access, VOLUME 9, 2021, Pages 39707 -39716.
[16] Bhalla R, Amandeep. A comparative analysis of application of proposed and the existing methodologies on a mobile phone survey. In Futuristic Trends in Networks and Computing Technologies: Second International Conference, FTNCT 2019, Chandigarh, India, November 22–23, 2019, Revised Selected Papers 2 2020 (pp. 107-115). Springer Singapore.
[17] Abdel-Basset, M., Gamal, A., Manogaran, G., Son, L. H., & Long, H. V. (2019). A novel group decision-making model based on neutrosophic sets for heart disease diagnosis. Multimedia Tools and Applications. doi:10.1007/s11042-019-07742-7.
[18] Bhalla R, Sharma A, Gupta J. An Ensemble-based Machine Learning Model for Accurate Predictions using Multiple Categorical Datasets. In2022 Second International Conference on Interdisciplinary Cyber Physical Systems (ICPS) 2022 May 9 (pp. 1-4). IEEE.
[19] R. Bhalla, A. Sharma, and G. Ganesan, "A Bibliographic Survey of Sentiment Classification using Hybrid Ensemble-based Machine Learning Approaches," in ceur workshop proceedings, 2022, vol. 3146, pp. 33–38, [Online]. Available: https://ceur-ws.org/Vol-3146/PAPER_04.pdf.
[20] Sivagowry, S.; Durairaj, M.; Persia, A. (2013). - An empirical study on applying data mining techniques for the analysis and prediction of heart disease., [IEEE 2013 International Conference on Information Communication and Embedded Systems (ICICES 2013) - doi:10.1109/ICICES.2013.6508204.