International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 04 | Apr 2022 www.irjet.net p-ISSN: 2395-0072

A COMPREHENSIVE SURVEY ON CARDIAC ARREST RISK LEVEL PREDICTION SYSTEM

Priyanka K M [1], Priyanka B G [2], R Tejonidhii [3], Raksha S Hegde [4], R Kasturi Rangan [5]

[1-4] Students, Information Science and Engineering
[5] Professor, Dept. of Information Science and Engineering
Vidyavardhaka College of Engineering, Mysuru, Karnataka, India

---

Abstract - *These days, Heart related problems are increasing day by day because of way of life, inherited. Most importantly coronary illness is the new normal, hence the lifespan of them is in greater danger. So, identification of Cardiac disease at early stages and providing proper treatments plays a vital role in the current medical sector. Many efficient techniques and algorithms are utilized for the disease prediction with interest in Cardiac sector. As per medically demonstrated results the ordinary Blood pressure is 120/90 and heart beat rate is 72bpm. The patient danger level is grouped utilizing machine learning and data mining arrangement procedures like Naive Bayes, Support Vector Machine, KNN, Logistic Regression, Decision Tree, Neural networks, Random Forest (RF).*

Key Words: *Cardiac disease, Data Mining, Naive Bayes, KNN, SVM, Logistic Regression, Decision Tree, Neural networks, Random Forest.*

1. INTRODUCTION

Nowadays, Heart related issues are growing day by day in view of lifestyle, acquired. Especially, coronary ailment has become more ordinary these days which makes people life difficult. The patient risk level is calculated using SVM, Logistic Regression, KNN and many more. A few non-modifiable variables are additionally present such as smoking, drinking additionally justification behind coronary illness. Whenever blood veins are overstretched, the danger meter of the veins are expanded which prompts the blood pressure.

Age is a factor that is non-modifiable and is a danger which additionally one of the justifications for coronary illness. Nearly more than 41% of demise due to heart sickness can be seen due to the smoking as the major issue. Since it majorly cut-offs the entry of oxygen to the blood thereby causing harm to the veins. Different mining strategies are used to foresee the risk of cardiac illness. The patient movement is observed carefully and often in the event that there are any progressions happen, the hazard level of illness is said to the patient by specialist

2. Generic Block Diagram

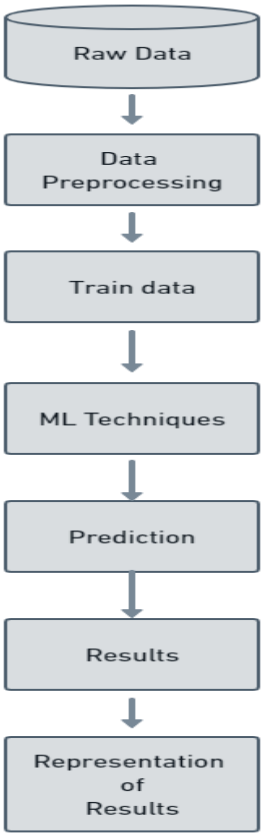

Fig 1: Flowchart for prediction

In the initial phase, we gather clinical information. In earlier years patients' information was gathered for handling. More than 5000 + informational collections are utilized for handling. Preparing informational indexes will contain patient subtleties and furthermore boundaries that are expected for expectation. Clinical information is dissected and only significant information is extricated. Required information extraction is done in light of the fact that whole preparation information is not needed for handling and assuming we input all information, it requires a lot of time for handling, so information handling is finished. Supervised Learning Technique: It's a prescient model utilized for the assignments where it includes a forecast of one worth involving different qualities in the informational index. We

have numerous calculations to construct models in regulated learning like KNN, Naive Bayes, and so forth. Depending on the necessity, marks, boundaries, and informational index we select the fitting calculation for expectations.

The calculation is utilized to construct a model that makes forecasts in light of proof within the sight of vulnerability. The doctor can access the module where the framework predicts the cardiac problems and types for the new patients. Results produced by the algorithm are checked with precision by utilizing the confusion matrix strategy, here we approve the outcomes produced by the calculation. The final results are displayed on GUI. At the point when clients get to sign in to the application, the framework predicts the sickness and presentations it on a GUI.

3. Challenges

There are several challenges in each stage of cardiac disease prediction:

1. Choosing the right dataset: Plenty of datasets are available of different varieties. Picking the right one which serves our purpose is crucial as the whole process depends on this phase.
2. Picking the right attribute for producing the accurate outcome: We may have 100+ fields in our data. But eliminating the fields whose absence doesn't affect the performance of the algorithm is important.
3. Distinguishing and taking care of the missing values. In information pre-processing, it is crucial to distinguish and accurately handle the missing parts, neglecting to do this, you could make erroneous and flawed determinations and inductions from the information. Obviously, this will hamper the venture.
4. Choosing the suitable algorithm to achieve a good accuracy score: When it comes to algorithms it is all about accuracy, so picking the right algorithm which yields good results is a must.
5. Assess the model's presentation and lay out benchmarks. Assessment incorporates model metric assessment, confusion matrix estimations, KPIs, model execution measurements, model quality estimations and a last assurance of whether the model can meet the laid-out business objectives.
6. Repeat and change the model: Even however the model is functional and you're constantly checking its exhibition, you're not done. With regards to carrying out advances, it's generally expected said that the recipe for progress is to begin little, imagine something truly mind-blowing and repeat frequently.

4. Motivation:

More number of individuals are getting coronary disease consistently and these are the greatest enemy of people all over the planet. The World Health Organization (WHO) examined that twelve million occurring happens overall because of Heart illnesses. Coronary illness is taking the lives of many individuals in a very frequent manner Clinical conclusion assumes an essential part but is a muddled task that should be executed proficiently and precisely. To decrease the cost of clinical tests, suitable PC-based data is required. Learning of the major risky parts associated with coronary disease helps administrations specialists to perceive patients at high risk of having cardiac illness. Finally for the purpose of human living benefits, spreading awareness to the society and helping them is the main motivation.

5. Literature Review

Different studies have been done and models has built for prediction of heart disease, various data mining and machine learning methods are used for predicting the disease and accuracy level is obtained for various methods

In [1] Theresa Princy, et al. The risk level for the patient is classified using machine learning techniques like KNN and ID3 algorithm. There are two phases namely Classification and regression, study consists of two modules i.e., classifier and prediction module. Using basic attribute such as BP, Cholesterol the accuracy level of the prediction was 40% and by adding smoking and family history the accuracy was increased to 80%. In [10] C. Kalaiselvi et al. the main objective is to diagnose heart disease with the reduced number. Classification and clustering algorithm are used in this method. The implementation part is done using the MATLAB 12 where data is classified as two types as getting heart disease and with no heart disease. The cost is reduced for different medical test and it help patients to take precautionary measures in advance. The accuracy for naive bayes classifier is about 90.72%, decision tree is of 96.62% and KNN is about 97%.

In [6] N. Komal Kumar et al. mainly determines about the prediction the ailment dependent on the side effects and conditions of patient and mainly deals with cardio vascular disease. The cleaned data is split into two types namely training and testing data and subjected to 5 classifiers and accuracy is found by confusion matrix thereby finding best classifier for predicting cardiovascular disease. Among all the predicted models, the higher accuracy was observed using random forest with 85.71%. The other models like decision tree and logistic regression have accuracy of 74.28%, SVM with 77.14% and KNN with 68.57%.

Mortality of the patients with heart disease is predicted using data mining methods [4]. ANN [Artificial neural

networks], logistic regression and decision tree are the three models used in predicting death of patients within 12 months of suffering from disease. In this model, RapidMiner studio was used for implementation. Model performance is dealt with some of the criteria like accuracy, specification and precision Among the three methods, neural network gives the higher accuracy of 74.3%. In [2] Sana Bharti, Dr. Saliendra Narayana Singh talks about how various algorithms can be used to predict the heart disease. The algorithms include data mining techniques like clustering, association rule for prediction. The basic flow of this model consists of database of patients, analysis of data, feature selection, optimization algorithm, training and classification. There are total of 14 attributes used in this method like age, gender, BP level, obesity, heart rate, sugar level etc. This type of prediction has an accuracy rate of 60.74%.

The [5] is about prediction of cardiac disease by using decision tree method with an accuracy rate of 73% where as other major algorithm like Naïve Bayes, KNN [K-Nearest neighbors], SVM, Logistic Regression which has a range of accuracy rate about 60-70%. The objective of this paper is to use a cardiac dataset which is grouped by using supervised machine learning algorithms. Some method to find out the accuracy of an algorithm include Accuracy prediction, precision, this method has proven more accuracy using decision tree. In [12] it is a study conducted by different algorithms like Naive Bayes, KNN, Logistic Regression. Regularization of model is used in this approach which helps in improving overall accuracy rate. Due to complex nature of cardiac disease, the models need professional advices, implementation of patient's issues to determine how severe is the attack on heart and also any previous attack info of the patient. Logistic regression has an accuracy of 77% and 81.48%

Accuracy with naïve bayes, highest accuracy is recorded in KNN model with 94.60%.

In [3] Seyedamin Puriyed et al. have provided aim to compare different machine learning techniques on dataset. The algorithms are applied individually and in combinations using precision(sensitivity). In [9] Khashman adan, et al. mainly determine about the discussion regarding the death rate infused by various cardiovascular heart disease, by designing an intelligent system for preventing misdiagnosis of heart disease. The dataset is divided into two categories namely train and test set. The dataset is in the ratio format of 60:40. The advantage of this type of division is to avoid underfitting of the data which occur if the training dataset is very smaller than the testing dataset. The designing of network includes standard ratio of dividing datasets. Neural network at once is provided for batch training's entire patterns. The accuracy was more when back propagation method is used with neural network (BPNN), gives accuracy of 85%.

In [8] Sanjida Reza Rafa et al. Principal Component Analysis (PCA) is used to reduce the count of attributes. K-means method is used for clustering the data. The dimensionality of dataset is reduced by PCA and then combined them with heuristic algorithm for better optimization, the accuracy for decision tree is about 99.62%.

A portion of the regulated AI strategies[11] utilized in the expectation of coronary illness are artificial neural network (ANN) has model with human like neurons, decision tree (DT) with classification and regression tree, random forest (RF), support vector machine (SVM) with one, two, three, four dimensions as point, line, plane and hyperplane respectively, naïve Bayes (NB) with independency of the attributes and k nearest neighbors algorithm .It also mentions about various kind of heart diseases with respect to different attributes that are considered. Grid search and random search algorithms are used for selecting features and hence provides better results. Decision tree has 86% accuracy during testing phase and 87% during training phase. Saba Bashir et al. predicts the heart disease using data science [7]. The research is based on feature selection methods and algorithms. The main objective is to provide an accurate way of diagnosing disease with reduced feature set. The results obtained shows accuracy of 84.85% by using Logistic regression (SVM) and 84.17% for Random Forest classifier.

In [13] the model purposes AI for foreseeing the gamble of coronary failure, as ML can anticipate precisely and simply decide. The proposed framework utilizes Cleveland coronary illness dataset and information mining techniques like regression and classification and AI strategies like Random Forest, Decision tree, hybrid breed model is utilized. The exactness pace of Decision tree is 79%, for random forest is around 81%, when hybrid of these both are taken the precision is more with 88%.

6. Dataset Description

The dataset has 3 sorts of characteristics specifically key, information and forecast attributes [1]. Out of many info credits, age and orientation are nonmodifiable qualities. Patient information gathered (2011 to 2017) at cardiovascular ward of Mostar medical clinic were utilized to fabricate the model [4], the dataset contains 507 patients with cardiovascular infection, it has 71 characteristics and two specials including patient ID, Vital status. Different traits are Peripheral Oedema, HbA1c, ACE inhibitors and so forth. In [12] the Cleveland coronary illness dataset from UCI vault is utilized; the dataset has 303 cases with 76 characteristics for certain inadequate properties. Six examples were dismissed in the concentrate because of missed esteem in task sections and hence total of 279 tests. In [2] the dataset has 14 ascribes for coronary illness forecast resting electrocardiographic outcomes, resting pulse, practice incited angina is few many among them.

Table -1: Literature review of survey papers

Sl. No	Citations	Year of Publication	Paper type	Dataset	Result
1	Theresa Princy et al. [1]	2016	Conference	UCI machine Learning data repository	Accuracy of prediction was at the range of 80%
2	Sana Bharathi et al. [2]	2015	Conference	Real time data	Accuracy: 61%
3	Seyedamin Puriyed et al. [3]	2017	Conference	UCI dataset repository and Cleveland dataset	KNN(K=15)->82.8% Naive bayes->83.49% Decision Tree->77.52% SVM->69.96%
4	Damir Imamovic et al. [4]	2020	Conference	UCI machine Learning Data Repository	Decision tree-> 65.13% Neural networks-> 74.34% Logistic regression->69.74%
5	Jane Preetha Princy et al. [5]	2020	Conference	UCI data repository and Cardiovascular dataset	Decision tree-> 73% Logistic Regression->72% SVM->72% KNN->60%
6	Sagarika et al. [6]	2020	Conference	UCI Data Repository Heart disease datasets	Random Forest-> 85.71 %,
7	Saba et al. [7]	2019	Conference	UCI Repository	Decision tree->82% Logistic Regression->82.5% Random Forest->84.2% Naive Bayes->84% Logistic Regression->85%
8	Sanjida et al. [8]	2020	Conference	UCI Machine Learning Repository heart disease dataset	Decision Tree (15 attributes)-99.62% Proposed system-94.06%.
9	Khashman et al. [9]	2015	Conference	UCI machine Learning Data Repository	Decision Tree->84.33% Naïve Bayes->84% BPNN->85%
10	Kalaiselvi et al. [10]	2015	Survey	UCI machine Learning Data Repository	NB->94.43% DT->96.1% KNN->96.5% 12 attributes used NB->90.72% DT->96.62% KNN->97%
11	Polipireddy et al. [11]	2020	Survey	Cleveland dataset UCI	Chronic disease diagnosis between 82% and 92%
12	Samir et al. [12]	2020	Conference	UCI Machine Learning Repository heart disease dataset	Logistic regression->77% Naïve Bayes->81.48 KNN->94.60% SVM, MLP->80.41%
13	Sai et al. [13]	2021	Conference	UCI Machine Learning Repository heart disease dataset	Decision Tree->79% Random Forest->81% Hybrid (Decision Tree+ Random Forest)-> 88%

Table 2 Dataset sample of survey papers

SL No	Author Name	No of attributes	Dataset type
1	Theresa Princy et al. [1]	5	UCI Machine Learning Dataset
2	Sana Bharathi et al. [2]	9	Real Time Date
3	Seyedamin Puriyed et al. [3]	13	UCI Data with Cleveland dataset
4	Damir Imamovic et al. [4]	8	UCI Machine Learning Data repository
5	R. Jane Preetha Princy et al. [5]	12	UCI Cardiovascular Dataset
6	N. Komal Kumar et al. [6]	10	UCI heart disease dataset
7	Saba Bashir et al. [7]	15	UCI Machine Learning repository
8	Md. Touhidul Islam et al. [8]	15	UCI Machine Learning Dataset
9	Khashman adan [9]	270 instances; 13 attributes	UCI Machine Learning Data repository
10	C.Kalaiselvi [10]	13	UCI Machine Learning Data repository
11	Rahul Katarya et al. [11]	13	UCI Data with Cleveland dataset
12	Samir S Yadav et al. [12]	297instances;13attributes	Heart disease dataset
13	Dr. M. Kavitha et al. [13]	303instances;14attributes	Heart disease dataset

7. CONCLUSIONS

As it is very hard to forecast the coronary illness in real time, some efficient research works required to predict cardiac disease in less time. Even though many works done on this cardiac disease, still some efficient and proper research works required to predict cardiac disease using efficient data science algorithms. Developing a real time system also helps in real time to predict cardiac disease in less time and better way and doctors can handle the patients in better way

REFERENCES

[1] Thomas, J., & Princy, R. T. (2016, March). Human heart disease prediction system using data mining techniques. In 2016 international conference on circuit, power and computing technologies (ICCPCT) (pp. 1-5). IEEE.

[2] Bharti, S., & Singh, S. N. (2015, May). Analytical study of heart disease prediction comparing with different algorithms. In International Conference on Computing, Communication & Automation (pp. 78-82). IEEE.

[3] A comprehensive investigation of machine learning technique in the Domain of Heart disease-Seyedamin Puriyed, sadavahid (2017)

[4] Prediction of mortality in patients with cardiovascular disease using data mining methods – Damir Imamovic, Elmir Babovic, Nina Bijedic (2020)

[5] Princy, R. J. P., Parthasarathy, S., Jose, P. S. H., Lakshminarayanan, A. R., & Jeganathan, S. (2020, May). Prediction of cardiac disease using supervised machine learning algorithms. In 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 570-575). IEEE.

[6] N. Komal Kumar, G. Sagarika Sindhu, D. Krishna Prashanthi, A Shaeen Sultana (2020) --Analysis and prediction of Cardio-vascular disease using Machine Learning classifiers

[7] Bashir, S., Khan, Z. S., Khan, F. H., Anjum, A., & Bashir, K. (2019, January). Improving heart disease prediction using feature selection approaches. In 2019 16th international bhurban conference on applied sciences and technology (IBCAST) (pp. 619-623). IEEE.

[8] Islam, M. T., Rafa, S. R., & Kibria, M. G. (2020, December). Early prediction of heart disease using PCA and hybrid genetic algorithm with k-means. In 2020 23rd International Conference on Computer and Information Technology (ICCIT) (pp. 1-6). IEEE.

[9] Olaniyi, E. O., Oyedotun, O. K., Helwan, A., & Adnan, K. (2015, September). Neural network diagnosis of heart disease. In 2015 International Conference on Advances in Biomedical Engineering (ICABME) (pp. 21-24). IEEE.

[10] Kalaiselvi, C., & Nasira, G. M. (2015). Prediction of heart diseases and cancer in diabetic patients using data mining techniques. Indian Journal of Science and Technology, 8(14), 1.

[11] Katarya, R., & Srinivas, P. (2020, July). Predicting heart disease at early stages using machine learning: A survey. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 302-305). IEEE.

[12] Yadav, S. S., Jadhav, S. M., Nagrale, S., & Patil, N. (2020, March). Application of machine learning for the detection of heart disease. In 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA) (pp. 165-172). IEEE.

[13] Kavitha, M., Gnaneswar, G., Dinesh, R., Sai, Y. R., & Suraj, R. S. (2021, January). Heart disease prediction using hybrid machine learning model. In 2021 6th International Conference on Inventive Computation Technologies (ICICT) (pp. 1329-1333). IEEE.

[14] Roopa, C. K., Harish, B. S., & Kasturi Rangan, R. (2022). Variants of Fuzzy C-Means on MRI Modality for Cancer Image Archives. In Recent Advances in Artificial Intelligence and Data Engineering (pp. 161-169). Springer, Singapore.