

# A Review on Prediction of Cardiovascular Disease on Machine Learning Techniques

Nivyamol P Varghese  
M.Tech, CSE Dept.  
Mangalam College of Engineering  
Ettumanoor, Kottayam, India

Dr. Sahaya Kingsly  
Prof., CSE Dept  
Mangalam College of Engineering  
Ettumanoor, Kottayam, India

**Abstract**—Cardiovascular diseases are considered due to the fact the most lifestyles- threatening syndromes with the best mortality rate globally. As in step with the distinctive evaluation about eighty-four million human beings in this USA be afflicted by a few shapes of cardiovascular ailment, inflicting approximately 2, two hundred deaths a day, averaging one loss of life every forty seconds. Almost one out of each 3 deaths consequence of cardiovascular ailment. In this task, we're the use the MaLcADD framework to get the most accuracy with immoderate precision and the validation of the framework is finished thru three benchmark datasets (i.e. Framingham, Heart Disease, and Cleveland), and the accuracies of 99.1%, 98.0%, and 95.5 % are completed respectively through the use of the KNN and logistic regression method. Finally, the comparative assessment proves that MaLcADD predictions are greater accurate (with a reduced set of features) in the evaluation of the winning contemporary-day methods. Therefore, MaLcADD is exceptionally reliable and can be carried out in a real environment for the early analysis of cardiovascular diseases.

**Keywords**— *Prediction; Cardiovascular Diseases; MaLcADD framework; HRFLM; Decision tree; Random forest; K-nearest neighbors; Machine Learning.*

## I. INTRODUCTION

It is durable to find out coronary heart sickness due to numerous contributory hazard elements including diabetes, excessive blood pressure, excessive cholesterol, atypical pulse price, and plenty of different elements. Various strategies in statistics mining and neural networks were hired to discover the severity of coronary heart sickness amongst humans. The severity of the sickness is classed primarily based totally on diverse techniques like K -Nearest Neighbor Algorithm (KNN), Decision Trees (DT), Genetic set of rules (GA), and Naive Bayes (NB). The nature of coronary heart sickness is complicated and hence, the sickness should be treated carefully. Not doing so may also influence the coronary heart or purpose untimely death. The angle of clinical technology and statistics mining is used for coming across diverse kinds of metabolic syndromes. Data mining with type performs a giant position withinside the prediction of coronary heart sickness and statistics investigation.

We have various techniques that were used for information abstraction with the aid of using the use recognized techniques of statistics mining for the prediction of coronary heart sickness. In this work, several readings were achieved to provide a prediction version of the use of now no longer most effective awesome based on the related

strategies however additionally with the aid of using touching on or extra strategies. These amalgamated new strategies are typically referred to as hybrid techniques. We introduce neural networks with the use of coronary heart price time series. This approach makes use of diverse medical data for prediction including Left package deal department block (LBBB), Right package deal department block (RBBB), Atrial fibrillation (AFIB), Normal Sinus Rhythm (NSR), Sinus bradycardia (SBR), Atrial flutter (AFL), Premature Ventricular Contraction (PVC), and Second diploma block (BII) to discover the precise circumstance of the affected person about coronary heart sickness. The dataset with a radial foundation characteristic community (RBFN) is used for type, in which 70% of the statistics are used for schooling and the last 30% is used for type. We additionally introduce a Computer-Aided Decision Support System (CADSS) withinside the discipline of medication and studies. In preceding work, the use of statistics mining strategies withinside the healthcare enterprise has been proven to take much less time for the prediction of sickness with extra correct consequences. We recommend the prognosis of coronary heart sickness the use of the GA. For experimental validation, we use the famous Cleveland dataset that is accumulated from a UCI device to get to know the repository. We will see in a while how our consequences show to be distinguished while in comparison to several of the recognized supervised getting to know strategies. The maximum effective evolutionary set of rules Particle Swarm Optimization (PSO) is delivered, and a few policies are generated for coronary heart sickness. The policies were carried out randomly with encoding strategies which bring about the development of the accuracy overall. Heart sickness is anticipated primarily based totally on signs namely, pulse price, sex, age, and plenty of others. Over a duration of time, they have ended up very common area and are overstretching the healthcare systems of nations. The results of coronary heart ailment and hazard elements may also display up in people as raised blood pressure, raised blood glucose, raised blood lipids, and obese and obesity. These “intermediate dangers elements” may be measured in number one care centers and imply an extended hazard of coronary heart attack, stroke, coronary heart failure, and different complications. The ML set of rules with Neural Networks is delivered, whose consequences are extra correct and dependable as we've visible in.

The common challenges in the prediction of cardiovascular diseases are improving accuracy and missing values in the data that highly affect the accuracy of the model. In addition to that, the problem of class imbalance. Similarly, features of the dataset largely affect the accuracy and computational complexity of the machine learning process.

The rest of the paper is organized in the following way. Section I represents an Introduction. Section II presents a literature review on the prediction of cardiovascular disease approaches. Section III provides the details of previous research models, performances, and their comparison. Section IV describes the conclusion.

## II. LITERATURE SURVEY

There is ample related work in the fields directly related to this paper. First, we are introducing the Ambedkar and Phalinkar paper in 2018, [1]. In that paper, Data evaluation performs a sizable position in dealing with a big quantity of information within the healthcare. The preceding scientific research was primarily based totally on dealing with and assimilating a massive quantity of medical institution information rather than prediction. Due to a considerable quantity of information increasing within the biomedical and healthcare subject the correct evaluation of scientific information turns into propitious for in advance detection of sickness and affected person care. However, the accuracy decreases while the scientific information is partly lacking. To conquer the hassle of lacking scientific information, we carry out information cleansing and imputation to convert the unfinished information to finish information. We are operating on coronary heart sickness prediction on the premise of the dataset with the assistance of Naïve Bayes and the KNN set of rules. To amplify this work, we endorse the sickness hazard prediction and the use of based information. We use a convolutional neural network-based unimodal sickness hazard prediction set of rules. The prediction accuracy of the CNN-UDRP set of rules reaches extra than 65%. Moreover, this machine solves the query associated with sickness that human beings face in their life.

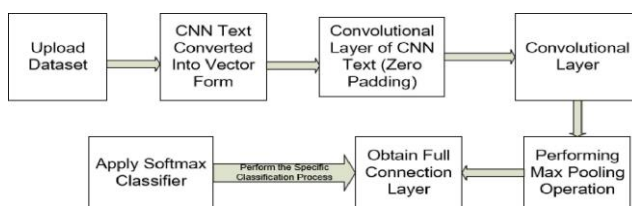


Fig 1. Block diagram of CNN-UDRP, Ambedkar, and Phalinkar [1]

The secondary related work is one of the famous papers published in 2018, [2] written by Mienye. The prediction of coronary heart ailment is a project in medical device gaining knowledge. Early detection of humans susceptible to the ailment is essential in stopping its progression. This paper proposes a deep learning model to gain knowledge of methods to reap advanced prediction of a coronary heart ailment. A greater stacked sparse autoencoder community (SSAE) is evolved to reap green function gaining knowledge. The community includes more than one sparse autoencoder and a SoftMax classifier. Additionally, in deep learning knowledge of models, the algorithm's parameters want to be optimized accurately to reap green overall performance. Hence, we suggest a particle swarm optimization (PSO) primarily based approach to optimize the parameters of the stacked sparse autoencoder. Optimization via way of means of the PSO improves the function gaining knowledge of and type overall performance of the SSAE. Meanwhile, the multilayer structure of autoencoders generally ends in an inner covariate shift, trouble that influences the generalization capacity of the community; hence, batch normalization is delivered to save you this trouble. The experimental effects display that the proposed technique efficiently predicts coronary heart ailment via way of means of acquiring a typing accuracy of 0.973 and 0.961 at the Framingham and Cleveland coronary heart ailment datasets, respectively, thereby outperforming different devices gaining knowledge of strategies and comparative studies.

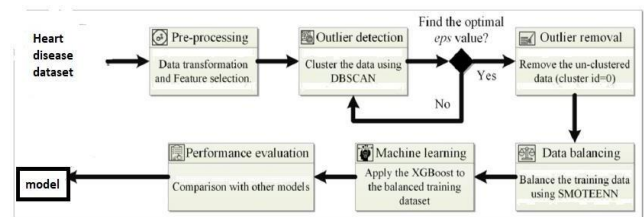


Fig 2. Block diagram of XG boost classifier, Mienye [2]

In 2020, [3] Samira diouzi research has been made on machine learning-based identification of patients with Cardiovascular sicknesses was for a long term one of the important clinical problems. As indicated by the World Health Association, coronary heart illnesses are the best factor of the ten main motives for death. Correct and early identity is an essential step in rehabilitation and treatment. To diagnose coronary heart defects, it might be vital to enforce a machine capable of expecting the lifestyles of coronary heart sicknesses. In the contemporary article, our important motivation is to increase a powerful clever clinical machine primarily based totally on device mastering techniques, to resource in figuring out an affected person's coronary heart circumstance and manual a medical doctor in making a correct prognosis of whether or not or now no longer an affected person has cardiovascular sicknesses. Using more than one information processing technique, we cope with the hassle of lacking information in addition to the hassle of imbalanced information within the publicly to be had UCI Heart Disease dataset and the Framingham dataset. Furthermore, we use device mastering to pick out the simplest set of rules for predicting cardiovascular sicknesses. In this work, victimization accommodative adaptive neuro-fuzzy inference system (ANFIS), We have tried to predict this unhealthiness. The speed and therefore the validity of the prompt formula is more than the opposite sensible strategies used. The strategy planned during this article, with a 10% validity increase throughout experimentation incorporates a better performance than previous sensible strategies. Other strategies for the detection of this illness are planned, together with strategies based on computing like fuzzy algorithms for pattern recognition in characteristic extraction, SVM formula, genetic formula, RBF neural networks, MLP neural networks, and strategies supported by the Bayesian model. A brand new technique for predicting polygenic disease, victimization accommodative Neural Fuzzy logical thinking system (ANFIS), is planned. Development and analysis of the model have been performed using real knowledge sets. Mathematician membership functions and hybrid algorithms have been used for network coaching. The required model was created by making ready the desired data concerning polygenic disease from existing information consistent with the results, the model is ready to predict polygenic disease with success. The strategy planned during this study is taken into account a brand new approach to the diagnosing of this unhealthiness that's quicker and additional correct than previous strategies methods. Different metrics, which include accuracy, sensitivity, F-measure, and precision, had been used to check our machine, demonstrating that the proposed technique is appreciable.

The last two papers are based on comprehensive strategy studies written by Wilson and sarulatha. Wilson papers were designed and executed within the MEDLINE, Embase, and Scopus databases from database inception through March 15, 2019, [4]. The primary outcome was a composite of the predictive ability of ML algorithms for coronary artery disease, heart failure, stroke, and cardiac arrhythmias. Of 344 total studies identified, 103 cohorts, with a total of 3,377,318 individuals, met our inclusion criteria. For the prediction of coronary artery disease, boosting algorithms had pooled area under the curve (AUC) of 0.88 (95% CI 0.84–0.91), and custom-built algorithms had a pooled AUC of 0.93 (95% CI 0.85–0.97). For the prediction of stroke, support vector machine (SVM) algorithms had a pooled AUC of 0.92 (95% CI 0.81–0.97), boosting algorithms had pooled AUC of 0.91 (95% CI 0.81–0.96), and convolutional neural network (CNN) algorithms had a pooled AUC of 0.90 (95% CI 0.83–0.95) same we are compared to the sarulatha most of the algorithms also [5].

#### A. DATASETS VARIABLE USED FOR PREDICTION OF CARDIOVASCULAR DISEASES

##### Predictable attribute

1. Diagnosis (value Heal: < 50% diameter narrowing (no heart disease); value Sick: > 50% diameter narrowing (has heart disease))

##### Key attribute

1. PatientID- Patient's identification number

##### Input attributes

- Sex (value 1: Male; value 0: Female)
- Chest Pain Type (value 1: typical type 1 angina, value typical type angina, value 3: non-angina pain; value asymptomatic)
- Fasting Blood Sugar (value 1: > 120 mg/dl; value 0: < 120 mg/dl)
- Restecg - resting electrographic results (value 0: normal; value 1: having reversible defect)
- Trest Blood Pressure (mm Hg on admission to the hospital)
- Serum Cholesterol (mg/dl)
- Thalach - maximum heart rate.

##### Predictable attribute

##### Diagnosis

- Value Healthy: No heart disease value
- Sick: has Heart disease

##### Reduced Input attributes:

- Type - Chest Pain Type
- Rbp - Resting blood pressure

- Eia - Exercise-induced angina
- Oldpk - Old peak
- Vsl - No. of vessels colored
- Thal - Maximum heart rate achieved

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	condition
69	1	0	160	234	1	2	131	0	0.1	1	1	0	0
69	0	0	140	239	0	0	151	0	1.8	0	2	0	0
66	0	0	150	226	0	0	114	0	2.6	2	0	0	0
65	1	0	138	282	1	2	174	0	1.4	1	1	0	1
64	1	0	110	211	0	2	144	1	1.8	1	0	0	0
64	1	0	170	227	0	2	155	0	0.6	1	0	2	0
63	1	0	145	233	1	2	150	0	2.3	2	0	1	0
61	1	0	134	234	0	0	145	0	2.6	1	2	0	1
60	0	0	150	240	0	0	171	0	0.9	0	0	0	0
59	1	0	178	270	0	2	145	0	4.2	2	0	2	0
59	1	0	170	288	0	2	159	0	0.2	1	0	2	1
59	1	0	160	273	0	2	125	0	0	0	0	0	1
59	1	0	134	204	0	0	162	0	0.8	0	2	0	1
58	0	0	150	283	1	2	162	0	1	0	0	0	0
56	1	0	120	193	0	2	162	0	1.9	1	0	2	0
52	1	0	118	186	0	2	190	0	0	1	0	1	0
52	1	0	152	298	1	0	178	0	1.2	1	0	2	0
51	1	0	125	213	0	2	125	1	1.4	0	1	0	0
45	1	0	110	264	0	0	132	0	1.2	1	0	2	1
42	1	0	148	244	0	2	178	0	0.8	0	2	0	0
40	1	0	140	199	0	0	178	1	1.4	0	0	2	0
38	1	0	120	231	0	0	182	1	3.8	1	0	2	1
34	1	0	118	182	0	2	174	0	0	0	0	0	0
74	0	1	120	269	0	2	121	1	0.2	0	1	0	0
71	0	1	160	302	0	0	162	0	0.4	0	2	0	0

Fig 3. Cleveland Dataset

### III . . COMPARISON AND DISCUSSION OF PREVIOUS RESEARCH

Methods	Datasets	Advantages	Disadvantages
HRFLM Approach	Cleveland dataset	Model emphasis on distinct combinations of parameters with classification techniques of ML.	Accuracy is less than 90% for the results.
RF algorithm	Cleveland database	Predict cardiovascular diseases with a small number of attributes.	Most difficult to reach results via algorithms.
CNN and Adaboost	CHDD dataset	CNN algorithm has achieved the highest accuracy. Deep learning, it can be extended with feature selection algorithms.	High error values and hard to evaluate the interpretation.
Yolo v5 and deep learning	The original data contains x-ray images	A visual VFM analysis method based on YOLO Model and Continuity Equation is designed. Positioning and tracking of the myocardial wall are realized based on the YOLO model and IBM. The proposed VFM method provides a new basis for the evaluation of cardiac function.	The results of the experiments on the evaluation of the Ultrasonic apical long-axis view show that the proposed method only improves the accuracy of VFM but also provides a new evaluation basis for cardiac function impairment.
Naive Bayes, SVM	CHDD dataset	A comparative analysis of various classification algorithms on the Cleveland dataset has been performed with good accuracy. This work has used ensemble algorithms such as bagging, boosting, and voting.	Feature selection has interpretation is difficult to find out and this leads to more chances to find out the more algorithms that implies the results have only a small improvement compared to another one.

Table 1: Comparative study of prediction of cardiovascular diseases based on previous research.

In this prediction models are developed with certain features, and it compares the accuracy, class error, precision, F-measure, sensitivity, and specificity. The maximum accuracy is completed through the HRFLM class technique in assessment with present techniques.

The UCI dataset is further classified into 8 types of datasets based on classification. Each dataset is further classified and processed by python coding. The results are generated by applying the classification rule for the dataset. The classification rules are generated based on the rule after data pre-processing is done. After pre-processing, the data's three best ML techniques are chosen, and the results are generated. The various datasets with DT, RF, and LM are applied to find out the best classification method. The results show that RF and LM are the best. The error rate for dataset 4 is high (20.9%) compared to the other datasets. The LM method for the dataset is the best (9.1%) compared to DT and RF methods. We combine the RF method with LM and propose a hybrid method to improve the results.

### IV. CONCLUSION

Specifying the processing of herbal healthcare facts of coronary heart facts will assist within the long time Saving of human lives and early detection of abnormalities in coronary heart conditions. Machine getting to know Techniques had been utilized in this to clear out facts and deliver a substitute and novel knowledge Towards coronary heart situation. Heart ailment projection is worrying and enormous within the medical. However, the mortality ratio may be considered regulated if the ailment is discovered at the primary Stages and preventive standards are normal as quickly as feasible. Further extension of this look could be very suitable to direct the investigations to real-international datasets in place of simply theoretical Approaches and simulations. The proposed hybrid HRFLM technique is used by combining the traits of Random Forest (RF) and Linear Method (LM). HRFLM proved to be correct within the prediction of coronary heart sickness. The destiny route of this study may be per- fashioned with various combos of devices getting to know strategies to higher prediction strategies. Furthermore, new feature- techniques may be evolved to get a broader perception of the sizable functions to boom the overall performance of coronary heart sickness prediction.

### REFERENCES

- [1] S. Ambedkar and R. Phalnikar, "Disease Risk Prediction by Using Convolutional Neural Network", Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018.
- [2] Ibomoiye I Mienye and Yanxia Sun, "Improved Heart Disease Prediction Using Particle Swarm Optimization Based Stacked Sparse Autoencoder", IEEE Access, vol. 6, 2018, pp.1-5.
- [3] Samira diouzi and B.S. Saini, "Detection of coronary heart disease by reduced features and extreme learning machine", Medicine and Pharmacy Reports, vol.91, pp.166-175, 2020.



- [4] W.H. Wilson Tang and Zhen Wang, “*Machine learning prediction in cardiovascular diseases: a meta-analysis*”, Springer, Charm, 2020.
- [5] M. Swathy and K. Saruladha, “*A comparative study of classification and prediction of Cardio-Vascular Diseases(CVD)using Machine Learning and Deep Learning techniques*”, in proc.11<sup>th</sup> Int.Conf. Hum. Syst.Interact.(HSI), July.2021, pp.1-8.
- [6] D. C. Yadav and S. Pal, “*Prediction of heart disease using feature selection and random forest ensemble method*,” Int. J. Pharmaceutical Res., vol. 12, no. 4, 2020.
- [7] N. Kumar and K. Sikamani, “*Prediction of chronic and infectious diseases using machine learning classifiers—A systematic approach*,” Int. J. Intell. Eng. Syst., vol. 13, no. 4, pp. 11–20, 2020.
- [8] K. Uyar and A. Ilhan, “*Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks*,” Procedia Comput. Sci., vol. 120, pp. 588–593, Jan. 2017.
- [9] L.Wang, W.Zhou, “*Deep ensemble detection of congestive heart failure using short-term RR intervals*,” IEEE Access, vol. 7, pp. 69559–69574, 2019.
- [10] G. O. Rashmi and U. M. A. Kumar, “*Machine learning methods for heart disease prediction*,” Int. J. Eng. Adv. Technol., vol. 8, no. 5S, pp. 220–223, May 2019.
- [11] F. Miao, Y.-P. Cai, Y.-X. Zhang, X.-M. Fan, and Y. Li, “*Predictive modeling of hospital mortality for patients with heart failure by using an improved random survival forest*,” IEEE Access, vol. 6, pp. 7244–7253, 2018.
- [12] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, “*A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms*,” Mobile Inf. Syst., vol. 2018, pp. 1–21, Dec. 2018.
- [13] W. Wiharto, H. Kusnanto, and H. Herianto, “*Hybrid system of tiered multivariate analysis and artificial neural network for coronary heart disease diagnosis*,” Int. J. Electr. Comput. Eng., vol. 7, no. 2, p. 1023, Apr. 2017.
- [14] C. A. Cheng and H. W. Chiu, “*An artificial neural network model for the evaluation of carotid artery stenting prognosis using a national-wide database*,” in Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Jul. 2017, pp. 2566–2569.
- [15] A. A. Shetty and C. Naik, “*Different data mining approaches for predicting heart disease*,” Int. J. Innov. Sci. Eng. Technol., vol. 5, pp. 277–281, May 2016.
- [16] K. Padmavathi and K. S. Ramakrishna, “*Classification of ECG signal during atrial fibrillation using autoregressive modeling*,” Procedia Comput. Sci., vol. 46, pp. 53–59, Jan. 2015.
- [17] A. M. D. Silva, Feature Selection, vol. 13. Berlin, Germany: Springer, 2015, pp. 1–13.
- [18] A. M. De Silva and P. H. W. Leong, *Grammar-Based Feature Generation for Time-Series Prediction*. Berlin, Germany: Springer, 2015.