

## A Survey on Heart Attack Prediction Using Machine Learning

P.Santhi <sup>a</sup>, R.Ajay<sup>b</sup>, D.Harshini <sup>c</sup> and S.S.Jamuna Sri <sup>d</sup>

<sup>a</sup> Department of Computer Science and Engineering , M.Kumarasamy College of Engineering, Thalavapalayam, Karur-639119

<sup>b,c,d</sup> Department of Computer Science and Engineering ,M.Kumarasamy College of Engineering, Thalavapalayam, Karur-639113

**Article History:** Received: 11 January 2021; Accepted: 27 February 2021; Published online: 5 April 2021

**Abstract:** Machine Learning Algorithm is used for many different diseases. Machine Learning is a learning of machine by own itself. And it is a part of AI that deals with to learn a machine according to their own. Now-a-days most are affected due to Heart attack it becomes head ache for doctors. In order to reduce the count of death we need to predict the Heart attack. For this problem Machine Learning play a major role in this paper. This prediction takes a people from the danger zone of their life. In this paper we use KNN algorithm and Random forest algorithm can predict the heart attack in advance.

**Keywords:** Machine Learning, Heart attack prediction, KNN algorithm and Random Forest algorithm, React and Flask Framework

### 1. Introduction

A big amount of complex data information in the healthcare Industry today about patients, medical devices, hospitals resources, and electronic patients records etc. Anyway large amount of data is useful for predict the data analysis and its used for minimize the cost[1]. As per WHO many heart Attack are come for people in abroad. In initial Stage The diseases is identifying correctly check in the vital time. It is really a head pain for the doctor's[5]. Many scientific technologies are adopting doctors now-a-days there are many technology machines are very useful for predicting many diseases [33]. Sometimes doctors may fail to take accurate decisions for heart attack prediction system which is use machine learning such as to get accurate results[6]. Important attributes like, age, gender, blood pressure, etc to predict the heart attack. Many heart strokes are happening because of accumulation of cholesterol in blood vessels or blood clot in blood vessels in arteries which supply blood to the heart muscles. External and internal image of heart figure 1 and figure 2 are given below.

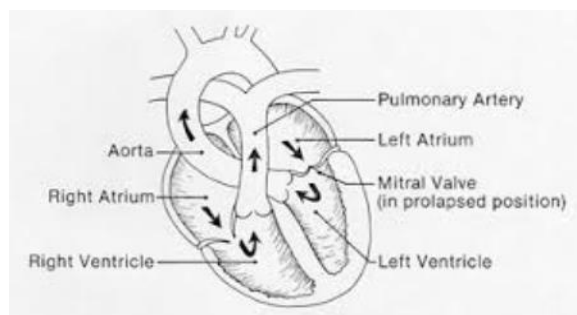


Figure 1. Internal view of the heart

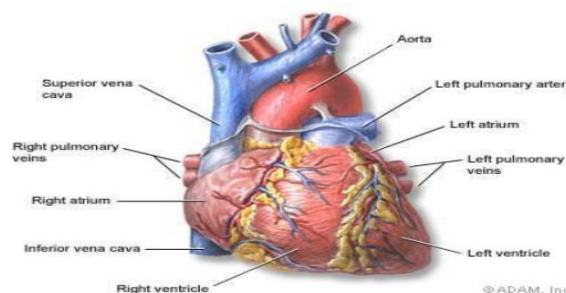


Figure 2. External view of the heart

### 2. Machine Learning Algorithms

In this paper we our focus is how we can train the Machine to learn from the medical data so it can predict and treat the disease[43]. Through Experience process of acquiring knowledge in called learning. From the minute of boring we have to start the process of getting knowledge[50]. The process of Learning will come out through our experience Machine Learning is the process of Learning and training the machine by own itself [21]. Machine

Learning has some statistics, Human psychology and brain modelling[12]. Neural models are used for understanding the human brain[9]. ML algorithms need a dataset, which is collection of records are instances where each instance consist of attributes[20].

The classifications of Machine Learning are Supervised Learning and Unsupervised Learning. The supervised Learning usually work with labelled data to learn mapping function that's turn input into output variable. The unsupervised Learning models work with the input data and not the output data for any given sample unlabelled data is used to model the data.

Algorithms with their accuracies are:

- Random Forest; 86.89% with 200, 500 & 1000 estimators
- K-Nearest Neighbours; 91.8% with 8 neighbours

### 3. KNN Algorithm

KNN means K Nearest Neighbours and it is very easy algorithm that give more practice can keep available data and give a new data[10]. The neighbour points is similar to the sample that added to the new point. KNN algorithm is used to website for searching similarities items.[31].The number of nearest is to be predict that needed in the algorithm if KNN algorithm of the K neighbours of the new point.KNN algorithm are otherwise defined as lazy learner because the ability of the model learning is pretty fast. It will learn through the practice and it will memorize the data very well.

This clustering algorithm, is one of the simplest and also a popular unsupervised algorithm. In other words, the K-means algorithm denotes k number of centroids, and then assigns every data point to the nearest cluster, while keeping the centroids as tiny as possible in the result. This K-means algorithm is mostly used for the classification process[36].

#### 3.1 How does the Algorithm Work

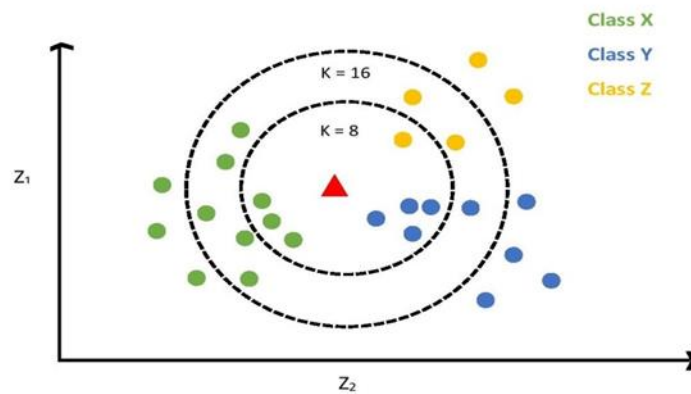


Figure 3:KNN Algorithm

First we need to select the value of K to predict the algorithm [37]. Based to figure 1,When the point is in the blue color that define as class y, points in in the green color define as class x, points is in yellow color defines as class z. When the k value is equal to neighbour points that represent triangle[18]. Based on the figure 1 when k=8 new points is near to on yellow point, four blue point, and three green point. We have to major of blue point, In this case, we can say that for k=8 is a new points is that belongs to class y.

If the k points is near to 16 we have to take different 16 points that near to closest point. We can test the several values of K with the technique of cress-validation in order to find the excellent k value. To find the low distance between neighbour point[27], Use either Euclidean distance otherwise Manhattan distance. In the Manhattan Distance will calculate the real vectors and the sum of their difference. Workflow of KNN algorithm describes in figure 3.

### 4. Data Collection

We can get the dataset from the Kaggale.com to predict the heart attack [2]. There are three classification of data's they are factual information's. And data in the dataset can be divided into two types categorical data and number of data. The attributes of dataset are given below table no 1.

Table no : 1 [Data set]

Categorical Data	Feature	Description
	Age days	Factual Information   age in days   int (days)
	Age year	Factual Information   age in Years   int (days)
	Height	Factual Information   Height   int (cm)
	Weight	Factual Inormation   weight   float (kg)

	ap_hi	blood pressure of Systolic   Examination Feature   int
	ap_lo	blood pressure of Diastolic  Examination Feature   int
Numerical Data	Gender	Factual Information 2:male ,1::female
	Cholesterol	Cholesterol   Examination Feature   Cholestrol
	glue	Glucose  Examination of Feature   glue
	smoke	Smoking's  subjective of Feature  Smoke's binary's
	alco	Alcohol of intake  Subjective of Features alco binary
	active	Physical activity  subjective of Feature active's binary's
	cardiology	Presence ortherwise absence of cardio disease's Target variable  cardio binary's

### 5. Biological Attributes

Features used for training machine learning models on, including the special binary class label target, describing whether heart disease was detected.

age: Age in years

ca: Classification of major blood vesselscular (0-3)

Chol: Serum of cholesterol's in mg/dl

Cp: Chest pain classification

Value 1: Typical angina's

Value 2: atypical anginas

Value 3: non-anginal pains

Value 4: Asymptomatic

Obs: (1 = true; 0 = no) fasting of blood sugar > 120 mg/dl

Old peak: ST depression's have also included by the exercise relative's to rest

Restecg: Resting the electro cardio graphic results

Value 0: Normal's

Value 1: Having ST-T wave's abnormality

Value 2: Showing probablelity

Sex: (1 = male; 0 = female)Sex

Slope: exercise of ST segment

Value 1: Up sloping's

Value 2: Flat's

Value 3: Down sloping's

Target: (0 = disease; 1 = no disease)Heart disease detection

Thal: Thallium stress test result

Value 3: normals

Value 6: fixed defects

Value 7: reversible defects

Thalach: Max of heart rate achieved in bpm

Trestbps: blood pressure's resting (in mmHg on admission to the hospital)

### 6. Random Forest Algorithm

Random forest algorithm is a supervised machine learning algorithm, which is helpful for both classification and regression problems. Similarly, with the help of the data samples, random forest algorithm creates the decision trees. With the samples from those decision trees the best solution will be predicted [36].

Random forest algorithm is more flexible and also easy to use algorithm that produces result without hyper-parameter tuning, which produces the greatest of all time. It is also one of the most used algorithms, because of its simplicity [6]. Random Forest algorithm can be considered to be one among bagging techniques and not boosting techniques. The random forests trees usually run in parallel. The trees in boosting algorithms will be trained sequentially [36].

### 7. React Framework

Compostable user interfaces is build by React library and it encourage creation of reusable UI components in which current data that changes over time. React is used as V in MVC and it abstracts away the DOM from you, offering a easier programming model and better performance[4]. It can also render on the server using Node, and powers native apps using React Native React is easier about traditional data binding and it is reduces the boiler plate.

### 8. Flask Framework

The web application framework is Know as Flask and it is written in Python. This is founded by Armin Ronacher; he is leads to python enthusiasts named Pocco. Flask is an refers to micro framework. It does not have

built-in function for database handling. It supports the extensions to add function to the application. The difference between the existing system and proposed system are given below table no 2,

Table no 2: Comparison table

Existing System	Proposed System
1. In the existing systems the prediction is done by Machine Learning .	1. Here the, prediction is done by Machine Learning using KNN algorithm and Random Forest algorithm .
2. The technologies that are used in the existing systems are Programming Languages.	2. The technologies that are used here is React and Flask.
3. Since the IOT devices are used it is not portable.	3. It is portable because the data are being stored in the server.
4. In the existing system only some of the parameters are being checked.	4. In the proposed system all the parameters will be checked so that it is more efficient.
5. The maintenance cost will be high because of the some hardware devices.	5. Here the maintenance cost is low when compared.

### 9. Case Study

Scope of this research is to predict whether a patient is affected by heart disease or not. This survey was done on supervised machine learning classification techniques using Random forest, and KNN algorithm. The Source Data for training the model are bootstrapped to generate a random large number of trees. Evaluation of predictor variables is done by how often they make successful predictions or how much they decrease node impurity when they are selected for the splits[36].

### 10. Conclusion

The major concern for the society is Heart Disease .It is hard to determine manually the odds of getting heart diseases based on risk factors like age etc. The useful technology used to predict the output from data is Machine Learning technology. In this survey, we have designed a system which is suited for real-time heart disease prediction and can be used by the users who have coronary disease[47]. The diagnosis system of the system is able to predict the heart disease by using ML algorithms and the prediction results are based on the heart disease dataset instance[44]. For checking the variances and raising the alarm if the user's heart rate rises than the normal rate of the heart. To prove the effectiveness of the system we have carried out experiments for both monitoring and diagnosis systems[50].

### References

1. Sonam Nikhar, A.M. Karandikar "Prediction of Heart Disease Using Machine Learning Algorithms" in International Journal of Advanced Engineering, Management and Science (IJAEMS) June2016 vol-2
2. Deeanna Kelley "Heart Disease: Causes, Prevention, and Current Research" in JCCC Honors Journal
3. Nabil Alshurafa, Costas Sideris, Mohammad Pourhomayoun, Haik Kalantarian, Majid Sarrafzadeh "Remote Health Monitoring Outcome Success Prediction using Baseline and First Month Intervention Data" in IEEE Journal of Biomedical and Health Informatics
4. Murugesan, M., Thilagamani, S. ," Efficient anomaly detection in surveillance videos based on multi layer perception recurrent neural network", Journal of Microprocessors and Microsystems, Volume 79, Issue November 2020, <https://doi.org/10.1016/j.micpro.2020.103303>
5. DhafarHamed, Jwan K. Alwan, Mohamed Ibrahim, Mohammad B. Naeem "The Utilisation of Machine Learning Approaches for Medical Data Classification" in Annual Conference on New Trends in Information & Communications Technology Applications - march2017
6. Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients Mai Shouman, Tim Turner, and Rob Stocker International Journal of Information and Education Technology, Vol. 2, No. 3, June 2012
7. Amudhavel, J., Padmapriya, S., Nandhini, R., Kavipriya, G., Dhavachelvan, P., Venkatachalapathy, V.S.K., "Recursive ant colony optimization routing in wireless mesh network", (2016) Advances in Intelligent Systems and Computing, 381, pp. 341-351.
8. Thilagamani, S., Nandhakumar, C. ." Implementing green revolution for organic plant forming using KNN-classification technique", International Journal of Advanced Science and Technology, Volume 29 , Issue 7S, pp. 1707–1712

9. Amudhavel, J., Inbavalli, P., Bhuvaneshwari, B., Anandaraj, B., Vengattaraman, T., Premkumar, K., "An effective analysis on harmony search optimization approaches", (2015) *International Journal of Applied Engineering Research*, 10 (3), pp. 2035-2038.
10. Amudhavel, J., Kathavate, P., Reddy, L.S.S., Bhuvaneshwari Aadharshini, A., "Assessment on authentication mechanisms in distributed system: A case study", (2017) *Journal of Advanced Research in Dynamical and Control Systems*, 9 (Special Issue 12), pp. 1437-1448.
11. Amudhavel, J., Kodeeshwari, C., Premkumar, K., Jaiganesh, S., Rajaguru, D., Vengattaraman, T., Haripriya, R., "Comprehensive analysis on information dissemination protocols in vehicular ad hoc networks", (2015) *International Journal of Applied Engineering Research*, 10 (3), pp. 2058-2061.
12. Amudhavel, J., Kathavate, P., Reddy, L.S.S., Satyanarayana, K.V.V., "Effects, challenges, opportunities and analysis on security based cloud resource virtualization", (2017) *Journal of Advanced Research in Dynamical and Control Systems*, 9 (Special Issue 12), pp. 1458-1463.
13. Thilagamani, S., Shanti, N., "Gaussian and gabor filter approach for object segmentation", *Journal of Computing and Information Science in Engineering*, 2014, 14(2), 021006, <https://doi.org/10.1115/1.4026458>
14. B.L Deekshatulu Priti Chandra "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm ", M.Akhil jabbar *International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013*.
15. S. Shilaskar and A. Ghatol, "Feature selection for medical diagnosis: Evaluation for cardiovascular diseases," *Expert Syst. Appl.*, vol. 40, no. 10, pp. 4146–4153, Aug. 2013.
16. American Cancer Society. *Breast Cancer Facts & Figures 2005-2006*. Atlanta: American Cancer Society, Inc. (<http://www.cancer.org/>).
17. N. Satyanarayana, CH. Ramalingaswamy, and Y. Ramadevi, 2014. Survey of Classification Techniques in Data Mining, *International Journal of Innovative Science, Engineering & Technology*, Vol. 1 Issue 9, November 2014.
18. C. Christin, H. C. Hoefsloot, A. K. Smilde, B. Hoekman, F. Suits, R. Bischoff, and P. Horvatovich, "A critical assessment of feature selection methods for biomarker discovery in clinical proteomics, " *Molecular & Cellular Proteomics*, vol. 12, no. 1, pp. 263-276, 2013.
19. High blood pressure prediction based on AAA++ using machine-learning algorithms, Satyanarayana Nimmala, Y. Ramadevi, R. Sahith & Ramalingaswamy Cheruku, *Cogent Engineering* (2018), 5: 1497114.
20. Rhagini, A., Thilagamani, S. ,"Women defence system for detecting interpersonal crimes", *International Journal of Advanced Science and Technology*, 2020, Volume 29, Issue 7S, pp. 1669–1675
21. Avinash Golande, Pavan Kumar T, Heart Disease Prediction Using Effective Machine Learning Techniques, *International Journal of Recent Technology and Engineering*, Vol 8, pp.944-950, 2019.
22. T.Nagamani, S.Logeswari, B.Gomathy, Heart Disease Prediction using Data Mining with Mapreduce Algorithm, *International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075*, Volume-8 Issue-3, January 2019.
23. P. Pandiaraja, N Deepa 2019 ,” A Novel Data Privacy-Preserving Protocol for Multi-data Users by using genetic algorithm” , *Journal of Soft Computing* , Springer , Volume 23 ,Issue 18, Pages 8539-8553
24. Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin, Design And Implementation Heart Disease Prediction Using Naives Bayesian, *International Conference on Trends in Electronics and Information(ICOEI 2019)*.
25. N Deepa , P. Pandiaraja, 2020 ,” Hybrid Context Aware Recommendation System for E-Health Care by merkle hash tree from cloud using evolutionary algorithm” , *Journal of Soft Computing* , Springer , Volume 24 ,Issue 10, Pages 7149–7161
26. Nagaraj M Lutimath, Chethan C, Basavaraj S Pol., Prediction Of Heart Disease using Machine Learning, *International journal Of Recent Technology and Engineering*, 8,(2S10), pp 474-477, 2019.
27. N Deepa , P. Pandiaraja, 2020 , “ E health care data privacy preserving efficient file retrieval from the cloud service provider using attribute based file encryption “, *Journal of Ambient Intelligence and Humanized Computing* , Springer , <https://doi.org/10.1007/s12652-020-01911-5>.
28. Sayali Ambekar, Rashmi Phalnikar, Disease Risk Prediction by Using Convolutional Neural Network, 2018 Fourth International Conference on Computing Communication Control and Automation.
29. C. B. Rjeily, G. Badr, E. Hassani, A. H., and E. Andres, Medical Data Mining for Heart Diseases and the Future of Sequential Mining in Medical Field, in *Machine Learning Paradigms*, 2019, pp. 7199.

30. Vijayakumar, P, Pandiaraja, P, Balamurugan, B & Karuppiah, M 2019, 'A Novel Performance enhancing Task Scheduling Algorithm for Cloud based E-Health Environment', International Journal of E-Health and Medical Communications (IJEHMC), Vol 10, Issue 2, pp 102-117
31. Fajr Ibrahim Alarsan., and Mamoon Younes Analysis and classification of heart diseases using heartbeat features and machine learning algorithms, Journal Of Big Data, 2019; 6:81.
32. Vijayakumar, P, Pandiaraja, P, Karuppiah, M & Deborah, LJ 2017, 'An Efficient Secure Communication for Healthcare System using Wearable Devices', Journal of Computers and Electrical Engineering, Elsevier, Vol .No 63, October 2017, pp 232-245
33. Decanna Kelley "Heart Disease: Causes, Prevention, and Current Research" in JCCC Honors Journal
34. K Sumathi, P Pandiaraja 2019, "Dynamic alternate buffer switching and congestion control in wireless multimedia sensor networks", Journal of Peer-to-Peer Networking and Applications, Springer, Volume 13, Issue 6, Pages 2001-2010
35. Ponrathi Athilingam, Bradlee Jenkins, Marcia Johansson, Miguel Labrador "A Mobile Health Intervention to Improve Self-Care in Patients With Heart Failure: Pilot Randomized Control Trial" in JMIR Cardio 2017, vol. 1, issue 2, pg no:1
36. Deepa. K, LekhaSree. R, Renuga Devi. B, Sadhana. V, Virgin Jenifer. S, "Cervical Cancer Classification", International Journal of Emerging Trends in Engineering Research, 2020, 8(3), pp. 804–807 <https://doi.org/10.30534/ijeter/2020/32832020>
37. Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients Mai Shouman, Tim Turner, and Rob Stocker International Journal of Information and Education Technology, Vol. 2, No. 3, June 2012
38. Amudhavel, J., Padmapriya, S., Nandhini, R., Kavipriya, G., Dha-vachelvan, P., Venkatachalapathy, V.S.K., "Recursive ant colony optimization routing in wireless mesh network", (2016) Advances in Intelligent Systems and Computing, 381, pp. 341-351.
39. Santhi, P., Lavanya, S., Prediction of diabetes using neural networks, International Journal of Advanced Science and Technology, 2020, 29(7 Special Issue), pp. 1160–1168.
40. Amudhavel, J., Inbavalli, P., Bhuvaneshwari, B., Anandaraj, B., Vengattaraman, T., Premkumar, K., "An effective analysis on harmony search optimization approaches", (2015) International Journal of Applied Engineering Research, 10 (3), pp. 2035-2038.
41. Amudhavel, J., Kathavate, P., Reddy, L.S.S., Bhuvaneshwari Aadharshini, A., "Assessment on authentication mechanisms in distributed system: A case study", (2017) Journal of Advanced Research in Dynamical and Control Systems, 9 (Special Issue 12), pp. 1437-1448.
42. Amudhavel, J., Kodeeshwari, C., Premkumar, K., Jaiganesh, S., Rajaguru, D., Vengattaraman, T., Haripriya, R., "Comprehensive analysis on information dissemination protocols in vehicular ad hoc networks", (2015) International Journal of Applied Engineering Research, 10 (3), pp. 2058-2061.
43. Santhi, P., Mahalakshmi, G., Classification of magnetic resonance images using eight directions gray level co-occurrence matrix (8dglcm) based feature extraction, International Journal of Engineering and Advanced Technology, 2019, 8(4), pp. 839–846
44. Amudhavel, J., Ilamathi, R., Moganarangan, N., Ravishankar, V., Baskaran, R., Premkumar, K., "Performance analysis in cloud auditing: An analysis of the state-of-the-art", (2015) International Journal of Applied Engineering Research, 10 (3), pp. 2043-2046.
45. Krishnan J Santhana and S Geetha, "Prediction of Heart Disease using Machine Learning Algorithms", ICICT, 2019.
46. Aditi Gavhane, Gouthami Kokkula, Isha Panday and Kailash Devadkar, "Prediction of Heart Disease using Machine Learning", Proceedings of the 2<sup>nd</sup> International conference on Electronics Communication and Aerospace Technology (ICECA), 2018.
47. Senthil Kumar Mohan, Chandrasegar Thirumalai and Gautam Srivastva, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", IEEE Access, 2019.
48. Himanshu Sharma and M A Rizvi, "Prediction of Heart Disease using Machine Learning Algorithms: A Survey", International Journal on Recent and Innovation Trends in Computing and Communication, vol. 5, no. 8, August 2017.
49. M. Nikhil Kumar, K. V. S. Koushik and K. Deepak, "Prediction of Heart Diseases Using Data Mining and Machine Learning Algorithms and Tools", International Journal of Scientific Research in Computer Science Engineering and Information Technology IJSRCSEIT, 2019.
50. Amandeep Kaur and Jyoti Arora, "Heart Diseases Prediction using Data Mining Techniques: A survey", International Journal of Advanced Research in Computer Science IJARCS, 2015–2019.