

International Journal of Advanced Research in Computer and Communication Engineering

Heart Disease Prediction with Machine Learning Classifiers

P.V.R.D. Prasada Rao¹, Pabbathi Sai Vaishnavi², Vakkalagadda Sai Mani Deep³,

Harshavardhan Samineni⁴, Madasu Paul Revanth⁵

Professor, CSE, Koneru Lakshmaiah Education Foundation, Guntur, AP, India¹
Student, CSE, Koneru Lakshmaiah Education Foundation, Guntur, AP, India²
Student, CSE, Koneru Lakshmaiah Education Foundation, Guntur, AP, India³
Student, CSE, Koneru Lakshmaiah Education Foundation, Guntur, AP, India⁴
Student, CSE, Koneru Lakshmaiah Education Foundation, Guntur, AP, India⁵

Abstract: This research papers gives an analysation over the features of heart disease analysis module that deals with a wide range of complications. Heart can be called as the most crucial and useful organ in the body of any organism especially in human body. It plays a very important and purpose full role. Diagnosing the heart is very important and regular check-ups regarding heart and it's related diseases can be helpful in living a healthy life. To tackle such heart complications and to help people be cautious regarding their heart conditions there exists an essential urge of heart disease prediction system. Machine learning provides the noticeable support in finding any type of the event which demands a structural training from the naturally occurring events. For the purpose of disease analysis, machine learning is thought to be the most well-known and notorious platform, while the convenience of utilising it along with computational domination in the array of modules that grabs the attention millions of patients all over globe, which also demands the several technical concerns in different platforms. This research article aims to analyze the predictive power of several machine learning algorithms for cardiac disorders.

Keywords: Heart Disease.prediction, machine learning, analysis

I. INTRODUCTION

The utilization of the disease analysis systems has data analysis increased Over the past few years, so far data analysis is the most famous and the popular analysis platform and it provides the convenience of using the analysis systems along with the computational convenience to handle the wide array of modules. The heart disease analysis platform provides us with the easy structure and demands the mark up languages and the technologies like data science and data analysis to function as desired. The movement that is discreetly designed functions to deliver a solitary information for the benefit of the interface of the patient and a service which on a whole completes the assigned contextual work. The use of the smart devices for the health monitoring day by day has increased rapidly and the majority of the global population already possess these facilities.

The usage of heart diseases analysis has elevated exponentially in the very recent years. Prediction of heart disease is regarded as one of the greatest significant functions in health care and pharma industry. A tiny mistake or the human error can cause sever complication while dealing with heart and heart diseases. Because of inefficient treatment and lazy diagnosis are causing a rapid growth in the heart complications and treatment failures day by day. Heart is the most essential organ so the care for heart is the top priority in the health industry. Comparative study is highly needed in the stream of predicting the heart diseases.

The majority of the heart patients are worldwide as a result of the late recognition of the diseases or due to lack of accuracy of severity of the disease so There's a critical need of more efficient methods and algorithms for predicting the heart diseases. Heart can be called as the most crucial and useful organ in the body of any organism especially in the human body. Diagnosis of heart is very important and regular check-ups regarding heart and heart related diseases can be helpful in living a healthy life. To tackle such heart complications and to help people be cautious regarding their heart conditions there exists an essential urge of heart disease prediction system.

International Journal of Advanced Research in Computer and Communication Engineering

M

Impact Factor 8.102 😤 Peer-reviewed & Refereed journal 😤 Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.13432

Machine learning can be termed as among the best efficient and useful technology in present days for the sake of testing and training. ML is the branch of the AI which can be taken into the consideration as a broad field of learning machines will be emulating the abilities of man. Machine learning modules are trained in a manner where they can pick up skills to utilise and process the data, so the combination of both the artificial intelligence and machine learning is called as the machine intelligence. Taking this feature machine learning into the consideration, we have taken the biological parameters as the testing data such as blood pressure, cholesterol, and body fat, gender, age and so on. Based up on these biological parameters a comparison was performed for the sake of accuracy of different machine learning algorithms. We utilised four algorithms such as Knn, decision tree, SVM and linear regression. Based up on the calculations regarding the accuracy of these algorithms we have concluded the best algorithm heart disease prediction.

II. LITERATURE REVIEW

Heart disease analysis are expensive and tough to maintain. Many reports are segregated because of which causes in the lack of resources from the doctor side. data analysis modules structure gives capability and the framework needed, data analysis and libraries. The usage of disease analysis systems has elevated exponentially in the very recent years. data analysis is the most infamous smartphone platform, while the convenience of utilise with the computational domination in the array of modules that grabs the attention millions of patients all over globe, which also demands the security concerns on these platforms. This research analysis provides us the understanding over the heart analysis method based on data analysis platform on a variety of devices. As the technology of data analysis evolving on this planet as we possess a very data analysis connection and network to instantly interconnect to one other.

The intermediate interface which comes under the category of programming interface that acts as an interface amongst the modules. This gives us the immediacy, automation, automation, and adaption. The disease analysis system gives us the sources of different disease gathered from various individuals in various locations on this planet and it tries to update the module with the latest information that is available. a new and updated design language was generated within the domain of data analysis in the name of material design which is completely dependent on the system of cards which further uses the grind-based layouts, depth effects, padding and the animation data. With the system the patients can acquire the better patient experience and can combine the resources at a single place all along with maintaining the patient's integrity.

In this particular module the researchers have used the data science methods for the benefit of the better experience for the patient. Even after using this data analysis It's possible that they can't reach out to the maximum point of the resources and to rectify that issue we can utilise the doctor's panel where the doctor can give their own health reports. Not only to write but to manage and to delete it. This contain will have a very separate database and the access bar which will be labelled as patient's health condition. A module will be implemented where the utiliser will certainly have various health reports based on the patient.

The chronic disease prediction platform provides an easy structure and demands the mark up languages and the technologies like data science and data analysis to function as desired. The movement that is discreetly designed functions to deliver a solitary information for the benefit of the interface of the individual and a service which on a whole completes the assigned contextual work. The module that we worked for the benefit of this research paper can be considered and compared Upon completion of the day, we can really add the modules that will be built Eventually.

The use of disease analysis systems has data analysis increased in the recent years, so far data analysis is the most famousy and the popular analysis platform and it provides the convenience of using the analysis systems along with the computational convenience to handle the wide array of modules .Heart disease analysis platform provides us with the easy structure and demands the mark up languages and the technologies like data science and data analysis to function as desired. The movement that is discreetly designed functions to deliver a solitary information for the benefit of the interface of the patient and a service which on a whole completes the assigned contextual work.

III. IMPLEMENTING ANY STRATEGY

Regarding your neutral working model to be successfully deployed, you must take several important actions while implementing a machine learning strategy. The fundamental actions you should think about are as follows:

1)Select the right algorithm: Pick the machine learning algorithm or algorithms that will be most beneficial to you solve your challenge. Consider factors including the task's kind (classification, regression, or clustering), the volume of data, and the available computer resources.

2)Data collection and preparation: Assemble pertinent information from numerous sources. Make sure the information is accurate, organized, and pertinent to the issue at hand. To develop useful features for modeling, perform data preprocessing, which may include Insufficient value handling, outlier detection, and feature engineering are all possible. For developing and assessing the model, divide the data into training, validation, and test sets.



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 $\,\,st\,$ Peer-reviewed & Refereed journal $\,\,st\,$ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.13432

3)Model Development: Utilize the training data to create, train, and improve your machine learning model(s). Use methods like grid search or randomized search to fine-tune hyperparameters. Utilize the right measurements and validation methods to assess the model's performance.

4)Model Evaluation: Utilize metrics pertinent to your issue to evaluate the model's performance on the validation dataset (examples: accuracy, TP,TN). Consider potential problems like overfitting and underfitting and make the required model revisions.

5)Monitoring and Maintenance: Keep an eye on the model's functionality in the real-world setting. Create notifications for abnormalities or performance degradation. To keep the model current, retrain it frequently with fresh data

IV. METHODOLOGY

One of the most effective technologies is machine learning, which depends entirely on two concepts: testing and training. This implies that the module receives direct training of the desired occurrences from the data and experience. The test will be applied to various demands based on this training in accordance with the algorithm's requirements. Three categories of machine learning algorithms are known to exist:

- 1. Supervised learning
- 2. Unsupervised learning
- 3. Reframed learning



Fig 1: Basic Machine Learning Classifiers

A) SUPERVISED MACHINE LEARNING:

Supervised machine learning refers to the learning process with a qualified tutor, or you can also term it as the learning about an event under the supervision of a teacher. We posses the training dataset that can be considered as the teacher for the sake of prediction of the given dataset that acts as the data set for testing. Supervised learning is completely dependent on the concept of "train me".

Supervised learning consists of the below mentioned methods.

- Regression
- Decision tree
- Random forest
- Classification

In relation to recognising the patterns and when it calculates the probability of the outcomes that are uninterruptable, can be called as the phenomenon of regression. The system posses the ability to recognise the values, the numbers and the grouping sense of numbers such as height and the width and so on.

The following are the supervised ML algorithms.

- Logistic regression
- Linear regression
- Neural networks
- SVM
- Gradient boosted tree

215



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 😤 Peer-reviewed & Refereed journal 😤 Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.13432

- Random forest
- Naïve bayes
- Decision tree

B) UNSUPERVISED MACHINE LEARNING:

With no sort of supervised training or with no guidance or in very simple words with no help of teacher. In the case of the unsupervised learning the data set will be provided that works automatically on the dataset to find the relationship and pattern and within the relationships that were created, whenever a newer information is provided it classifies and store then in created relations. Unsupervised learning is usually based on the idea of "self sufficient"

Below algorithms are few unsupervised learning algorithms

- K-means clustering
- PCA
- t-SNE

C) DATA COLLECTION

The first and the foremost step in the method by which prediction can be described as collection of data and categorising regarding the testing and the training dataset. In regards of this research we have considered 73 percent of training dataset and 37 percent of testing dataset of the whole system. Categorizing the testing and training datasets in the whole project.

SELECTION OF ATTRIBUTES

The dataset's attribute is property of the dataset which can be used In consideration of the system and for the benifit of heart, several attributes such as heart rate of the patient, patient's gender, patient's age and so on.

V. RESULTS

LINEAR REGRESSION

One type of supervised machine learning algorithm is linear regression. This method depends entirely on the relationship that exists between the independent and dependent variables.



Fig 2: Points the X and Y axis

Because the relationship between the variables X and Y, which are independent and dependent variables was established by a linear equation found in nature, the approach used to analyze the data is known as linear regression.

DECISION TREE

Decision trees are a type of supervised learning algorithm and can be thought of as the graphical representation of the dataset. In order to facilitate the construction of the tree, we employed the entropy of the data's attributes. The various nodes and attribute root were then determined based on this information.

Entropy= - Σ Pij log Pij (1)

© <u>IJARCCE</u> This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 😤 Peer-reviewed & Refereed journal 😤 Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.13432

The node that possesses the calculation of highest entropy can be considered as the root node and the processes repeats itself till all the tree nodes will be calculated or till it generates the trees.



Fig 3: Identifies the tree structure of dataset

One possible explanation for the decision tree's lower accuracy rate as compared to linear regression is that it generates an overfitting problem that is not appropriate for the computation due to an inappropriate node balance.

SUPPORT VECTIOR MACHINE (SVM)

It may be regarded as one of the unique areas of machine learning techniques that works with the concept of a hyperplane, which

divides the entire set of data by creating a hyperplane between them.



Fig 4: Identify the hyperplane

Sample of the training dataset Yi, Xi where i=1,2,3,...n and Xi regarded as the ith vector or the target vector. The number of hyper plans will be used to determine the support vector. If a line will be utilised as the hyper plan then the method referred to as the linear support vector

KNN (K NEAREST NEIGHBOUR)

This algorithm functions based up on the distinct data and distance between the place where the information is segregated with one other. Every group of the data can be termed as the neighbour of each other, and it can be called as the neighbour it can be decided by the user who plays the vital role in the process of analysation of the data



International Journal of Advanced Research in Computer and Communication Engineering



Fig 5: k=3 represents that there are 3 neighbours.

CALCULATION OF ACCURACY



Fig 6: Accuracy in graph

Algorithm's accuracy is completely dependent on four of the values which are namely called as

- 1. Amount of people that possess heart disease: True positive (TP)
- 2. Amount of people that possess no heart disease: False positive (FP)
- 3. The number of individuals with cardiac disease and those without it: True negative (TN)
- 4. The number of individuals without heart disease and those who do: False Negative (FN)

Accuracy = (TP+FN)/(FP+TP+FN+TN) (2)

VI. CONCLUSION

After executing all necessary machine learning techniques for training and testing, we discovered that the Knn algorithm's accuracy might be regarded as the quickest of all. As we mentioned the accuracy should be calculated according to the confusion matrices. The value of TN, TP, FN, FP are mentioned using the equation of accuracy. The accuracy value was obtained and through the final values we obtained we can conclude that Knn can be the most effective method for machine learning with 87% accuracy.

218



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 😤 Peer-reviewed & Refereed journal 😤 Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.13432

ALGORITHM	ACCURACY
SVM	83 %
Decision tree	78%
Logistic regression	79%
Knn	87%

fig : The accuracy of various machine learning algorithms are compared and displayed.

REFERENCES

- [1]. Santhana Krishnan J and Geetha S, "Prediction of Heart Disease using Machine Learning Algorithms" ICIICT, 2019.
- [2]. Aditi Gavhane, Gouthami Kokkula, Isha Panday, Prof. Kailash Devadkar, "Prediction of Heart Disease using Machine Learning", Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology(ICECA), 2018.
- Senthil kumar mohan, chandrasegar thirumalai and aDisease Prediction Using Hybrid Machine Learning Techniques" IEEE Access 2019.
- [3]. 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 Volume: 5 Issue: 8, IJRITCC August 2017.
- [4]. M. Nikhil Kumar, K. V. S. Koushik, 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, JJSRCSEIT 2019.
- [5]. 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.
- [6]. Pahulpreet Singh Kohli and Shriya Arora, "Application of Machine Learning in Diseases Prediction", 4th International Conference on Computing Communication And Automation(ICCCA), 2018.
- [7]. M. Akhil, B. L. Deekshatulu, and P. Chandra, "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm," Procedia Technol., vol. 10, pp. 85–94, 2013.
- [8]. S. Kumra, R. Saxena, and S. Mehta, "An Extensive Review on Swarm Robotics," pp. 140-145, 2009.
- [9]. Hazra, A., Mandal, S., Gupta, A. and Mukherjee, "A Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review" Advances in Computational Sciences and Technology , 2017.
- [10]. Patel, J., Upadhyay, P. and Patel, "Heart Disease Prediction Using Machine learning and Data Mining Technique" Journals of Computer Science & Electronics, 2016.
- [11]. Chavan Patil, A.B. and Sonawane, P."To Predict Heart Disease Risk and Medications Using Data Mining Techniques with an IoT Based Monitoring System for Post-Operative Heart Disease Patients" International Journal on Emerging Trends in Technology, 2017.
- [12]. V. Kirubha and S. M. Priya, "Survey on Data Mining Algorithms in Disease Prediction," vol. 38, no. 3, pp. 124–128, 2016.
- [13]. M. A. Jabbar, P. Chandra, and B. L. Deekshatulu, "Prediction of risk score for heart disease using associative classification and hybrid feature subset selection," Int. Conf. Intell. Syst. Des. Appl. ISDA, pp. 628–634, 2012.