



CARDIAC INFARCTION CATEGORIZATION AND DIAGNOSIS USING EXPERT SYSTEM MODELS

Aswini P¹, Akshaya R², Mareedu Kundanika³, Vengateshwaran M⁴

^{1,2,3} UG student, ⁴ Assistant Professor in CSE

Department of Computer science & Engineering

Agni College of Technology, Chennai, India

Abstract

In today's world, heart disease is a leading cause of death to human. This affects both female and male category. People of age group 25-69 may have a heart disease. Now a day's people are not taking care of their health because of their busy life. So they are not aware of their health condition and ended up getting various diseases. Those disease can be get cured if they were aware of it earlier. Heart disease is one of them. It can be cured if the patient is get diagnosed correctly. The main challenge is to look for a system which can predict the heart disease and give correct diagnose for it. In this paper, we have come up with system which can predict the disease accurately using machine learning algorithm. This can help doctors to diagnose a patient.

Keywords:- Heart disease, machine learning.

I. INTRODUCTION

1.1 MACHINE LEARNING:

Machine learning is a technology that builds intelligence system. These systems have ability to learn from experience and provide results according to it. Machine learning came up to reduce human work. A task can be done without human involvement.

A large area of artificial intelligence is machine learning. This enables system to induce new knowledge from experience. The system can learn things without being explicitly programmed.

Applications of AI:

Though artificial intelligence arouse thoughts of science fiction, artificial intelligence already has many uses in our day to day life, for example:

- **Email filtering:** Email services use artificial intelligence for filtering incoming emails. Users can able to train their spam filters by marking emails as "spam".
- **Personalization:** Online services use artificial intelligence to personalize your experiences. Services, like Amazon or Netflix, "predict" from your previous purchases and purchases of other users in order to recommend relevant content for you.
- **Fraud detection:** Banks use artificial intelligence to determine whether there is strange activity on your account. Some activities, such as foreign transactions, might be flagged by the algorithm.
- **Speech recognition:** Applications use artificial intelligence for speech recognition to optimize functions. Examples include intelligent personal assistants, such as Amazon's "Alexa" or Apple's "Siri".

1.2 MACHINE LEARNING ALGORITHM:

Algorithms are sequence of instructions used to solve the problems. Algorithms, developed by programmers are used to instruct computers in new tasks, these are the building blocks of the advanced digital world we see today. Computer algorithms organize large amounts of data into services and information, based on certain instructions and rules.

Instead of programming the computer every step, this approach gives the instructions to computer that allow computer to learn from data without new step-by-step instructions. This means computers can be used for new complicated tasks that cannot be manually programmed. The basic process of machine learning is to train data from learning algorithm. Learning algorithm later generates new set of protocols and rules, based on inferences from the data. By using different training data, the same learning algorithm might be used to generate different models. For example, the same learning algorithm may be used to teach the computer how to translate languages or predict the stock market. Inferring new instructions from data is one of the core strength of machine learning. It also showcases the critical role of data: the more data available to train the algorithm, the more it learns. In fact, many recent advances in AI have not been due to radical innovations in learning algorithms, but instead the enormous amount of data enabled by the Internet.

How machines understand things and start to learn:

Although a machine learning model may apply a mixture of different techniques, the methods for learning can typically be categorized as three general types they are:

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning

Supervised learning:

In Supervised learning, we train the machine using data which is well “labelled”. It means some data is already tagged with the correct answer. It is compared to learning which takes place in the presence of a supervisor or a teacher or on a monitoring environment. A supervised learning algorithm learns from labelled training data, that helps you to predict outcomes for unforeseen data. Successfully scaling, building and deploying accurate supervised machine learning data science model takes time and technical expertise from a team of highly skilled data Scientists. Moreover, Data scientist must rebuild models to make sure the insights given remains true until its data changes.

- ☐ Two Types of Supervised Machine learning techniques
- ☐ Regression
- ☐ Classification

Regression:

Regression technique predicts single output value by using the training data.

Example: You can use regression technique to predict the house price from training data. The input variables will be size of a house, locality etc.

Classification:

Classification means to group the output inside a class. If the algorithm tries to label input into two distinct classes, it is called as binary classification. Selecting between two or more classes is referred as multiclass classification.

Example: Determining whether someone will be a defaulter of the loan.

Strengths: Outputs always have a probabilistic interpretation, and the algorithm can be regularized to avoid over fitting.

Weaknesses: Logistic regression may underperform when there are multiple or non-linear decision boundaries. This method is not flexible, so it does not capture more complex relationships.

Unsupervised machine Learning:

Unsupervised learning is also one of the machines learning technique, in which you do not need to supervise the model. Instead, you need to allow the model to work on its own to discover information. It mainly deals with the unlabelled data set. Unsupervised learning algorithms allow you to perform more complex and deeper processing tasks compared to supervised learning. Even though, unsupervised learning can be more unpredictable compared with other natural learning deep learning and reinforcement learning methods.

Two Types of Unsupervised Machine Learning Techniques

- ☐ Clustering
- ☐ Association.

Clustering:

Clustering is an important concept when it comes to unsupervised learning. It mainly deals with finding a structure or pattern in the collection of uncategorized data. Clustering algorithms will process your data and find natural clusters(groups) if they exist. You can also modify how many clusters your algorithm can identify.

Association:

Association rules allow you to establish associations among the data objects inside large set of databases. Unsupervised learning technique is about discovering exciting relationships between variables in large set of databases. For example, people who buy a new home most likely to buy new furniture.

□ **Reinforcement learning:** This algorithm interacts with a dynamic environment that provides feedback in terms of rewards and punishments. Consider an example, self-driving cars being rewarded to stay on the road.

1.3 HEALTH CARE IN ML:

Heart disease is increasing drastically in our modern world. According to WHO (world health organization), an estimated 17.9 million people died from heart diseases. It is necessary to know the cause of disease such as changing in lifestyle, smoking, food habits, physical activity, obesity, diabetics, and blood pressure and so on. For this reason, it is important to find system which helps to make correct and efficient decision for diagnosing a patient.

Tradition way of predicting the heart disease are using ECG, occultation, blood pressure, blood sugar level and cholesterol. But these techniques are expensive and time consuming task. Machine learning algorithms are used for making diagnosis in a less time and it also increases the accuracy.

The classification algorithms are SVM(support vector machine), KNN (K nearest neighbour), NB(Naive Beys), and RF(Random forest). This is used to classify the datasets.

Data preprocessing technique is included in this process to improve the data quality and extract useful information. This identify the missing values in the dataset and replace them by it's mean value or default value instead of ignoring them. We have used WEKA(waikato environment for knowledge analysis) tool for preprocessing .

The main object of this study is to predict the heart disease and comparison of different machine learning algorithms. This tells us which gives better result.

II. SYSTEM STUDY

Several studies and resources are conducted in heart disease data sets using various approaches.

[1] Nabaouia Lourdi and Meryem Amar proposed to predict CVD(Cardio vascular disease) using machine learning. They had predicted the disease using various approaches such as support vector machine (SVM), naive bays (NB) and k nearest neighbor (KNN). Data preprocessing technique used in this experiment is to remove the noises in the dataset. They have used normalization technique for it. The result proved that the accuracy of SVM gives 86.8% which higher than all the other approaches.

[2] M.Ganesan proposed a IOT(internet of things) based heart disease prediction system using machine learning models. The iot gadget is used to monitor the health condition of a patient and it record and examine reports. The monitored data can be transmitted to information platform to diagnose a patient in real time or stored in datasets. This information is stored in a cloud platform for safety and can be accessed by health applications. The classifiers used are J48 classifier, logistic regression and multi layer perception. He calculated F score, precision, recall and kappa values for each classifier and compared them. In this study, j48 classifier gives 91.4% accuracy.

[3] Reddy prasad, pidapathi, Anjali and Deepa were proposed a system which is used to predict heart disease using logistic regression algorithm. The system have a data which is classified if the patient have heart disease or not according to features in it. They have predicted the disease using logistic algorithm gives 86.89% accuracy and it is compared with the other classifiers such as naive bays(86%), decision tree(78.69%). This says that logistic regression have high accuracy.

[4] Youness khourdifi and Mohamed bahaj are presented a study that says a heart disease prediction system can give more accuracy if the dataset is optimized. They have optimized the dataset using PSO/ACO(Particle swarm optimization / Ant colony optimization). The features selection process was carried out by using FCBF(Fast

correlation- based features selection) to eliminate redundant features. After all these process , classification technique was used to classify the patient records as presence and absence of heart disease. The proposed optimized model by FCBF , PSO and ACO gives 99.65% accuracy which is used KNN classification algorithm.

[5] Avinash Golande and Pavan kumar had researched an effective machine learning algorithm that is used to predict the heart disease. They have discussed about various algorithms and tools. The Machine learning algorithms used are decision tree, classification and clustering . They concluded that the accuracy of the system depend on the tools diagnosed for execution. They have used data mining technique also to extract the data from the real world.

III. SYSTEM ANALYSIS

3.1 Problem statement:

Among various life threatening diseases, heart disease has a great attention in medical research. Diagnosis of heart disease is a challenging task in today's world. The disease need to be predicted for a future treatment. So that a patient can be cured.

3.2 Existing system:

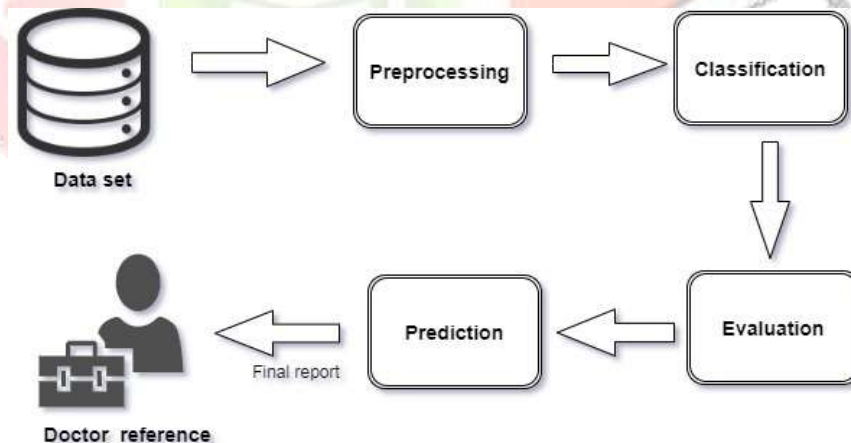
The before all existing system works on various machine learning algorithms such as support vector machine , k nearest neighbour , naive bays , random forest and so on. But the accuracy of all the algorithms is not sufficient to predict the disease. Data preprocessing, data optimization techniques and feature selection processes are not taken into consideration. So the system is considered as weak.

3.3 Proposed system

Our goal is to identify the presence of absence of disease using an effective machine learning algorithm which can give more accuracy. This is also used to compare various algorithms to find a better one.

IV. SYSTEM ARCHITECTURE

Architecture diagram:



UCI dataset:

This study uses UCI heart disease dataset to predict the patient who is having heart disease. Heart disease data set was given by Robert detrano. MD, PhD., This dataset contains 303 instances and 13 attributes. The UCI data set provides various amount of datasets from four institutions.

1. Cleveland clinic foundation
2. Hungarian institute of cardiolog , Budapet
3. V.A medical center, Long Beach
4. University hospital , Switzerland

Data pre-processing:

This is an important step in data classification. Pre-processing is used to find out the missing data in patient records. The missing values can be replaced by it's mean values or default values. It allows the system to give more accuracy.

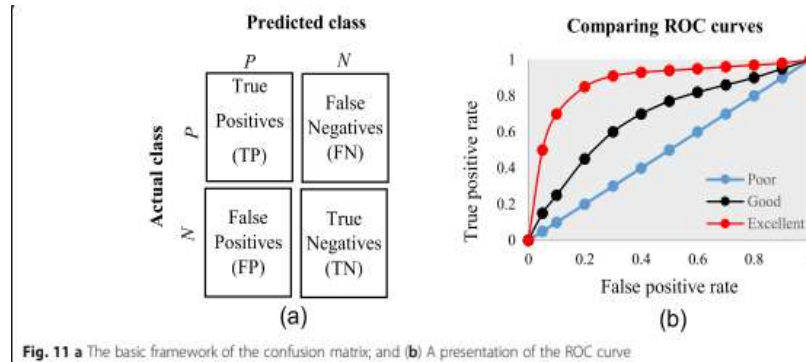


Fig. 11 a The basic framework of the confusion matrix; and (b) A presentation of the ROC curve

Classification:

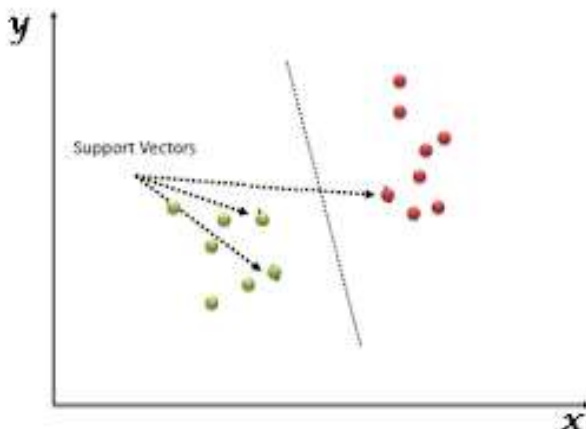
Classification techniques used in this paper are followings.

1. Support vector machine:

Support vector machine is a supervised learning model which helps to analysis and classification of data. It separates the data distinctly by hyperplane. In our system this separate the patient with disease and without disease.

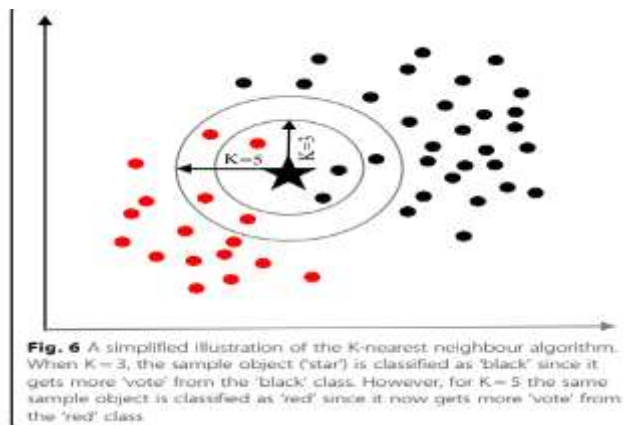
SVM are classified into two catageries .

- Linear SVM: In linear SVM, the training data separate classifier by hyperplane
- Non linear SVM: In non linear SVM, the training data cannot be separate classifiers by hyperplane .



2. K nearest neighbour:

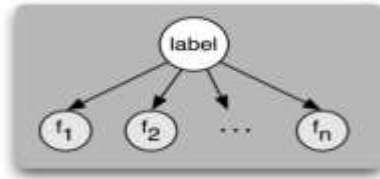
K nearest neighbour algorithm is widely used for classification and regression task. It finds similarities in the dataset. K factor is used to find the similarities. The more no of k factor increases the accuracy and make the system more stable.



3. Naive Bayes:

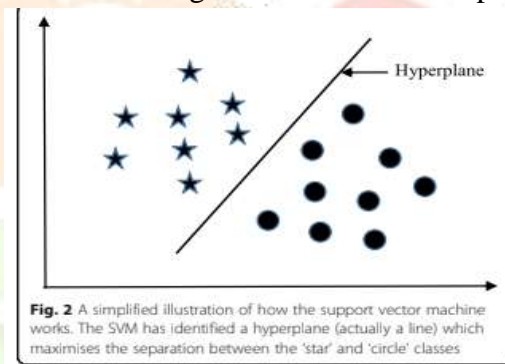
Naive Bayes is a classifier which uses Bayes theorem. It acts independently among predictors. It is a probabilistic graphical model to represent random variables with conditional independence. Naive Bayes is used to represent the probabilistic relationship between the diseases and symptoms.

Naïve Bayes Classifier



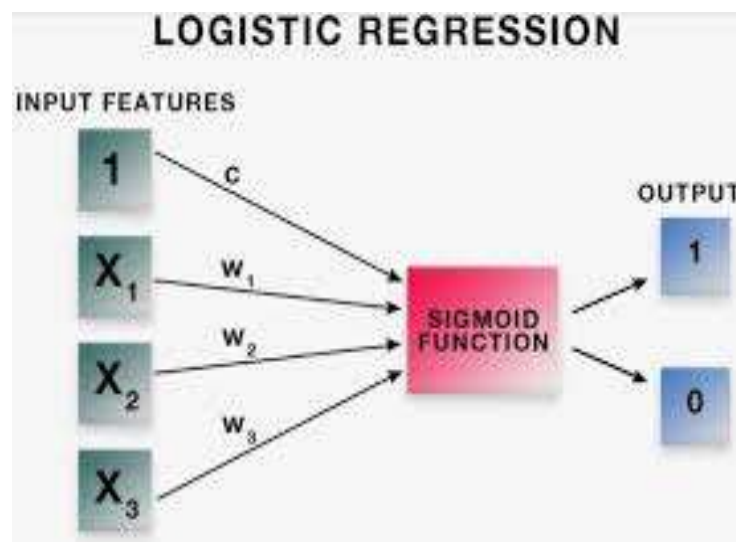
4. Random forest:

Random forest uses a bagging approach. It creates a bunch of decision trees by using a random subset of data. These data sets are needed to be trained several times in order to achieve good prediction performance. In this ensemble learning method, the output of all decision trees are combined together to make a final prediction.



5. Logistic regression:

Logistic regression is used for classification task and not for regression task. Regression means the linear model fit into the feature space. It uses a logical function to a linear combination of features. This is needed to predict the outcome of a dependent variable.



Prediction and evaluation:

This step is used to identify the presence or absence of heart disease in a patient based on the medical data. Prediction is done based on the classification task. Accuracy helps to check the efficiency of a build model. We have compared the accuracy given by all the classifiers. The final result tells us whether the patient is normal or abnormal.

V.CONCLUSION

Heart Disease is one of the major threats now days in our society. It is difficult to find the odds of getting heart disease based on risk factors. However, machine learning algorithms are useful to predict the output from existing data. This work is the first step to predict the odds of getting heart disease using machine learning algorithm. Our future work is to find new approaches in this domain.

Reference:

- [1] Priyan Malarvizhi Kumar, Usha Devi Gandhi, "A novel threetier Internet of Things and architecture with machine learning algorithm for early detection of heart diseases", Computers and Electrical Engineering, Vol.65, pp. 222–235, 2018.
- [2] Prabal Verma, Sandeep K. Sood, "Cloud-centric IoT based disease diagnosis healthcare framework", J, Parrallel Distrib. Comput., 2018.
- [3] Kakria P, Tripathi N.K, Kitipawang P, "A real-time health monitoring system for remote cardiac patients using smartphone and wearable sensors", Int. J. Telemed. Appl. (2015).
- [4] G.Parthiban,S.K.Srivatsa.(2012) Applying Machine Learning Methods in Diagnosing Heart Disease for Diabetic Patients', International Journal of Applied Information Systems (IJAIS), ISSN : 2249-0868 , Volume 3– No.7.
- [5] Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, Juan Gutierrez. (2017) A Comprehensive Investigation and Comparison of Machine Learning Techniques in The Domain of Heart Disease', Published in: Computers and Communications (ISCC), 2017 IEEE Symposium on 3-6 July 2017, DOI: 10.1109/ISCC.2017.8024530.
- [6] Sanjay Kumar Sen. (2017) Predicting and Diagnosing of Heart Disease Using Machine Learning', International Journal of Engineering and Computer Science', ISSN:2319-7242 Volume 6 Issue 6, Page No. 21623-21631.
- [7] Alice M. Arnold and Richard A. Kronmal,"Multipleimputation of baseline data in the cardiovascular helath study," American Journal of Epidemiology, Volume 157, Issue 1, 1 January 2003, Pages 74– 84, <https://doi.org/10.1093/aje/kwf156>

Author's Profile:

P. Aswini B.E.,
Final year CSE
Agni College of Technology, Chennai



R. Akshaya B.E.,
Final year CSE
Agni College of Technology, Chennai



Mareedu Kundanika B.E.,
Final year CSE
Agni College of Technology, Chennai



Mr. M. Vengateshwaran M.E.,
Assistant Professor in CSE
Agni College of Technology, Chennai
Area: Machine Learning, Big Data, Data mining, IR