Comparative Analysis of Machine Learning Models for Heart Disease Prediction

*K.Ishwarya[1], Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Semmancheri, Chennai 600119, TamilNadu, India, ishwarya.k.cse@sathyabama.ac.in

C.Sai Varun[2], Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Semmancheri, Chennai 600119, TamilNadu, India, saivarunchandrashekar@gmail.com

Abstract-Heart disease is an appellation that offer a large number of heart related medical actions. The heart and its parts are directly swayed by quirky heart actions which are portrayed by various medical actions. Major health stumbling block in today's time is health disease. Various machine learning algorithms are set in motion for heart disease prediction during recent years. The aim of this paper is scrutinizing various machine learning algorithms. Various investigations reveal that Neural networks with 15 attributes has surpassed over all other machine learning algorithms. The peroration obtained from the investigation states that decision tree is considered when there is a complex dataset and has shown good accuracy with the aid of genetic algorithm.

Keywords: Heart Disease; Genetic Algorithm; Decision Tree; Neural Network; Naive bayes;

1. INTRODUCTION

Machine Learning algorithms are the programs that can learn the hidden patterns from the data, predict the output, and improve the performance from experiences and on their own. Machine learning algorithms are increasingly getting admired in the field of healthcare. Large amount of complex data about patients, hospital resources, disease registries electronic patient records, medical equipment, etc. are generated by healthcare industry. The vast volume of data is a crucial resource that must be handled and analyzed for knowledge extraction that supports cost reduction and decision making. Medical experts can apply machine learning in healthcare to create better diagnostic tools for examining medical pictures. Figure 1 depicts the basic machine learning algorithm model.

One in three persons globally have high blood pressure, which accounts for around half of all stroke and heart disease fatalities, according to the World Health Statistics 2012 report. Cardiovascular disease (CVD), another name for heart disease, refers to a variety of illnesses that affect the heart, not just heart attacks. Heart dysfunction also refers to structural issues with the heart's valves or irregular heartbeats. Heart failure, arrhythmias, and a variety of other issues can result from these issues.

Heart disease prediction in the healthcare industry may benefit from effective and efficient automated solutions. Our work makes an effort to give a thorough analysis of the many machine learning algorithms that might be used in these automated systems. A patient will undergo fewer tests as a result of this automation. As a result, it will save both analysts' and patients' time in addition to money.

2. METHODOLOGY

The central aim of this paper is utilizing various machine learning techniques which can be obliging for medical analysis or for explicit heart disease diagnosis. The central methodology used for our work was by investigating the publications, journals and reviews in the field of computer science and engineering, machine learning and cardiovascular disease in neoteric times [5].

3. RESEARCH OBSERVATIONS

3.1 Machine Learning and Neural Networks

An artificial neural network (ANN) is a mathematical model or computational model based on a biological neural network, commonly known as a "neural network". It is based, in other words, on the imitation of the biological neurological system.

This study has established a strategy for predicting heart disease based on 15 features [4]. 13 factors were previously utilised for prediction, but this research included 2 more factors obesity and smoking for an accurate diagnosis of heart disease.

One of the most straightforward and efficient classification algorithms is the Naïve Bayes Classifier, which aids in the development of quick machine learning models capable of making accurate predictions. Additionally, a database of heart diseases has been used to analyse different machine learning techniques. For each classifier, a confusion matrix is found.

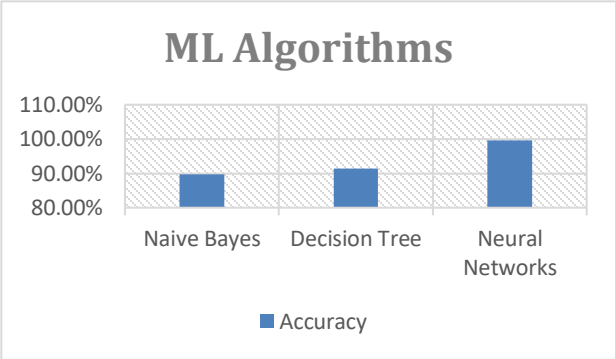
The results of this research are shown in Table 1, which demonstrates that neural networks have performed better than other machine learning techniques.

Classification Techniques	Accuracy
Naive Bayes	**89.75%**
Decision Trees	**91.43%**
Neural Networks	**99.58%**

Table 1: Comparison of various machine learning Algorithms

S.No.	Attributes	Support Set	
		Heart Patients	Non – Heart Patients
1.	chest pain	4	1,2,3
2.	Rbps	134-153	142-154
3.	Exang	Yes	No
4.	Oldpeak	2.06 -6.2	<2.06
5.	Thalach	71-136	136-168
6.	Ca	1,2,3	0

Table 2: Values of the features in the support set



Graph 1: Comparison of various machine learning Algorithms

3.3 Fuzzy Logic and Genetic Algorithm

The method that is being advocated in this study is an expanded version of the model that combines fuzzy expert systems and genetic algorithms for efficient feature selection and classification. For the development of knowledge-based systems in healthcare for disease diagnosis, fuzzy set theory and fuzzy logic are ideal [2].

In MATLAB, experiments are carried out with the fuzzy tool. The Mamdani model of fuzzy system is employed in this. Based on their knowledge of this field, the specialists who created the fuzzy rules. We use a dataset from the UCI machine learning repository, and we find that only 6 variables are both useful and essential for predicting heart disease. The set of all the chosen features serves as the input to the proposed system, and the system's output is a value of 0 or 1, which denotes whether a patient has heart disease or not.

As part of the fuzzy logic process, the input data is first fuzzified by being gathered as a crisp set and then transformed into a fuzzy set utilising fuzzy linguistic variables, fuzzy linguistic words, and membership functions. Following that, an inference is drawn in accordance with a set of rules, and finally, the defuzzification process is carried out [2]. Based on the acquired support sets, this system constructs the fuzzy rules. This support set is shown in Table 2.

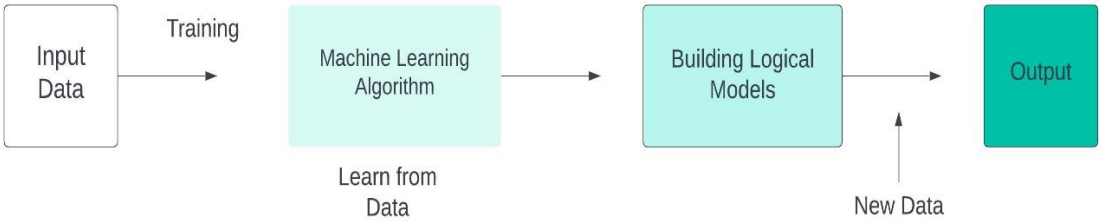


Figure 1: Machine Learning Model

3.4 Machine Learning and Supervised Machine Learning Algorithms

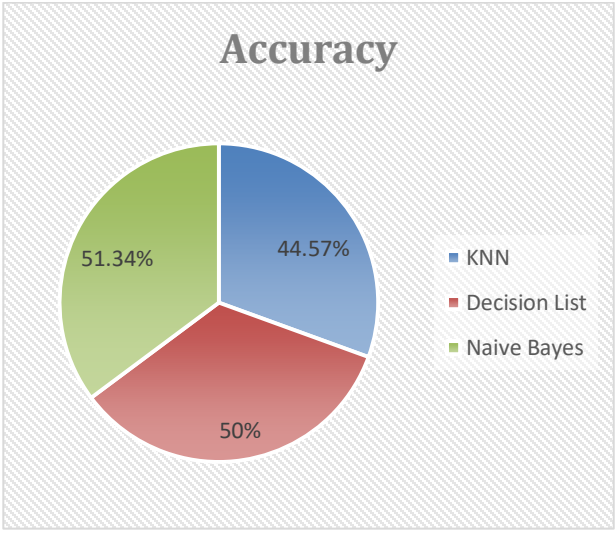
This research work has presented the data classification based on various supervised machine learning algorithms, namely, Naïve Bayes, Decision List and KNN. TANAGRA tool is used to classify the data and the data is evaluated using 10- fold cross validation.

TANAGRA [19] is a machine learning framework built for educational and research applications. TANAGRA is a machine learning tool intended for use in teaching and research.

It has an easy-to-use interface, direct access to data warehouses, and a wide range of machine learning algorithms for analysis and experimentation. TANAGRA makes it possible to evaluate and compare algorithm performance based on accuracy and calculation time. A training dataset of 3000 cases with 14 attributes is used to conduct experiments. Based on the attributes, this dataset is divided into two parts, with 70% of the data utilised for training and the remaining 30% for testing. The performance of each method is compared and assessed based on accuracy and calculation time [11]. Surprisingly, the Nave Bayes method outperformed the other two. Table 3 presents a detailed description of the algorithm performance study.

Algorithm Used	Accuracy	Time Taken
KNN	**44.57%**	**1000ms**
Decision List	**50%**	**710ms**
Naive Bayes	**51.34%**	**600ms**

Table 3: Performance analysis of various Algorithms



Graph 2: Performance analysis of various Algorithms

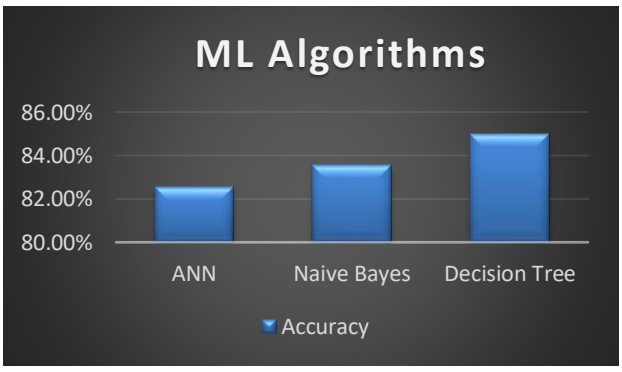
3.5 Machine Learning and Genetic Algorithm

The goal of this research was to decrease the number of qualities that were used to diagnose heart disease. In the past, 13 attributes were utilised to make this prediction, but this study work used a genetic algorithm and feature subset selection to minimise the number of attributes to just six. With no traits and a starting population of randomly generated rules, the genetic search began. The idea of "survival of the fittest" was used to produce new populations that matched the fittest rules in the existing populations and their progeny. Crossover and mutation, two genetic operators, were used to produce offspring. The generation process went on indefinitely till it evolved.

In addition to the genetic algorithm, the CFS Evaluator is also used. Weka 3.6.0 tool is used to conduct the observations. 909 records with 13 attributes made up the first data set. For simplicity, all qualities were made categorical, and contradictions were removed. Several classifiers are employed on the dataset corresponding to the 6 attributes after the 13 variables were reduced to 6, in order to predict heart disease. Table 4 displays a performance study of various classifiers. The table clearly shows that the Decision Tree outperformed the other models in terms of accuracy and mean absolute error.

ML Algorithms	Accuracy
ANN	**82.53%**
Naive Bayes	**83.53%**
Decision Tree	**85%**

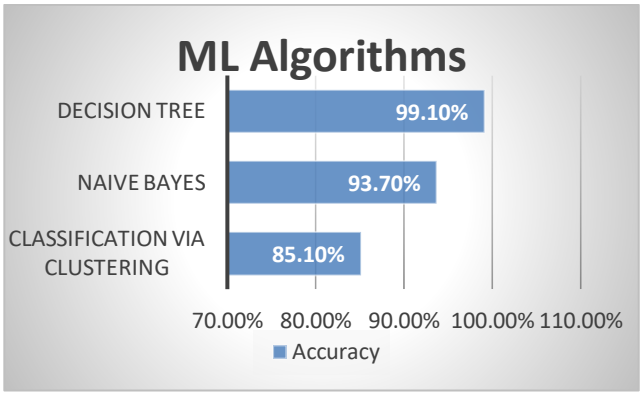
Table 4: Comparison Table for three Classifiers



Graph 3: Performance Analysis of ML Algorithms

ML Algorithms	Accuracy	Model Construction Time	Mean Absolute Error
Classification via Clustering	**85.1%**	**0.06s**	**0.117**
Naive Bayes	**93.7%**	**0.02s**	**0.044**
Decision Tree	**99.1%**	**0.09s**	**0.00016**

Table 5: Performance analysis of IHDP



Graph 4: Comparison Table for three Classifiers

3.6 IHDPS and Machine Learning Algorithms

In this study, a prototype Intelligent Heart Disease Prediction System (IHDPS) was created using Decision Trees, Naïve Bayes, and Neural Networks, three state of the art techniques. The .NET framework is used to create the web-based, user-friendly, scalable, dependasxble and expandable IHDPS system [14].

Hidden knowledge linked with heart disease can be enduced from a historical heart disease dataset using IHDPS. Baffling queries for diagnosing heart disease can be answered. It enables healthcare analysts and practitioners to make deft clinical judgements that are not possible with conventional decision support tools. By offering efficient therapies, it also helps to save healthcare expenses. Moreover, Both tabular and graphical forms of the results are manifested. This IHDPS is based on 15 attributes.

The Cleveland Heart Disease database yielded a total of 909 records. Two datasets, a training dataset (455 records) and a testing dataset (454 records), were created from the records. The investigation revealed that Naive Bayes had the highest percentage of accurate predictions for patients with heart disease (86.53%), followed by Neural Networks

(85.53%), and Decision Trees (85.53%). However, when predicting people without heart disease, decision trees proved to be the most accurate, scoring (89%) compared to the other two models.

4 RESULTS

Result of each machine learning algorithm has been shown clearly in different tables for better apprehension. Various classifiers are used in combination with many machine learning algorithms for heart disease prediction. As per observation, same classifiers have shown different accuracy in different machine learning algorithms in few cases.

5 CONCLUSIONS

Our work's primary goal is to give a study of several machine learning algorithms that can be used in automated heart disease prediction systems. In this paper, various methods and machine learning classifiers that have become available for the quick and accurate diagnosis of heart disease are defined. The investigation reveals that a neural network with 15 attributes has so far demonstrated the maximum accuracy, or 100%. Using 15 attributes, Decision Tree has likewise done well, achieving an accuracy rate of 99.62%. Additionally, Decision Tree has demonstrated 99.2% efficiency when combined with Genetic Algorithm and 6 characteristics.

6 REFERENCES

[1] Anooj PK. Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules. Journal of King Saud University-Computer and Information Sciences. 2012 Jan 1;24(1):27-40.

[2] Ephzibah EP, Sundarapandian DV. Framing fuzzy rules using support sets for effective heart disease diagnosis. International Journal of Fuzzy Logic Systems (IJFLS) Vol. 2012 Feb;2.

[3] Battineni G, Sagaro GG, Chinatalapudi N, Amenta F. Applications of machine learning predictive models in the chronic disease diagnosis. Journal of personalized medicine. 2020 Mar 31;10(2):21.

[4] Jindal H, Agrawal S, Khera R, Jain R, Nagrath P. Heart disease prediction using machine learning algorithms. InIOP conference series: materials science and engineering 2021 (Vol. 1022, No. 1, p. 012072). IOP Publishing.

[5] Godara S, Singh R. Evaluation of predictive machine learning techniques as expert systems in medical diagnosis. Indian Journal of Science and Technology. 2016 Mar;9(10):1-4.

[6] Anbarasi M, Anupriya E, Iyengar NC. Enhanced prediction of heart disease with feature subset selection using genetic algorithm. International Journal of Engineering Science and Technology. 2010 Oct 14;2(10):5370-6.

[7] Esfandiari, Babavalian MR, Moghadam AM, Tabar VK. Knowledge discovery in medicine: Current issue and future trend. Expert Systems with Applications. 2014 Jul 1;41(9):4434-63.

[8] Ahmed U, Rasool G, Zafar S, Maqbool HF. Fuzzy rule based diagnostic system to detect the lung cancer. In2018 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube) 2018 Nov 12 (pp. 1-6). IEEE.

[9] Khourdifi Y, Baha M. Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. International journal of Intelligent engineering & systems. 2019 Jan 1;12(1).

[10] Boadh R, Grover R, Dahiya M, Kumar A, Rathee R, Rajoria YK, Rawat M, Rani S. Study of fuzzy expert system for the diagnosis of various types of cancer. Materials Today: Proceedings. 2022 Jan 31.

[11] Kumar P, Maurya LS, Gupta HK. Diagnosis Of Heart Disease Using Machine Learning Algorithm. Webology (ISSN: 1735-188X). 2022;19(5).

[12] Haq AU, Li JP, Memon MH, Nazir S, Sun R. A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. Mobile information systems. 2018 Dec 2;2018:1-21.

[13] Q Yuan Z, Chen H, Xie P, Zhang P, Liu J, Li T. Attribute reduction methods in fuzzy rough set theory: An overview, comparative experiments, and new directions. Applied Soft Computing. 2021 Aug 1;107:107353.

[14] Shinde R, Arjun S, Patil P, Waghmare J. An intelligent heart disease prediction system using k-means clustering and Naïve Bayes algorithm. International Journal of Computer Science and Information Technologies. 2015 Feb;6(1):637-9.

[15] Cleveland database: http://archive.ics.uci.edu/ml/datasets/Heart+Disease

[16] Han, El Misilmani HM, Naous T. Machine learning in antenna design: An overview on machine learning concept and algorithms. In2019 International Conference on High Performance Computing & Simulation (HPCS) 2019 Jul 15 (pp. 600-607). IEEE.

[17] Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Chang AR, Cheng S, Delling FN, Djousse L. Heart disease and stroke statistics—2020 update: a report from the American Heart Association. Circulation. 2020 Mar 3;141(9):e139-596.

[18] Statlog database: http://archive.ics.uci.edu/ml/machine-learning-databases/statlog/heart

[19] http://eric.univ-lyon2.fr/~ricco/tanagra/