A Survey on Machine Learning Methods for Heart Disease Prediction

Amit Kumar Gupta, Prof. Bhavana Gupta
Computer Science and Engineering Department
SISTEC, RGPV, Bhopal, India

Abstract— *The healthcare field is usually the diagnosis of the disease. The healthcare environment comprises an enormous amount of data, such as clinical information, genetic data, and data generated from electronic health records (EHR). In case, Machine learning, Data mining and deep learning methods provide the methodology and technology to extract valuable knowledge for decision making. Heart disease (HD) is one of the cardiovascular diseases of the heart and blood vessel system. Extensive research in all aspects of heart disease (diagnosis, therapy, ECG, ECHO) has led to the generation of vast amounts of data. The disease is diagnosed earlier, and many lives might be rescued. The present study aims to conduct a systematic review of the applications of machine learning tools in the field of Heart disease research concerning Heart disease complications, prediction, saving time and diagnosis. As heart disease is usually the most recognised killer in the present day, it might be one of the most challenging diseases to diagnose. Our study's main aim is some recent works related to the usage of machine learning in predicting heart-related disease. This review forms the basis of understanding the complexity of the domain, tools and techniques employed by the researchers and the amount of efficiency achieved by the various methods recently.*

Keywords: Machine Learning, Machine Learning Classification Techniques, Heart Disease, Heart Disease Prediction, Diagnosis.

I. Introduction

The data produced by the health care industry is not mined. Data mining techniques can be used to build an intelligent medical model using data sets involving patient risk factors. Most people today experience an unhealthy and fast living style that, according to the studies, gives a jolt to the heart. The heart is the organ that pumps blood into various parts of the body through the vessels with a proper amount of oxygen and other essential nutrients. The survival of any organism relies solely on the proper functioning of the heart, and if the heart's pumping operation is troublesome, the body's main bodies, such as the brain and kidneys, undergo adverse effects. When the heart's work ceases, the person's death occurs within minutes. Various diseases attributed to our unhealthy lifestyles include heart disease, angina pectoris, congestive heart failure, cardiomyopathy, congenital heart defects, arrhythmias, myocarditis, cardiac attack, and cardiac cancer. For coronary heart disease, the cardiovascular does not have enough blood to deliver the blood to the heart because of cholesterol and fat within its arterial wall. In case of

heart attacks, the direction of the coronary artery is blocked due to the coagulation of the blood on the heart's wall. During angina, pain in the chest is caused by a blood flow that does not function properly in the heart. Other causes of cardiac disease include coronary artery disease, valvular heart disease, stroke, and high blood pressure.

Several schemes are suggested for predicting heart disease using different techniques and algorithms. The key and daunting challenge for healthcare institutions tend to provide quality treatment at an affordable price. The exact diagnosis of patients and an appropriate dose of medications must be rendered to deliver quality services on a par. Unwanted and insufficient outcomes may produce poor quality clinical diagnosis and care. Using computer-generated data or using decision support systems may be a method for reducing costs across health care facilities. The method of disease diagnosis in the field can be considered a decision process in which a medical practitioner diagnoses a new and unknown case using the knowledge of clinical evidence and their clinical experience [1].
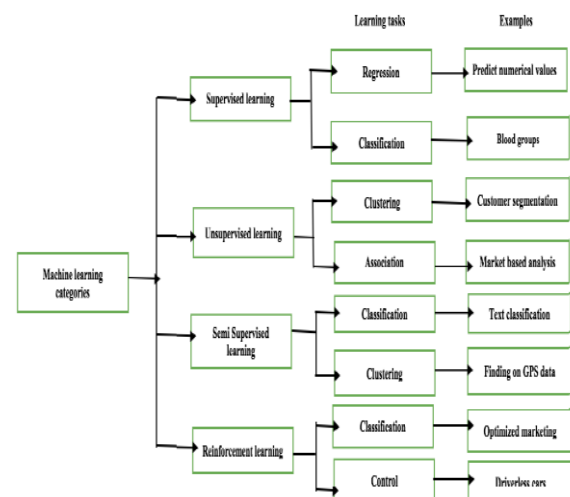

Fig1. Machine learning categories

The process can be automated to optimise this decision-making process's cost, facility, speed, precision, and reliability. Data mining is a technology implicit in the available data and has never been previously known, making it accurate in the future. In short, it analyses data from different points and gathers knowledge from it. Data mining blends statistical analysis, machine learning, and database technology to extract unknown patterns and relationships from an extensive database. Learning knowledge can be helpful in medical applications, i.e. health care. Health Informatics is a fast-

growing field in which evolving computer science and information technology are interested in medical and health data. Medical Data Mining is a challenging area that involves many misconceptions and uncertainties[2].

1.1 Categories of Machine Learning
Machine Learning Algorithms divided into categories according to their purpose are shown in Fig1[3].

1. Supervised Learning: Supervised learning is the system in which input and desired output are provided for future data processing. There are two kinds of learning tasks: regression and classification. Some of the most common algorithms are Support Vector Machines (SVM), Genetic algorithms, Decision Trees (DT), k-Nearest Neighbors (k-NN) and Artificial Neural Networks (ANN).

2. Unsupervised Learning: Unsupervised learning is used to draw inferences from datasets consisting of input data without labelled responses. In this, there are two learning tasks, Association and clustering. In order to find the correlations in the objects of a database, R Agarwal proposed Association learning. The most standard algorithm used in association rule is Apriori, and clustering groups similar kinds of datasets in which some of the most common algorithms are k-means clustering and association rule learning algorithm[4].

3. Semi-supervised Learning: Semi-supervised learning combines labelled and unlabeled data, which falls between supervised and unsupervised learning. This learning is mainly used in classifying webpage, genetic sequencing and speech recognition. Semi-supervised learning is broadly classified into two learning tasks, classification and clustering.

4. Reinforcement Learning: Reinforcement Learning is a kind of Machine Learning method. It is concerned with how software agents automatically determine the ideal behaviour within a specific context to maximise its performance. The reinforcement signal sends the reward feedback for the agent to learn its behaviour. It consists of two learning tasks, classification and control. Some applications are computer-played board games, robotic hands, and self-driving cars. The most commonly used algorithms are Q-learning, Temporal difference, and Deep Adversarial networks [5].

Heart disease: The heart is a vital organ of all living things that are essential in pumping blood into the remaining organs through the circulatory system's blood vessels. If the blood circulation in the body is improper, the organs like the brain suffer, the heart stops functioning altogether, and death occurs. Life depends entirely on the proper functioning of the heart. The term heart disease refers to diseases of the heart and cardiovascular system. Factors that increase the risk of heart disease are Family history, Age, Smoking, Poor diet, High blood pressure, High blood cholesterol, Obesity, Physical inactivity, and Hypertension.

In many cases, diagnosis is generally based on the patient's current test report & doctor's experience. Thus the diagnosis is a complex task that requires high skill & much experience. The most popular dataset used by the researchers is the Cleveland heart disease dataset obtained from the University of California, Irvine (UCI) online for machine learning. It is comprised of 303 samples, with 6 samples having missing values. The data, in its original form, have 76 features, but all the published work is likely to refer to 13 features, and the other feature outlines the effect of the disease. The most common types of heart disease are[6]:

a) Coronary artery disease (CAD) occurs when problems arise with the blood vessels due to high cholesterol, diabetes, smoking, high blood pressure and inherited from parents. Angina is a symptom of CAD, Which a chest pain and occurs discomfort in the neck, arms, shoulders, and back and even feels like indigestion.

b) Congestive heart failure (CHF) The heart's function is to pump blood, and CHF is found when the heart is not pumping at an average level.

c) Abnormal heart rhythms are a problem in the heart with electrical activity, making the heart beat too fast or too slow. Wrong rhythms stop pumping blood in the heart. Heart diseases are also called silent killers because symptoms are complicated to detect. Common symptoms include shortness of breath, uncontrolled heart palpitations, chest pain, skin discolouration, leg swelling, and dizziness. HD can be diagnosed in many ways based on the doctor's recommendation and symptoms. A few of the commonly recommended tests are given below.

1. Echocardiogram: In this, ultrasound waves monitor blood flow through the heart.
2. Holter monitoring (HM):-A portable device for 24 to 72hours worn by a patient to record the continuous ECG.
3. Electrocardiogram (ECG): This is used to diagnose the problem related to the heart's rhythm.
4. Cardiac computerised tomography (CT) scan: With the help of an X-ray, cross-sectional views of the heart are captured.
5. Cardiac magnetic resonance imaging (MRI): In this, powerful magnets and radio waves create an image of the heart and tissues surrounding it [7].

II. Related Work
Our objective of this audit segment is to introduce a writing study of picture division strategies. The fundamental objective is to feature the benefits and restrictions of these techniques are essential for

machine learning methods for heart disease prediction. According to A. N. Repaka, et.al.[8] discussed that the significant intent of this approach was to detect the coronary disorder considering primary data and information. Thus, the SHDP technique was constructed using NB, which assisted in predicting the risk factors regarding heart disease. Mobile healthcare was developing rapidly as the technology was advanced these days. The essential data were combined in a standardised form. The probability of heart disease was predicted in a patient based on extracting some attributes. Multiple knowledge abstraction methods were introduced and explained using DM models. The implementation of these DM methods was performed to predict heart-related disorders. The outcomes demonstrated that the risk factors of heart diseases were predicted efficiently using the established diagnostic system. Ankita Dewan et al. [9] presented a competent GA with a Back Propagation algorithm for predicting health-related diseases. A prototype was utilised to accomplish the significant goal of determining and extracting the indefinite knowledge of heart illness from a historical database. The presented approach effectively tackled the complicated queries to predict health-related disorders. Like this, it supported the medical practitioners in intelligent decision-making related to the clinic as the conventional decision support systems were incapable of doing so. This technique offered effective treatments due to decreased expenses for treatment. Monika Gandhi et al. [10] discussed that an enormous volume of information was extracted from clinical organisations. However, there was no proper deployment of this information. The practical methods to determine these associations and patterns were unavailable in clinical data. DM schemes were efficient and utilised as a solution in this case. Therefore, the DM methods were adopted. A comprehensive concept of diverse methods utilised to extract the information was presented with the implementation of DM. The health diseases were predicted using these methods. Several algorithms were analysed using the medical datasets. The quantification and deployment of these techniques were done in this work. Rashmi G Saboji et al. [11] established a scalable mechanism for predicting health diseases based on some attributes. For this, the clinical data were taken into consideration. The primary purpose of this work was that heart disease detection was predicted using a few features. This solution applied Apache Spark for implementing the RF algorithm. An enormous opportunity was provided to the medical experts who exploited this answer on the accessible vast database, which was frequently used to make insightful decisions. The results depicted that the recommended framework provided 98% accuracy. The comparative analysis was conducted on the RF and NB. A significant margin proved the superiority of the RF algorithm compared to traditional techniques. According to T.

John, Peter et al. [12] introduced pattern recognition and DM techniques. The medical risk predictive systems were adopted to implement these techniques. The data were modelled and classified using the DM technique. The input set consisted of essential linear mixtures of variables. Thus, they could not model nonlinear compound relations within clinical sectors. The main drawback of traditional medical scoring systems. Thus, the classification models were employed to overcome this drawback as they had the potential for detecting the complex. Nonlinear connections among diverse kinds of variables also, these models were utilised for detecting the possible interactions among predictor variables. Cincy Raju et al. [13] examined that heart disease causes life loss. A severe disability for an extended period occurred due to this disease. The individuals were quickly affected by this disease. Hence, it was essential to diagnose the patient accurately at the initial stage because this was a practical function for medical support. The status of some hospitals was in danger due to the unauthentic diagnosis. The challenging biomedical issue was to analyse the heart disorder accurately. An effectual treatment was developed using DM techniques to provide curative situations. Moreover, the deployment of various DM techniques comprising DTs, NNs, Bayesian classifiers, SVMs, Association Rule, and KNN classification was done to detect heart disorders. The best outcomes were obtained using SVM, among other algorithms. Aakash Chauhan et al. [14] examined the present world suffered from the heart disease due to which deaths occurred. Numerous patients selected the healthcare system to attain results at an accurate and quick rate. A considerable amount of data was extracted and collected daily from the medicinal organisation. The procedures were atomised to extract the data in the innovation to achieve effective information. The manual task was eliminated, and the data was extracted directly from the electronic records using the Weighted Association rule data mining method. The cost of services was reduced, and patients' lives were saved with it. This paper investigated the rule to predict the risk in patients with coronary disease. It was indicated in the outcomes that the best prediction of coronary illness was made successfully by most of the rules. According to Senthilkumar et al. [15], Machine learning (ML) effectively makes decisions and predictions from the large quantity of data produced by the healthcare industry. ML techniques are being used in recent developments in the Internet of Things (IoT) areas. Various studies give only a glimpse into predicting heart disease with ML techniques. The authors proposed a novel method to find significant features by applying machine learning techniques, improving the accuracy in predicting cardiovascular disease. The prediction model is introduced with different combinations of features and several known classification techniques. We produce an enhanced performance level with an

accuracy level of 88.7% through the prediction model for heart disease with the mixed random forest with a linear model (HRFLM).

IV. CONCLUSION

Heart disease has been one of the global health challenges in recent years. Currently, many research works are carried out to predict and diagnose heart diseases. In this study, a systematic effort was made to identify and review machine learning, data mining and deep learning approaches applied to HD research. In our study, the work can be enhanced by developing an internet application supported by the random forest algorithm and employing a large dataset compared to the one utilised in this analysis, which can provide better results and help health professionals predict the disease effectively and efficiently. Data analysis and machine learning methods have been used to forecast heart disease events and have been summarised. Determine each algorithm's prediction output and apply the method proposed for the area needed. To boost the actual performance of algorithms, use more specific feature selection methods. There are many treatment methods if patients are diagnosed with a specific type of heart disease. Our research work finds the best answer and accurate data analysis using ML.

REFERENCES

[1]. Dangare, Chaitrali S., and Sulabha S. Apte. "Improved study of heart disease prediction system using data mining classification techniques." International Journal of Computer Applications 47, no. 10: 44-48, 2012.

[2]. Thomas, J., and R. Theresa Prince. "Human heart disease prediction system using data mining techniques." In 2016 international conference on circuit, power and computing technologies (ICCPCT), pp. 1-5. IEEE, 2016.

[3]. Soni, Jyoti, Ujma Ansari, Dipesh Sharma, and Sunita Soni. "Predictive data mining for medical diagnosis: An overview of heart disease prediction." International Journal of Computer Applications 17, no. 8: 43-48, 2011.

[4]. Mirmozaffari, Mirpouya, Alireza Alinezhad, and Azadeh Gilanpour. "Heart disease prediction with data mining clustering algorithms." Int'l Journal of Computing, Communications & Instrumentation Engg 4, no. 1: 16-19, 2017.

[5]. Ramalingam, V. V., Ayantan Dandapath, and M. Karthik Raja. "Heart disease prediction using machine learning techniques: a survey." International Journal of Engineering & Technology 7, no. 2.8: 684-687, 2018.

[6]. Bhatla, Nidhi, and Kiran Jyoti. "An analysis of heart disease prediction using different data mining techniques." International Journal of Engineering 1, no. 8: 1-4, 2012.

[7]. Kanchan, B. Dhomse, and M. Mahale Kishor. "Study machine learning algorithms for special disease prediction using the principle of component analysis." In 2016 international conference on global trends in signal processing, information computing and communication (ICGTSPICC), pp. 5-10. IEEE, 2016.

[8]. Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin, "Design and Implementing Heart Disease Prediction Using Naive Bayesian", 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Vol. 1, No. 9, pp.209-217, 2019.

[9]. Ankita Dewan, Meghna Sharma, "Prediction of heart disease using a hybrid technique in data mining classification", 2nd International Conference on Computing for Sustainable Global Development (INDIACom), Vol. 4, No. 21, pp. 534-542, 2015.

[10]. Monika Gandhi, Shailendra Narayan Singh, "Predictions in heart disease using data mining", International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), Vol. 56, No. 9, pp. 4652-4660, 2015.

[11]. Rashmi G Saboji, "A scalable solution for heart disease prediction using classification mining technique", International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)Vol. 10, No. 46, pp. 676-684, 2017.

[12]. T. John Peter, K. Somasundaram, "An empirical study on prediction of heart disease using classification data mining techniques", IEEE-International Conference on Advances in Engineering, Science and Management (ICAESM - 2012), Vol. 4, No. 78, pp. 922-929, 2012.

[13]. Cincy Raju, E Philips, Siji Chacko, L Padma Suresh, S Deepa Rajan, "A Survey on Predicting Heart Disease using Data Mining Techniques", Conference on Emerging Devices and Smart Systems (ICEDSS), Vol. 61, No. 5, pp. 454-462, 2018.

[14]. Aakash Chauhan, Aditya Jain, Purushottam Sharma, Vikas Deep, "Heart Disease Prediction using Evolutionary Rule Learning", 4th International Conference on Computational Intelligence & Communication Technology (CICT), Vol. 7, No. 34, pp. 100-108, 2018.

[15]. Mohan, Senthilkumar & Thirumalai, Chandra Segar & Srivastava, Gautam," Effective Heart Disease Prediction using Hybrid Machine Learning Techniques", IEEE Access. PP. 1-1, 2019.