A Review on Heart Disease Prediction using Machine Learning

Akhand Pratap Singh
Department of Electrical and Electronics Engineering
Ajay Kumar Garg Engineering College, Ghaziabad, India
Email: akhandsingh1395@gmail.com

Dr. Bhupal Singh
Department of Electrical and Electronics Engineering
Ajay Kumar Garg Engineering College, Ghaziabad, India
Email: bhupals_21@yahoo.co.in

Abstract - Heart disease is one of the world's most dangerous illnesses for humans and has a significant effect on human life. During heart failure the heart lags pulling required amount of blood needed to other areas of the body. Precise diagnosis of heart disease on time is critical for the prevention and treatment of heart failure. It was shown that machine learning techniques are effective in helping to determine and predict a large number of data produced by the medical industry. In various research studies, the diagnosis of heart disease through the machine-learning system has been published. In this paper, a review of some recent works related to usage of machine learning in prediction of heart related disease is predicted. 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 – Heart Disease, Machine Learning, Prediction, Classification

I. INTRODUCTION

The majority of people today experience an unhealthy and fast living style that according to the studies if giving 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 brain and kidneys will undergo adverse effects. When the heart's work ceases, the death of the person takes place within minutes. Various diseases which can be attributed to our unhealthy lifestyles heart disease, angina pectoris, congestive heart failure, cardiomyopathy, congenital heart defects, arrhythmias, myocarditis, cardiac attack; cardiac cancer, etc. 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, where 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, heart valvular disease, stroke, high blood pressure, etc.

With regard to the prediction of heart disease, several schemes are suggested that are used by different techniques and algorithms. The key and daunting challenge for healthcare institutions tends to provide quality treatment at an affordable price. The exact diagnosis of patients and an appropriate dose of medications must be rendered for delivering quality services on a par. Unwanted and insufficient outcomes may produce poor quality clinical diagnosis and care. The use of computer generated data or the use of decision support systems may be a method for reducing costs across health care facilities. The method of diseases diagnosis in the field can be considered as a decision process in which a medical practitioner makes his diagnosis of a new and unknown case using the knowledge given by clinical evidence and his / her clinical experience. To optimize cost, facility, speed, precision, and reliability of this decision-making process, the process can be automated.

The rest of the paper is organized as follows. Further in this paper some previous research done in this field has been in section II. A general Framework has been presented in Section III. Concluding remarks are given in section IV.
According to SenthilKumar et.al [1], Machine learning (ML) has been shown to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry. ML techniques are being used in recent developments in different areas of the Internet of Things (IoT). Various studies give only a glimpse into predicting heart disease with ML techniques. Authors proposed a novel method that aims at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of 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 hybrid random forest with a linear model (HRFLM).

Anjan Nikhil Repaka, Sai Deepak Ravikanti and Ramya G Franklin[2] used Naive Bayesian method to design and implement heart disease prediction. To achieve this SHDP (Smart Heart Disease Prediction) is built via Navies Bayesian in order to predict risk factors concerning heart disease. The speedy advancement of technology has led to remarkable rise in mobile health technology that being one of the web application. The required data is assembled in a standardized form. For predicting the chances of heart disease in a patient, the following attributes are being fetched from the medical profiles, these include: age, BP, cholesterol, sex, blood sugar etc... The collected attributes acts as input for the Navies Bayesian classification for predicting heart disease. The dataset utilized is split into two sections, 80% dataset is utilized for training and rest 20% is utilized for testing. The proposed approach includes following stages: dataset collection, user registration and login, classification via Navies Bayesian, prediction and secure data transfer by employing Advanced Encryption Standard. Thereafter result is produced. The research elaborates and presents multiple knowledge abstraction techniques by making use of data mining methods which are adopted for heart disease prediction. The output reveals that the established diagnostic system effectively assists in predicting risk factors concerning heart diseases.

According to Ed-Daoudy[3], early detection of heart diseases and continuous monitoring can reduce the mortality rate. The exponential growth of data from different sources such as wearable sensor devices used in Internet of Things health monitoring, streaming system and others have been generating an enormous amount of data on a continuous basis. The combination of streaming big data analytics and machine learning is a breakthrough technology that can have a significant impact in healthcare field especially early detection of heart disease. This technology can be more powerful and less expensive. A real-time heart disease prediction system has been proposed based on apache Spark. The system consists of two main sub parts, namely streaming processing and data storage and visualization. The first uses Spark MLlib with Spark streaming and applies classification model on data events to predict heart disease. The seconds uses Apache Cassandra for storing the large volume of generated data.

Amin Ul Haq et.al[4] have come up with Hybrid Intelligent System Framework for the Prediction of Heart Diseases. The authors asserted that noninvasive-based methods such as machine learning are reliable and efficient. A Machine-learning-based diagnosis system for heart disease prediction by using heart disease dataset was developed using seven popular machine learning algorithms, three feature selection algorithms, the cross-validation method, and seven classifiers performance evaluation metrics such as classification accuracy, specificity, sensitivity, Matthews’ correlation coefficient, and execution time. The proposed system can easily identify and classify people with heart disease from healthy people. Additionally, receiver optimistic curves and area under the curves for each classifier was computed. The authors validated performance of the proposed system on full features and on a reduced set of features. The features reduction has an impact on classifiers performance in terms of accuracy and execution time of classifiers.

Md. Shahriare Satu et.al [5] present that Heart Disease is one of the leading diseases that causes enormous loss of lives all over the world. Some unusual approaches to find out significant factors of heart diseases have been considered by the authors. They have used two heart disease data (Cleveland & Hungarian) and both of them are divided into 33%, 65% and 100% data. Values of different range of individual attributes in these data are determined to find out relevant factors of this disease. Then, different semi supervised learning algorithms such as Collective Wrapper, Filtered Collective and Yet Another Semi Supervised Idea are used to analyze heart disease data. Metrics of these classifiers like accuracy, f-measure and area under ROC have been calculated to justify individual classifiers and specify the best semi supervised learning algorithm. This algorithm is explored significant and irrelevant factors of heart disease by removing attributes one after another sequentially and observing the outcomes of classification. Experimental results on two real data demonstrates the effectiveness and efficiency of the analysis.

R. Sharmila et al, [6] suggested using a non-linear heart disease prediction classification algorithm. Big data tools like the Hadoop Distributed File System (HDFS), Mapreduce and SVM are proposed for cardiac disease prediction with optimized attribute description. This thesis explored the use of various methods of data mining for cardiac...
disease prediction. This proposes to use HDFS to store large amounts of data in different nodes and run the prediction algorithm concurrently using SVM in more than one node. SVM is used in parallel, providing better time than sequential SVM computing.

The prediction and study of Heart Disease Use Techniques for Data Mining was suggested by Chala Beyene et al, [7]. The principal aim is to predict that heart disease will arise in a limited period of time for an automatic early diagnosis of the disease. In the health system with professionals who have no experience and expertise, the suggested approach is also important. It uses various medical features including blood sugar and heart rate, age, sex, some of the features to decide whether you have heart disease. WEKA software is used to measure the performance of data sets.

C. Sowmiya and P.Sumitra[8] believe that It is essential to have a frame work that can effectually recognize the prevalence of heart disease in thousands of samples instantly. The authors evaluated the potential of nine (9) classification techniques for prediction of heart disease. Namely decision tree, naïve Bayesian neural network, SVM,ANN, KNN. My proposed algorithm of Apriori algorithm and SVM (support vector machine) in heart disease prediction. Using medical profiles such as a age, sex, blood pressure, chest pain type, fasting blood sugar. It can predict likelihood of patients getting heart disease. Based on this, medical society takes part interest in detecting and preventing the heart disease. From the analysis it have proved that classification based techniques contribute high effectiveness and obtain high accuracy compare than the previous methods.

P.Sai Chandrasekhar Reddy et al, [9] suggested predicting heart disease using the data mining ANN algorithm. As the cost of diagnosing cardiovascular diseases has increased, a new method needs to be built to predict cardiac diseases. After an assessment on the basis of various parameters such as pulse rate, blood pressures, cholesterol etc., a prediction model can be used. The system's accuracy is shown in java.

Sarath Babu et al.[10] have suggested that medical data mining will explore the secret trends of medical data sets with great potential. These trends may be used to diagnose clinically. The data must be obtained in a structured manner. From the healthcare profiles, the patient can be predicted by 14 characteristics, such as age, sex, blood pressure and blood sugar etc. Such attributes are fed in the prediction of heart disease with K-means algorithms, MAFIA algorithms and decision tree classification, using data mining technology to diagnose heart disease; it can achieve as much precision as it can when diagnosing a heart condition. By this medical industries could offer better diagnosis and treatment of the patient to attain a good quality of services. The main advantages pointed in the research are: early detection of heart disease and its diagnosis correctly on time and providing treatment with affordable cost.

M.A.Zabbar and Shirina Samreen[11] said that heart disease is a world-wide leading cause of death. According to authors, cardiac diagnosis is a tedious task and a smart decision-making support system for prediction of diseases is required. Methods of data mining are also used to assess the patient's normal or heart condition. Hidden Naïve Bayes is a type of data mining that relaxes the conventional independence of the Naïve Bayes. The model proposed notes that the Hidden Naïve Bayes can be applied to the classification of cardiac diseases.

Marjia et al, [12] developed heart disease prediction using KStar, j48, SMO, and Bayes Net and Multilayer perception using WEKA software. Based on performance from different factor SMO and Bayes Net achieve optimum performance than KStar, Multilayer perception and J48 techniques using kfold cross validation. The accuracy of these algorithms is also not satisfactory. The exactness of the results is therefore enhanced to enable a better diagnostic decision.

M.Raihan et.al[13] has developed a simple approach to forecasting the probability of ISD using smartphone. The application of clinical data collected from patients admitted with IHD has developed Android-based prototype applications. Clinical data from 787 patients were examined and associated with risk factors such as high blood pressure, diabetes, elevated cholesterol, smoking, family history, obesity, depression and current clinical symptoms that may indicate the underlying unrecognized IHDs. Data were extracted from and a score generated using data mining technology Risks for IHD are graded as low, medium and strong. The authors found that a substantial association between low-and medium-and high-grade cardiac event was observed for patients whose data were collected to produce the ratings, respectively; p=0.0001 and 0.0001. They provide an simple way of detecting the danger of IHD and are prone to cardiology in order to avoid sudden deaths. At present there are some restrictions on the available resources that make them unused by the population.

III. GENERAL FRAMEWORK FOR HEART DISEASE CLASSIFICATION:
The general framework which has been employed to perform the heart disease classification has been shown in Figure 1 below. The dataset of heart disease consists of pre collected data of patients and is available on the internet. There are several dataset repositories like the UCI Machine Learning repository, Cleveland dataset etc. which have been used by the researchers.
The dataset contains various parameters of the patients like age, sex, blood pressure, type of chest pain etc. As shown in the figure the dataset is first preprocessed and cleaned to remove any missing values and uniform the parameters. Then it is passed on to feature extraction stage in which the features are extracted for all the individuals. The classifiers are then utilized to classify the various features based on ground truth taken from other sources for classification. Machine learning models can be developed using numerous techniques like SVM, KNN, Artificial Neural Network etc. These machine learning models are trained based on the features to classify the dataset as belonging to either healthy or affected patients.

IV. CONCLUSION

Diagnoses of cardiac disease are the sternest of challenge in the medical profession. It is based on the thorough review by medical experts of the various clinical and medical data of the patient. Because of the advances in machine learning and IT, researchers and medical experts are interested in creating a highly accurate, efficient and supportive predictive framework for the prediction in heart disease. Data analysis and machine learning methods have been used to forecast heart disease events and have summarized. Determine each algorithm's prediction output and apply the method proposed for the area needed. To boost the exact performance of algorithms, using more specific methods of feature selection. If patients are diagnosed with the specific type of heart disease, there are many treatment methods.

REFERENCES


