International Journal of Theoretical & Applied Sciences, 9(2): 333-337(2017)

ISSN No. (Print): 0975-1718
ISSN No. (Online): 2249-3247

Cardiovascular Disease Prediction using Machine Learning Techniques: A Review

Hardeep Singh* and Sandeep Kaur
Department of Computer Science & Engineering,
Guru Nanak Dev University Regional Campus, Sathiala, (Punjab) India.

(Corresponding author: Hardeep Singh)*
(Received 05 September, 2017, accepted 16 December, 2017)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Cardiovascular disease (CVD) is a blanket term for problems with the heart and blood vessels. Common side effects include atherosclerosis, an accumulation of fat in the arteries, and a higher risk of blood clots. According to estimates, cardiovascular disease is the main reason people get sick and die. In order to keep people healthy and save lives, clinical data analysis and healthcare professionals struggle greatly to anticipate and diagnose disease. The healthcare sector collects enormous amounts of information that can be utilized to make knowledgeable decisions concerning the diagnosis of cardiovascular disease. Deep neural networks and AI algorithms may also be used to research and diagnose cardiovascular disease. Deep learning is essential in the medical sector for effectively identifying various ailments and treating health issues. This article will discuss the subject of cardiovascular disease. There are many ways to classify data, including utilizing KNN, SVM, Nave Bayes, or Random Forest. In this article, we evaluate the benefits, drawbacks, and applications of each approach for heart disease prediction.

Keywords: Cardiovascular disease, Machine learning, Supervised learning, Unsupervised learning.

I. INTRODUCTION

In modern medicine, cardiac illness is one of the biggest worldwide health issues. According to the proverb of the twenty-first century, life expectancy is increasing and there has been a substantial global shift in the causes of heart disease bereavement [1]. Today, it is estimated that there will be a global decline of around 30%, with the high-income country seeing a decrease of about 40% and the low- and middle-income countries experiencing a decrease of about 28%. This ongoing transformation is happening globally among all races, ethnic groups, and nations at an even greater rate than the last century, driven by economic development, suburbanization, and related circadian life changes. Heart failure rates have increased tremendously recently as a result of modern lifestyle changes. According to a recent study, heart failure symptoms have increased over the past 25 years [2].

According to a recent study, chronic noninfectious diseases like heart illness are one of the main causes of death worldwide. Globally, there has been a tremendous shift in people's health status, which has led to an increase in cardiac illness. The leading cause of death worldwide every day is now heart illness. The substantial shift in people's health status around the world is what causes the global increase in cardiac illness. Over the past two decades, heart disorders have alarmingly increased day by day and are now one of the leading causes of death in the majority of the world's nations [3]. Heart failure prognosis has historically been a challenging undertaking on the eve of high-cost ratios. The expense of the many clinical methods and contemporary imaging for the diagnosis of heart disease is too high. Chest discomfort, dyspnea, tiredness, edema, palpations, syncope, cough, hemoptysis, and cyanosis are a few examples of the primary symptoms linked to cardiac disease [4].

Different Types of Cardiac Diseases. Heart disorders fall into a number of categories. Based on the clinical data, these groups are broadly categorised as myocardial infarction, heart failure, heart arrhythmia, angina pectoris, cardiomyopathy, and atrial fibrillation [5].

1. **Coronary Artery Disease:** The discomfort of coronary artery disease is brought on by reduced blood flow. The veins will be harmed by the arteries' reduced blood flow, which will also make the heart's normal systolic and diastolic functions uncomfortable.
2. **Acute myocardial infarction:** Acute myocardial infarction is the medical term for cardiac arrest. Fatty substances in the blood can alter the pace of flow during a cardiac arrest, causing tissue damage to the arteries. Atherosclerosis may prevent the arteries from supplying the body with oxygenated blood, which will cause other organs to malfunction.
3. **Chest Pain (Angina):** Angina is the medical term for chest pressure. Medical assistance is frequently required as an emergency for the patients. If we feel this type of discomfort, patients must be immediately treated with ventilators. The pressure on the blood vessel walls and insufficient blood flow will affect the blood vessels. It will result in chest pain by increasing strain on the blood vessels.

Risk Factors of Heart Diseases. Numerous factors are escalating the risk of developing heart disease [6]. Age, gender, previous patient case reports, cholesterol levels, smoking, and diabetes are stated as the components.

1. **Family History of Patients:** Previous cases of patients reporting cardiac discomfort along with erratic blood sugar levels increased their susceptibility to heart disease. Obese patients frequently have aberrant lipid profiles and high blood pressure; these alterations increase the risk of developing atherosclerotic artery plaques.
2. **High Blood Pressure and High Cholesterol:** The idea that uneven blood flow increases the risk of heart illness is explained in Civicioglu and Besdok [6]. Higher blood cholesterol levels increase the risk of heart disease. High-cholesterol fatty compounds cause atherosclerosis. Lipid profiles come in two different varieties. Low level lipid profiles and high-level lipid profiles are what they are.
3. **Smoking:** Smoking increases the likelihood of heart dysfunction by causing patient blood enzymes to coagulate with one another, which is a substantial risk factor for cardiac arrest.
4. **Diabetes Mellitus:** The molecular causes, suggested treatments, and epidemiological aspects of diabetes are explained in Civicioglu and Besdok [6]. Diabetics develop the condition as a result of the coordination of blood sugar flow. If diabetic individuals do not maintain their blood sugar, blood pressure, and cholesterol levels, they run the risk of developing heart disease.

Requirement for Medical Technology Advancements. Modern electronic health technologies have made it possible for both humans and machines to produce massive amounts of data in daily life. Online data can be kept for later examination and is created at a pace of 2.5 quintillion bytes per day for every online transaction. Models for data extraction and categorization utilizing machine learning and statistical methodologies are developed to infer knowledge from the huge and complicated datasets. Modern technologies including grid computing, distributed computing, cloud computing, social computing, wireless communication, parallel computing, green computing, cluster computing, mobile communication, telecommunication, and satellite communication are essential to the practice of modern medicine [6]. Internet-based technologies have a significant impact on telemedicine.

Overview of Machine Learning Techniques. The uses for maintaining medical data are immensely diverse, and at a higher level, one can look at a patient's private data like temperature, glucose, blood pressure, and sugar level to detect and treat diseases sooner [7]. To give appropriate treatments, a single characteristic, such as blood pressure or blood sugar level, can be evaluated at the low level. To obtain the relevant information from the already-existing, raw data from the real world, many approaches and techniques are applied in all domains. In this article, we concentrate on various machine learning methods that are useful for predicting cardiac disease. Kamal Kant [8] explains how machine learning techniques are essential to the processes of data analysis and prediction. However, due to their size and complexity, standard approaches cannot handle the extraction of medical data.

Machine learning is widely used in daily life to aid in crucial decision-making processes like clinical and medical diagnosis. It is the greatest choice for giving difficult issues with plenty of data and features a solution. These algorithms carry out the classification process by creating the classification models from the input dataset. The models of machine learning techniques are displayed in Fig. 1. It uses approaches for both supervised and unsupervised learning. In the case of supervised models, classification models are constructed with known input and output data. While unsupervised learning focuses on the data's hidden patterns [9].

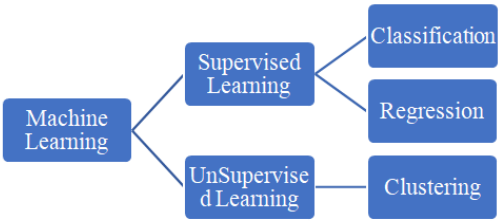

Fig. 1. Machine Learning Models.

Supervised techniques. These methods use largely certain known evidence to generate the training model. In supervised learning, classification and regression techniques are used to create the predictive models. Support vector machines (SVM), naïve bayes classification methods, decision trees, closest neighbor, logistic regression, discriminant analysis, and neural networks are a few examples of common classification algorithms [10].

Unsupervised techniques. The outcome is predicted using unsupervised learning on the unlabeled datasets. The most popular unsupervised learning technique is clustering. Some of the simple clustering strategies include hierarchical clustering, the K-means algorithm, the K-medoids algorithm, and hidden Markov models. Numerous supervised and unsupervised learning algorithms are available; however, their applications vary greatly depending on the situation. Therefore, choosing a suitable machine learning method yields superior predictions and classification outcomes. But selecting the best algorithm can be challenging. The process of accurately classifying the incoming data and assigning it to the appropriate classifications is known as classification in predictive analysis. There are often two categories of data, labelled and unlabeled. There are many predictor attributes in the labelled data and just one target attribute. The class label is represented by each value of the target characteristics. Only the predictor attributes are present in the unlabeled attributes.

Using classification models created from labelled examples (historical data), the primary goal of the classification process is to correctly predict the class of the unlabeled data [11].

As the training data model, a summary of the relationships between the data points is constructed. When the target values are unknown, this training data model is utilized to forecast the target values. The accuracy of the classification process is then determined by comparing the anticipated values to the known values or labelled data. The process of testing a data model is described as such, and the data used for testing is referred to as the test data or the evaluation data. It assesses how well the prediction procedure worked [12].

II. LITERATURE REVIEW

The goal of this paper is to show how current research on using computational intelligence approaches to diagnose cardiac problems has evolved. Over 1.7 billion people worldwide have died as a result of this exceedingly deadly illness. The factors of uncertainty, early prognosis, and disease diagnosis are important. Diagnostic systems are now digitalized thanks to recent developments in medical technology, advanced computing methods, inexpensive storage methods, and internet access. A new era in the field of diagnostics has been created by the digitization of medical data. These unprocessed patient medical records are crucial to analyze, investigate, and use with various classification approaches given the enormous expansion of digital information. For the benefit of users of electronic health records, health informatics integrates mathematical models, algorithms, and analysis to improve the quality of healthcare services. Machine learning is an area of health informatics that is essential to the management and analysis of health data. This essay primarily focuses on categorization based on forecasts of heart disease. In this section, numerous machine learning techniques that use medical diagnosis are briefly reviewed.

Rashedur [13] discovered a classifier model of advancing feature inclusion along with back-elimination using a variety of data sets, including datasets on arrhythmia, heart illness, and ECG data. Experiential findings show that the feature choices improved classification methods' accuracy and reduced the volume of in-takes.

Sanjay and Rutuparna [14] explains an ANN-based fuzzy inference system that uses linear LDA and ANFIS to forecast endangerment and gives endangerment information. The hybrid approach of classification is used by the inference system, and the findings indicate that it offers relatively superior efficiency and performance measure than the existing methodologies. Additionally, it aids in the earlier diagnosis and treatment of coronary heart disorders.

Shanthakumar [15] proposed a design for a system that makes precise predictions using a firefly-based algorithm using rough sets. The difficulties of uncertainties and high-dimensional elements in heart disease datasets are lessened by the fusion of fuzzy and roughest theoretical principles. The fuzzy learning model with the roughest foundation makes it simpler to identify ideal solutions while requiring less processing. When compared to the support vector machine and ANN for heart disease prediction and the pharmaceutical procedure, the results are better.

Shilaskar and Ghatol [16] created a method to anticipate ventricular arrhythmia. In order to forecast discomfort, this work offers a fully integrated ECG signal processor. In order to predict the occurrence of ventricular arrhythmia in people, a specific set of ECG features is used. In order to examine and extract the fiducial points, it recognizes and marks the ECG waves (PQRST). Real-time, adaptive strategies are used to create this process. The application of this technique gives high sensitivity and precise measurements while handling the variability in ECG signals effectively. The American Heart Association's database of recorded cardiovascular signals is used to assess the system's performance. In comparison to the current approaches, the simulation findings offer comparable greater accuracy measures.

A method for acute diagnosis and prediction of coronary diseases is given in Shouman *et al.* [17]. The primary objective of the work is to recognize heart disease and to assist medical professionals in earlier diagnosis and treatments. This review clearly states that the use of machine learning techniques across healthcare systems dramatically assists the medical professionals to make clear and concrete decision and diagnosis. Table 1 shows the summary of Machine learning algorithms in healthcare.

III. GAPS IN LITERATURE

According to the available literature, current machine learning models have at least one of the following problems:

1. The majority of recent studies have ignored the impact of feature selection methods during training and testing. The literature has shown that an effective feature selection method has the potential to improve the performance of machine learning models.
2. The adoption of hopeful-sounding strategies to improve the performance of current machine learning models for cardiovascular disease prediction has been largely disregarded by researchers.
3. Another significant gap in the literature is the tweaking of parameters. An effective adjustment of the parameters might enhance the performance even more.
4. It has been noted that the bulk of academics currently working on this issue have concentrated on developing heuristic machine learning models to forecast cardiovascular illness.
5. The majority of the approaches now in use make generic predictions without identifying the precise illness.

There are many methods available to categorize, cluster, and find the hidden patterns in the health data. In addition, applying machine learning techniques to medical information systems has further effects on privacy, heterogeneous data management techniques, interoperability, and the processing of both organised and unstructured medical records. The medical information system makes it necessary to use the right methodology and machine learning approaches to address new issues in the healthcare industry. The main topics of the dissertation are methods for managing and predicting cardiac disease.

CONCLUSION

The cardiovascular disorders and their many forms were the focus of this study. Additionally, we looked at a number of approaches and models that help fill the gap between medical professionals and services in terms of accuracy and earlier medication concerns. In this work, we compared cutting-edge machine learning methods for predicting and diagnosing cardiovascular disorders. In our upcoming work, we'll create the framework for predicting cardiovascular disorders using a variety of machine learning techniques.

Table 1: Summary of Machine Learning Algorithms in Healthcare.

Application of Classification Algorithms in Heart Disease Prediction				
Classifier	**Algorithm Models**	**Advantages**	**Disadvantages**	**Applicability**
Decision Tree	C4.5, ID3, C5, Random Forest	Simple, easy to implement, good with categorical data	Model overfitting, computationally expensive and complex	High
Artificial Neural Networks	Back propagation, Convolutional Neural Networks	Efficient, varies with different types of ANN used	Hard to implement, expensive	High
Swarm Intelligence	Meta-heuristic, Particle Swarm Optimization	Use Lesser memory, Efficient and optimized solution	Search ability degrades with complex datasets	Medium
K-Nearest Neighbour	K-medoids and k-means	Simple and widely used method	Difficult to assume K value, Time consuming	Medium
Bayesian Classifier	Naive Bayes	Simple, computationally fast and work well with different dimensions	Accuracy is highly concerned with assumptions	High
Support Vector Machine	Linear and Non-linear SVM	Performs well with linear separations, effectively handle high dimension data	Sometimes suspected to model overfitting	High
Logistic Regression	Multinomial, ordinal Regression	Low variance, provides probability of outcomes	Not a classifier but performs classification, widely usedfor performance analysis	Low

REFERENCES

[1]. Ashwini, S. and Naik, C. (2016). Different data mining approaches for predicting heart disease. *International journal of innovative research in science, engineering and technology*, **3**(2), 277–281.

[2]. Aydin, S. (2016). Comparison and evaluation data mining techniques in the diagnosis of heart disease. *Indian journal of science and technology*, **6**(1), 420–423.

[3]. Bhatla, N. and Jyoti (2012). An analysis of heart disease prediction using different data mining techniques. *International journal of engineering research and technology*, **1**(8), 10–14.

[4]. Chandna, D. (2014), 'Diagnosis of heart disease using data mining algorithm. *International journal of computer science and information technologies*, **5**(2), 1678–1680.

[5]. Chebbi, A. (2016). Heart disease prediction using data mining techniques. *International journal of research in advent technology*, **25**(3), 781–794.

[6]. Civicioglu, P. and Besdok, E. (2013). A conceptual comparison of the cuckoo-search, particle swarm optimization, artificial bee colony algorithms. *Artificial Intelligence Review*, **1**, 32–38.

[7]. Eberhart and Kennedy (1995). Particle swarm optimization. *Journal of IEEE transactions on swarm optimization*, **5**(4), 1942–1948.

[8]. Kamal Kant, D. G. (2014). Review of heart disease prediction using data mining classifications. *International Journal for Scientific Research*, **2**(3), 509–517.

[9]. Kavitha, M. K. (2010). Modeling and design of evolutionary neural network for heart disease detection. *International journal of computer science*, **7**(5), 78–84.

[10]. Manimekalai, M. (2014). Study of heart disease prediction using data mining. *International journal of advanced research in computer science and software engineering*, **4**(1), 121–128.

[11]. Patil, J. (2016). Heart disease prediction using machine learning and data mining technique. *International journal of computer science*, **7**(1), 129–137.

[12]. Pencina, M. J. and Derpan (2010). Heart disease diagnosis using machine learning approach. *Journal of NIH public access*, **119**(24), 3078–3084.

[13]. Rashedur, Rahman, F. (2013). Comparison of various classification techniques using different data mining tools for diabetes diagnosis. *Journal of Software Engineering and Applications*, **6**(1), 85–97.

[14]. Sanjay Agrawal and Rutuparna Panda, S. B. (2013). Entropy based optimal multilevel thresholding using cuckoo search algorithm. *Swarm and Evolutionary Computation*, **11**, 16–30.

[15]. Shanthakumar, P. (2009). Intelligent and effective heart attack prediction system using Data mining and artificial neural network. *European journal of scientific research*, **31**(4), 642–656.

[16]. Shilaskar, S. and Ghatol, A. (2013). Expert systems with applications feature selection for medical diagnosis: evaluation for cardiovascular Diseases. *Journal of expert system with application*, **4**(10), 4146–4153.

[17]. Shouman, M., Turner, T. and Stocker, R. (2012). Applying k-nearest neighbor in diagnosing heart disease patients. *International journal of information and education technology*, **2**(3), 220–223.