STM JOURNALS

Research & Reviews: A Journal of
Medical Science and Technology
https://journals.stmjournals.com/rrjomst/

ISSN: 2319-3417 (Online)
ISSN: 2349-1272 (Print)
Volume 13, Issue 1, 2024
DOI (Journal): 10.37591/RRJoMST

Review RRJoMST

Review on Machine Learning Techniques for Heart Failure Analysis in Health Industries

Amit Arya[1]*, Vineet Richhariya[2], Sadhna K. Mishra[3]

Abstract

There are few bodily components as crucial as the heart. It aids in the filtration and distribution of blood to every area of a body. The world's biggest cause of death is heart disease. It has been reported that symptoms include breathing difficulties, fast heartbeat, and chest discomfort. They analyze this data on a regular basis. This review begins with a brief introduction of cardiac disease and the present methods used to treat it. It also provides a concise overview of the most important ML methods currently published for the forecasting of CVD. Data analytics is helpful for making predictions with more data, and it aids the medical Center in forecasting a variety of ailments. The monthly data retention rate is quite high. The collected information may serve as a foundation for disease outbreak prediction. Predictions and judgements have become feasible because to the massive amounts of data produced by the healthcare business. Cardiovascular disease prediction and prevention is the greatest data analytic problem. The abundance of data generated by healthcare facilities has prompted the development of machine learning algorithms that can make accurate forecasts and sound decisions. The area of Machine Learning (ML) within AI focuses on teaching computers new skills and tasks with little to no human oversight. Finding patterns and making predictions are the goals of data analysis and statistical approaches. This research compared many Machine Learning models to find the most effective one for making more accurate predictions of cardiovascular disease (CVD). Finally, the survey delves into several research gaps and difficulties, providing researchers with valuable information to inspire better future work on HD prediction using ML models.

Keywords: Heart failure, Cardiovascular, Risk factor, Machine learning, Cardiovascular Disease.

INTRODUCTION

The human body's most vital organ is the heart. It's a muscular organ that sits just beneath the breastbone, to the left of it. Heart disease accounts for more than 31% of all deaths annually, making it the world's leading cause of death, according to the WHO. This figure comes to around 17.9 million individuals. Heart and blood artery diseases, such as rheumatic heart disease and cerebrovascular disease, are together referred to as cardiovascular diseases. A heart attack or stroke accounts for four out of every five fatalities caused by cardiovascular disease. These days, heart disease is the leading cause of death worldwide. So, it's crucial to detect heart disease early on, and leading a healthy lifestyle may help you avoid it [1].

***Author for Correspondence**
Amit Arya
E-mail: amitaryan371@gmail.com

[1]Research Scholar, Department of Computer Science Engineering, Lakshmi Narain College of Technology, Bhopal, Madhya Pradesh, India
[2]Professor, Department of Computer Science Engineering, Lakshmi Narain College of Technology, Bhopal, Madhya Pradesh, India
[3]HOD, Department of Computer Science Engineering, Lakshmi Narain College of Technology, Bhopal, Madhya Pradesh, India

Received Date: February 23, 2024
Accepted Date: February 29, 2024
Published Date: April 16, 2024.

Citation: Amit Arya, Vineet Richhariya, Sadhna K. Mishra. Review on Machine Learning Techniques for Heart Failure Analysis in Health Industries. Research & Reviews: A Journal of Medical Science and Technology. 2024; 13(1): 29–43p.

The cardiovascular disease group includes heart failure. Without an adequate supply of blood, the heart cannot pump blood effectively, leading to heart failure. Hospitalization is the norm for individual with heart failure. In both the US and Europe, heart failure affects around 12% of the

population. The National Health Survey indicated that cardiovascular disease was a leading cause of death in Indonesia (26.4%) and ranked ninth overall; nonetheless, there is a lack of epidemiological data about heart failure in Indonesia. The importance of predicting heart failure has grown among medical professionals as a result of the high monetary worth of organs like the heart. But up to this point, clinical heart failure prediction hasn't been very accurate [2]. The World Health Organization (WHO) lists diabetes, heart disease, and breast cancer as the three most deadly chronic illnesses. The majority of this disease's victims are from wealthy nations. Coronary artery disease is another name for heart disease. Even while most people with heart disease have symptoms including chest pain and high or low blood pressure, everyone is different [3].

Problems in the heart's capacity to circulate blood throughout the body are hallmarks of cardiac irregularities or diseases. Symptoms, medical history, physical exam results, and laboratory tests must all be considered in order to arrive at a diagnosis of heart failure. However, machine-learning approaches may now be used to assist in the diagnosis and prognosis of this condition, thanks to technological advancements and machines' data-processing capabilities. Many in the medical field are interested in machine learning approaches because of their potential to improve cardiac disease diagnosis and prognosis. Machine learning algorithms can sift through mountains of patient records in search of trends and correlations that might otherwise go unnoticed by medical professionals [4]. Disease risk assessment may benefit greatly from the insights provided by these algorithms, which can use computer models to provide accurate forecasts.

What follows is the remainder of the paper: Heart failure and cardiovascular heart disorders are discussed in Sections II and III, respectively. The kinds of machine learning are analyzed in section IV. The literature review on heart disease forecasting utilizing ML is presented in section V. This paper concludes and discusses potential future directions in the last part.

OVERVIEW OF HEART FAILURE

Heart failure develops when the heart's minute blood pumping capacity is inadequate to provide the body with oxygen and nutrients. Any one or both sides of the heart might be affected by heart failure. As a outcomes of a heart's diminished pumping capacity, blood may leak out of the organs and tissues. Despite popular belief, heart failure does not indicate that the heart has ceased beating; rather, it indicates that the heart has become weak and unable to pump blood effectively. Heart failure is more frequent in the elderly and those with underlying health issues, however it may happen to anybody at any age. While heart disease often becomes worse with age, it is manageable with the right lifestyle choices and may be left untreated for a long time. More than 50% of Americans diagnosed with heart failure do not survive beyond 5 years after their diagnosis; 550,000 new patients are diagnosed with heart failure annually [5].

An estimated 26 million individuals throughout the globe are already living with heart failure, and the number of those affected is only going up. As the population ages, heart failure-related complications and deaths will increase. The number one killer on a global scale is cardiovascular disease. Heart disease and stroke kill more individuals every year than any other medical condition. The majority of cardiovascular diseases are preventable if risk factors are identified and removed. People with cardiac disease or high risk factors (e.g., hypertension, diabetes, hyperlipidemia, etc.) must be informed promptly so that they may get the therapy they need. At this early stage, treatment options are limited to medication or behavioural adjustments (risk factor removal). Cardiovascular disease prevention may be approached from two angles: the population-wide and the individual [Figure 1].

A comprehensive strategy to combat cardiovascular illnesses might include taxing goods that contribute to illness or increase risk factors, such as tobacco, certain foods, salts, sugars, etc. The next step is to inspire individuals to adopt beneficial lifestyle habits like going for walks and enjoying sports. The distribution and use of alcohol need to be reduced. The aforementioned examples should be included into teaching to promote healthy lifestyle choices in youngsters and indirectly contribute to a

decrease in patient numbers. Individuals with high overall cardiovascular risk or with levels of specific risk factors exceeding established limits, such as hypertension and hypercholesterolemia, should be the focus of health-care measures aimed at preventing initial heart attacks and strokes [6]. An AI method, namely ML or classification, may be used to identify cases of heart failure using data that has already been collected.

CARDIOVASCULAR DISEASES

A wide range of medical issues affecting the heart and blood vessels are collectively referred to as "cardiovascular disease" (CVD). Angina and heart attacks are both disorders that fall under the umbrella term of coronary artery disease (CAD). Heart failure, thromboembolic disease, peripheral artery disease, venous thrombosis, carditis, hypertension, rheumatic heart disease, aortic aneurysms, cardiomyopathy, abnormal heart rhythms, congenital heart disease, valvular heart disease, and carditis are all examples of cardiovascular diseases. Different diseases have different underlying processes. In the US, CVD is the leading cause of death, with dietary risk factors estimated to be accountable for 53% of CVD-related fatalities. One of the main causes of peripheral artery disease, heart disease, and stroke is atherosclerosis. Reasons for this include, but are not limited to, hypertension, smoking, diabetes mellitus, inactivity, obesity, high cholesterol levels, an unhealthy diet, excessive alcohol use, and inadequate sleep [7]. Approximately one-fifth of all CVD deaths are caused by hypertension; another nine percent are attributed to smoking, six percent to diabetes, six percent to a lack of physical exercise, and five percent to obesity. Strep throat may lead to rheumatic heart disease if not managed.

The prevalence of cardiovascular disease (CVD) may be reduced by as much as 90 percent if certain measures are taken. Positive lifestyle choices, such switching to a healthy diet, getting regular exercise, quitting smoking, and consuming less alcohol, may lower the risk factors for CVD. More than that, it aids in the management of risk factors including diabetes, hyperlipidemia, and hypertension. One possible benefit of antibiotic therapy for strep throat is a decreased risk of rheumatic heart disease in later years. Those who are otherwise healthy may not benefit from taking aspirin.

Types of Cardiovascular Disease

The following kinds of Cardiovascular heart diseases exist, and every one contains its symptoms and action. Figure 2 shows the types of heart disease.

Rheumatic Heart Disease (RHD)

The heart valves may be irreparably damaged by rheumatic fever, leading to rheumatic heart disease. In particular, it has the potential to impact the skin, brain, joints, heart, and connective tissues throughout the body. Typically affecting youngsters between the ages of 5 and 15, RHD may permanently damage the heart valves, which is a co-morbidity. In underdeveloped nations, streptococcal infections have been rare, although they may rise a risk of heart failure in rheumatic if not treated.

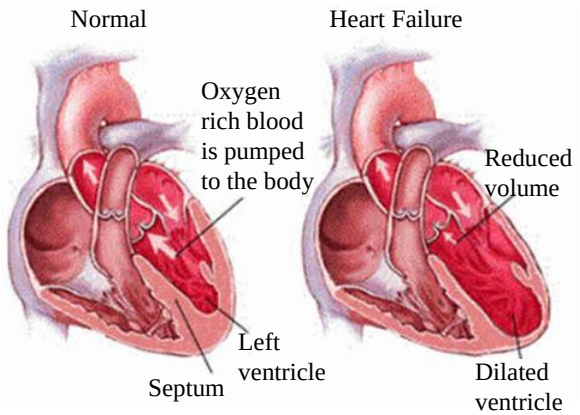

Figure 1. Heart failure.

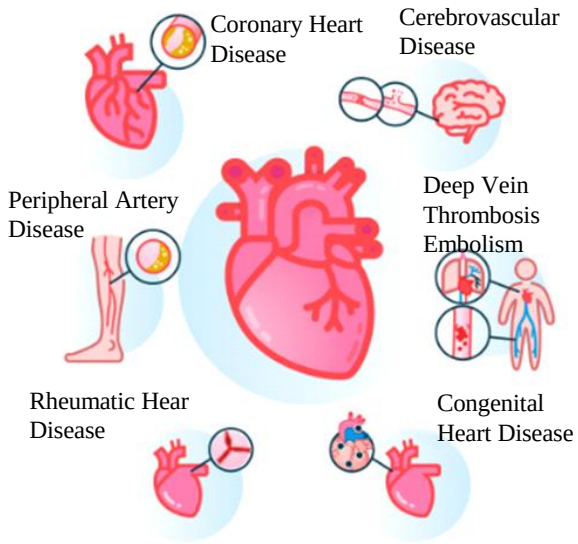

Figure 2. Types of Cardiovascular heart Disease

Ischemic Heart Disease (IHD)

An inability of a heart to get enough blood flow as a result of narrowing or occlusion of the coronary arteries is the hallmark of ischemic heart disease, which goes by many names. In cases with suspected IHD, such as angina pectoris, further diagnostic procedures, such as coronary angiography, may be necessary. Atherosclerosis, the buildup of plaque in blood vessels, is the most common cause of narrowing, but blood clots and constriction of the vessels may also contribute. A myocardial infarction (MI), sometimes referred to as a heart attack, happens when there is no longer any blood supply to the heart muscles, which results in cell necrosis.

Congenital Heart Disease

Congenital heart disease describes an inherited cardiac defect. About one percent of live newborns are affected by this most prevalent kind of congenital cardiac abnormality, which may disrupt the normal blood flow to the heart [8].

Cardiomyopathy

An assortment of heart conditions known as cardiomyopathies may be characterised by a variety of symptoms, including electrical or mechanical malfunction, inappropriate ventricular hypertrophy or dilatation, and hereditary predisposition [9].

Valvular Heart Disease (VHD)

These four heart valves—the mitral, aortic, tricuspid, and pulmonary—are all responsible for allowing blood to flow freely into the heart, but in people with valvular heart disease, these valves are congenitally defective. Diseases caused by degraded values include stenosis and regurgitation. In order to assess the patient's diagnosis, symptoms, and identification of VHD using auscultation, a clinical method is necessary. When it comes to diagnosing diseases, determining their prevalence, and predicting their outcomes, echocardiography is indispensable.

Atherosclerosis

The accumulation of plaque inside the walls of the arteries is known as atherosclerosis. A person's brain, heart, legs, arms, and pelvic arteries are all vulnerable to its effects.

Aortic Valve Sclerosis and Stenosis

Characteristics of aortic valve stenosis and sclerosis, which mostly affect the elderly, include calcification, stiffness, and an enlarged leaflet. Due to the similarities between atherosclerosis and aortic

stenosis, various biochemical and clinical variables associated with aortic stenosis have been discovered; these variables seem to align with the traditional risk factors for atherosclerosis. Myocardial infarction, heart failure, and cardiovascular mortality are much more likely in patients with aortic sclerosis, according to current research.

Risk Factors of Heart Disease

Risk factors are characteristics that indicate a higher likelihood of CVD [10]. Risk factors for cardiovascular disease and stroke may be changed or eliminated totally. They can reduce the likelihood of contracting diseases by altering risk factors. The following are examples of modifiable risk factors: cigarette smoking, hypertension, sedentary lifestyle, obesity, excessive alcohol intake, and poor dietary habits. Factors such as age and family history cannot be changed. In general, the likelihood of being sick increases as the number of risk factors increases. In order to maintain good heart health, it is essential to be aware of these risk factors[Figure 3]. Major contributors to the development of heart disease include:

Harmful Blood Cholesterol Levels

The human liver could generate sufficient for necessities of the body, other than more cholesterol is obtained frequently from eating foods [12]. If more cholesterol is made than the human body could utilize, the additional cholesterol is constructed in the arteries' walls containing those of the heart part and this is directed to the tapering of an arteries and blood flow can have reduced to the brain, heart, kidneys, and remaining parts.

Stress

Stress may also be a factor leading to heart disease. A risk of heart failure and heart attack may be affected by the effects of mental stress, behavioral behaviors and socioeconomic status. Unrelieved stress can even damage the arteries.

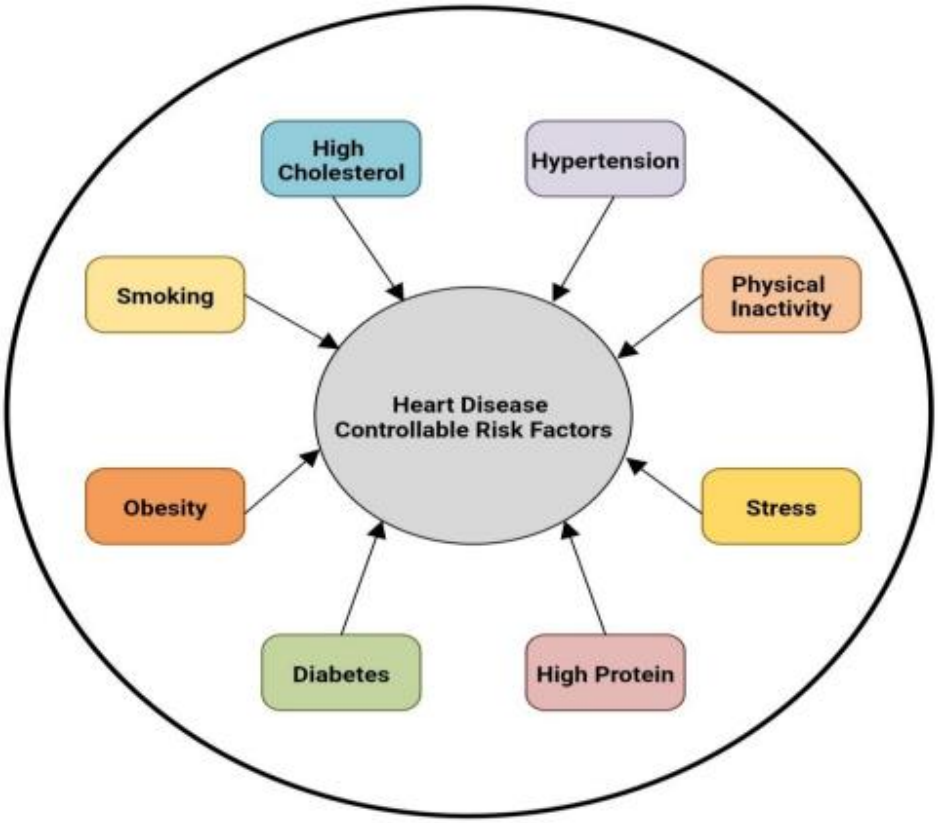

Figure 3. Risk Factor of Heart Disease [11].

Diabetes

Sugar may have happened by diabetes and it is built up in the blood. For adults, the death risks from heart issues with diabetes can be high compared to those who do not contain diabetes [13, 14].

Obesity

Obesity can be surplus fat in the human body. There is evidence linking this extra fat to triglyceride, bad cholesterol, and decreased good cholesterol levels. High blood pressure, diabetes, and heart issues have happened through heart disease [15].

Smoking

There are a lot of harmful health impacts associated with smoking since it introduces a number of pollutants into the body [16]. Injury to endothelial cells is one consequence of these impacts.

Physical Movement

Insufficient physical movement is directed to heart disease and other risk factors can also be happened such as diabetes, obesity, high blood pressure, and cholesterol. Regular physical movements will lessen the risk factor for a heart problem.

Signs and Symptoms of Heart Disease

Heart disease symptoms could be varied for women and men. For cases, men can be further probable to contain chest pain. Women might contain other signs and symptoms like chest uneasiness, shortness of breath, nausea, and severe exhaustion. Heart disease contains the following signs and symptoms [17]:

- Breathing Shortness
- Chest pressure, Chest ache, chest tension, and chest distress
- Ache in the chin, neckline, gullet, upper stomach
- Deadness, ache, flaw or chilliness in arms or arms when the blood vessels in those parts of the body have been tapering.

Need for Heart Disease Prediction

Around the world, the estimated death is around 17.5 million and that occurs due to cardiovascular disease. This cardiovascular disease mostly occurs in the middle-income and low-income people and it has resulted in 75 percentage of mortality rate. Moreover, 80 percent of the deaths happens due to cardiovascular disease because of heart attack and stroke. As per the report of WHO, India is one of the countries with increasing number of cardiovascular disease patients added in every year. It is mandatory to decrease the mortality rate that happens due to cardiovascular disease. Therefore, to overcome or to rectify these problems, researchers have been trying a lot to forecast a heart disease in an initial stage in order to give appropriate treatments to the patients in order to safeguard the human's life. The upgrading technology has been used by the researches to detect or predict the cardiovascular diseases. Data mining which is one among them provides a great benefits for the health care industry to authorize the health systems to utilize the data systematically and to identify the inefficiencies using analytics as well as it produces the best practices in order to upgrade the treatment and reducing the cost [18].

According to the researches, the usage of opportunity to upgrade the care and reducing cost constantly could be applied as much as 30 percentage of the entire medical spending. Data mining's success in other areas has increased its use across a variety of businesses and domains, including marketing, e-commerce, and retail.

OVERVIEW OF MACHINE LEARNING

The area of computer science known as machine learning (ML) makes use of AI to discover patterns or correlations in data automatically, without any previous definition of these things [19]. It is born out of the merging of two fields: computer science, which focuses on efficient algorithms for computing, and statistics, which aims to infer connections from data. Because it is extremely difficult, if not

impossible, for humans to extract meaningful information from the vast quantities of healthcare data stored and processed, machine learning is seeing increased use in this area to aid in disease diagnosis through data analysis, with the dual benefits of decreasing diagnostic time and increasing accuracy and efficiency. The main idea behind ML algorithms is that they construct models by discovering hidden patterns in the input information. Achieving that will allow them to make accurate predictions using datasets that are totally new to the algorithms. By using this route, computers may "learn" to get wiser and see patterns that a human being would have a very hard time spotting. ML methods and algorithms are able to process and forecast data from massive datasets [Figure 4].

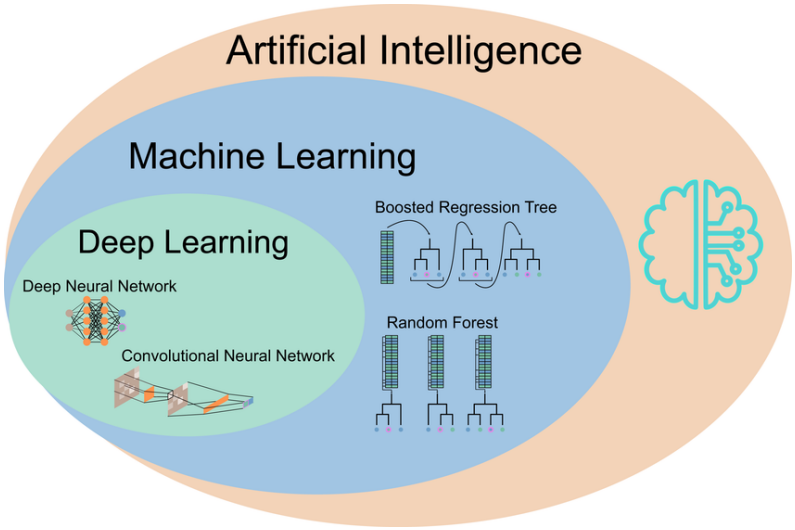

Figure 4. Machine learning.

The three main methods by which computers acquire new knowledge are reinforcement learning, unsupervised learning, and supervised learning.

Supervised Learning

Mathematical models are constructed by supervised learning algorithms using input-output data sets. Training data, sometimes called a collection of training examples, is typically comprised of a randomly chosen subset of the whole dataset. Training data typically consists of a supervisory signal, an output that is sought, and one or more inputs. The mathematical model is doing what it does by using a vector or array called a feature vector to represent each trainer dataset [20]. Information learned during training is often represented as a matrix. Through repeated optimization of an objective function, supervised learning algorithms retrieve knowledge from a model that contains functions that may be used to forecast outputs from fresh inputs. Making use of an optimum function, the method accurately predicts the outcome for inputs that were not included in the training data [Figure 5].

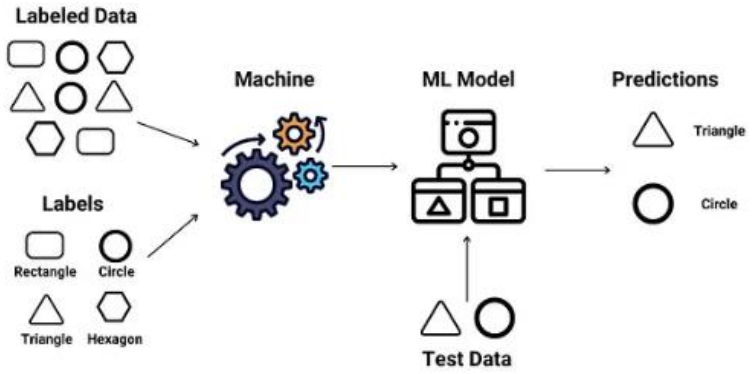

Figure 5. Supervised learning.

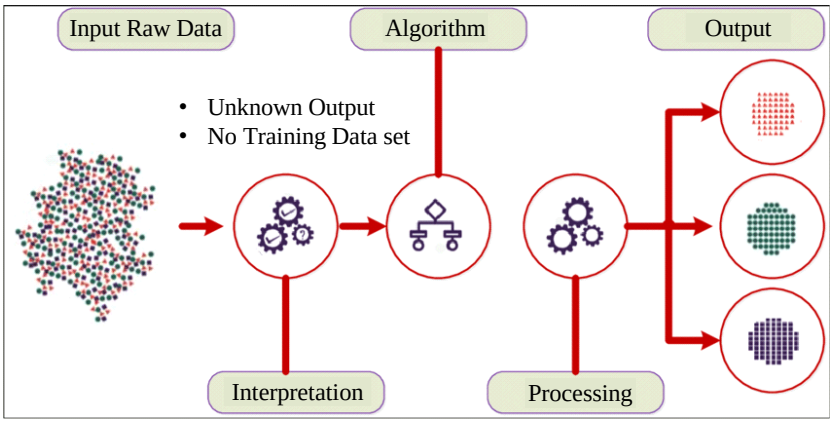

Figure 6. Unsupervised learning.

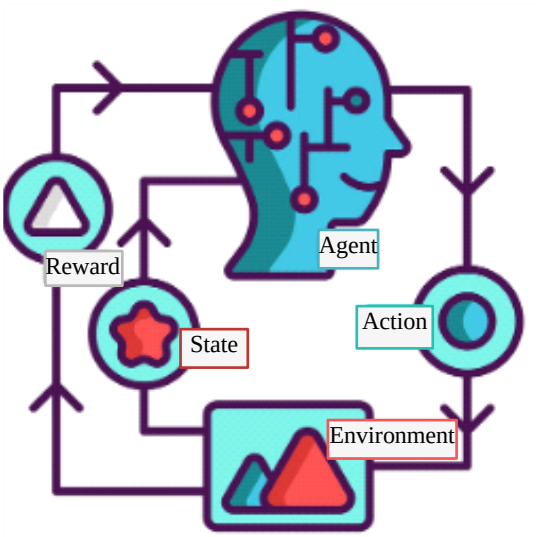

Figure 7. Reinforcement learning.

Unsupervised Machine Learning

An aim of unsupervised ML techniques is to uncover associations in a data structure without having a quantifiable output, making them especially effective in description jobs. Due to the absence of a response variable that may guide the study, this type of ML is known as unsupervised [21]. The purpose of unsupervised learning is to unearth hidden patterns and relationships within a dataset. While they will be focusing on supervised learning below, unsupervised learning encompasses several methods usually employed in psychological categorization and psychometric research, such as PCA, factor analysis, and mixture modeling. Therefore, unsupervised learning may be utilized to recognize latent or unobserved mental health dimensions and trajectories and to ascertain the optimal method for sub-categorizing dimensions (e.g., diagnostic groups) [Figure 6].

Reinforcement learning

In reinforcement learning, the goal is to teach a system to undertake actions that maximize the chance of positive outcomes, so that such events occur more often. Reinforcement learning (RL) is a method that is driven by the environment and may be used by software agents and robots to maximize performance in certain settings. Responding to environmental observations in a manner that maximizes favorable outcomes and minimizes negative ones is the purpose of RL. Utilizing AI models that have been trained, RL has the potential to automate and improve operations in a wide range of sectors, including supply manufacturing, autonomous driving, robots, and supply chain logistics [22] [Figure 7].

MACHINE LEARNING CLASSIFICATION TECHNIQUES FOR HEART FAILURE

The exponential growth of data has made machine learning a relatively new area of study. Machine learning makes short work of what is otherwise an insurmountable challenge: extracting actionable insights from massive datasets. The purpose of the research is to identify the health behavior's that lead to CVD and to rank the diagnostic tests in order of importance. Most notably, many performance indicators are used to evaluate various machine learning algorithms. The data used in this thesis is categorized by hand. The health benefits and risks of manual classification are debatable. Classification is a machine learning approach that forms the basis of 70% of the data used in this thesis, while 30% is used for testing purposes. The outcome of the predictions is then used to compare the various algorithms.

Naive Bayes

When it comes to predictive modelling, Naive Bayes is an unexpectedly strong algorithm. This statistical classifier aims to maximize the posterior probability in finding the class without assuming any relationship between characteristics. Although this classifier ought to have the lowest error rate in principle, this isn't necessarily the case in real-world situations. An absence of accessible probability data and the assumption of class conditional independence lead to inaccuracies [23].

Artificial Neural Networks

MLP and other ANN take their cues from the natural world and may mimic very complicated non-linear processes. ANNs are a powerful tool in the machine learning toolbox. "Neural" systems are those that mimic the way the human brain learns, as the name implies. As a general rule, neural networks have three layers: input, output, and hidden. Typically, the input is transformed to a pattern that the output layer may control by means of units in a hidden layer. ANNs are great at discovering patterns that would be difficult for a human programmer to extract and train to recognize because of their complexity or ambiguity. Since their inception in the 1940s, neural networks have grown in importance within the field of artificial intelligence. This is mostly because of a relatively recent method called "backpropagation" that allows networks to modify their hidden layers of neurons when the outcomes don't match the expectations of the developer [23] [Figure 8].

Support Vector Machine

SVM are equally adept at processing linear and non-linear datasets. In order to provide the training data more depth, a non-linear mapping method is used. An SVM uses a hyperplane—a kind of line—to split the space of input variables. The input variable space may be modified by applying the hyperplane so that points with classes of 0 and 1 are divided into two columns. The two-dimensional

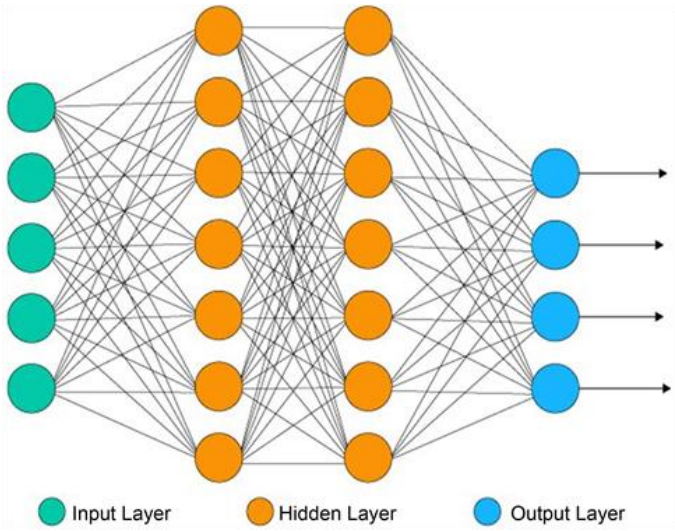

Figure 8. Diagram of an artificial neural network.

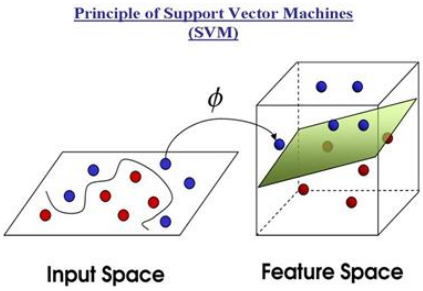

Figure 9. Principles of support vector machine operation.

equivalent of this would be a line, and it would be taken for granted that each input point would be entirely isolated by it. A data set's margin is the distance from its hyperplane to its neighboring locations. The optimal strategy for distinguishing between the two sets is to locate the hyperplane that has the largest margin. Support vectors describe or bolster the hyperplane, which is why these locations are so important. The optimal parameter values for maximizing the margin may be determined using an optimization algorithm in practice. You can see the feature transformation process in Figure 9 [23].

Simple Logistic Regression

LR is a ML approach that has its roots in statistics. Binary classification makes use of this strategy when a given value may be classified into two distinct groups. Logistic and linear regression have the same goal of finding the values of the input variables' coefficients. In this case, a logistic function—a non-linear function—is used to produce the output prediction, as opposed to linear regression. Any number between zero and one may be transformed using the logistic function. The likelihood of a data instance pertaining to class 0 or class 1 is determined by the predictions produced by logistic regression. When more justification for a forecast is required, this may be essential. Removing associated variables and those that are unrelated to the output variable improves the performance of logistic regression [23].

Decision Tree

A decision tree (DT) is an early supervised ML technique that, as its name suggests, aids in reaching a choice. Used mostly for classification issues, the DT-based ML model constructs a tree-like model with several nodes representing characteristics and branches indicating results of testing their related nodes. Traversing the classification tree from top to bottom allows for a thorough inspection of each node along the way, which ultimately results in the assignment of each sample to its expected class or label. The DT-based ML classifier has been tested in many research for the purpose of predicting the occurrence of cardiac diseases.

Random Forest

In the realm of ML, Random Forest is among the most well-known and potent algorithms. When it comes to machine learning, one kind of method is known as Bagging or Bootstrap Aggregation. One robust statistical method is the bootstrap, which may be used to estimate values from samples, such the mean. A better estimate of the true mean value may be obtained by taking several data samples, calculating their means, and then averaging them together. Using decision trees instead of calculating the mean of every data sample is the norm in bagging, but otherwise the procedure is the same. This part takes a large chunk of the training data and uses it to build models for each individual data point. They average the predictions from all the models to provide a clearer view of the real output value, which is necessary for all the data sets [23].

The steps for running the RF algorithm are as follows:

- Chooses K occurrences at random from the training set.
- Uses these k data points to construct a decision tree.
- Selects the Tree subset of the trees and proceeds with steps 1 and 2.
- Takes a majority vote into account when deciding the outcome or category.

k-Nearest Neighbour

The Nearest Neighbor algorithm is one of the simplest ML techniques. Remembering the training set and using it to forecast the label of each new occurrence based on its nearby neighbours' labels in the list is the main objective. This method is based on the assumption that the details used to define the domain points are relevant to their labelling in a way that makes it more likely that close points will have the same label. In addition to this, the process of computing the nearest neighbours can be performed very quickly under some conditions, even where the training set is very big.

Deep Learning

Deep learning (DL), now a relatively young but rapidly evolving branch of AI, has experienced a series of successes and made its way into real-world applications on a wide range of complex problems. DL has shown impressive progress in several fields such as image recognition, classification and segmentation, genomics, NLP, recognition of heart sounds and reconstruction of brain circuits. DL based on multilayer artificial neural networks is what constructs its core, which is to copy the way the human brain processes information. Deep learning (DL) employs data for exploring complex multi-tiered representations with different levels of abstraction. Revolutionary features of DL are now feasible due to the breakthroughs in graphics processing units and cloud computing, which makes it possible to run compute-intensive DL algorithms extensively. RNNs and CNNs are two well-known artificial neural network (DL) algorithms [24].

LITERATURE REVIEW

There are various different machine learning algorithms, among those, one algorithm is used to forecast a heart disease at an early stage. In this section, give a short review of recent related works.

This study determines the optimal collection of features for the chosen algorithm by analyzing performance across feature selections [25]. In this study, Decision Tree, KNN, and SVM were used for prediction, with the help of the stacking ensemble learning approach and K-fold validation. They finished data preparation and feature selection before they built the models. They have included metrics like accuracy, recall, and precision into their model to measure its performance. The dataset was best handled using Random Forest (RF) machine learning, which achieved an accuracy score of 99.02%.

In this study applying DL and ML Algorithms to a Kaggle Cardiovascular Disease Dataset[26]. Exercising lowers the risk of cardiovascular disease, however behaviors like smoking and drinking alcohol increase the likelihood of coronary artery disease. Their approach incorporates four models for CVD prediction: LR, NB, DL, and RF. The DL model outperformed the other 3 models with an accuracy of 73.78%.

In this work used survey data from more than 400k US citizens in 2020 to suggest six machine learning models [27]. A detailed comparison has been carried out between the 6 ML models-Adaboost, Xgboost, RF, DT, LR, and NB. They were able to increase their performance with a 91.57% accuracy rate in predicting heart illnesses using a LR model thanks to the heart disease forecasting model.

This study uses the K-Means clustering method to provide a prediction by classifying the UCI Heart Failure Clinical Records data into six categories [28]. In addition, a 93.33% success rate was attained while using SVM on this cluster data to predict HF in patients according to risk variables. The accuracy of the model's forecasts was improved by adjusting its hyperparameters using grid search.

In proposed methods improve the model's performance in both training and testing scenarios, leading to more precise diagnoses of heart failure [29]. The suggested method employs the gradient boosting approach, a kind of supervised learning, to identify cases of heart failure. For both model training and testing, the suggested diagnostic method employs the gradient boosting algorithm (GB). For the purpose of cardiac diagnostics, a gradient boosting classifier is drawn upon. Using the Cleveland Dataset, this study aimed to identify heart failure. The suggested approach outperforms competing techniques with an accuracy of 97.10 percent.

Table 1. Comparative analysis of related work for heart failure prediction analysis

Reference	Proposed work	Dataset	Key Findings/Results	Research Gap/limitations
Julian *et al.* [30]	algorithms for data mining, like RF, LR, DT, and NB.	UCI Cleveland dataset	90.16%	Build a web app using the RF algorithm to enhance a work in a future.
Qadri *et al.*, [31]	machine learning-based algorithms for comparison	heart failure dataset from the repository Kaggle	performance outcomes of 99% utilizing a CMSFL-Net	The medical community stands to benefit much from their planned study's substantial scientific contributions.
(Saboor *et al.*, [32]	ML classifiers, like XGB, CART, SVM, LDA, RF, MNB, LR, ET, and AB.	Cleveland heart disease dataset, Z-Alizadeh Sani dataset, Stat Log Heart, Hungarian, Long Beach VA, and Kaggle Framingham dataset.	90.16%	Detailed clinical data is lacking in these datasets, which means they do not include enough risk factors or features. Before this study, these restrictions were not well thought out.
(Kavitha *et al.*,[33]	RF, DT, hybrid of RF and DT	UCI HD Cleveland repository	88.7%	Lack of discussion on potential bias in the UCI HD Cleveland repository.
Bharti *et al.*[34]	ML algorithms and DL are applied	UCI Machine Learning Heart Disease dataset	94.2%	On rare occasions, the training model becomes overfit, leading to insufficient accuracy when tested on real-world data issues.
Atallah and Al-Mousa, [35]	Majority voting ensemble model	Medical test data	90%	Lack of discussion on potential sources of error in the medical test data.

Research gap

Several important research gaps and limitations are highlighted by the comparative study of comparable efforts for heart failure prediction. First, there is an immense demand for the creation of intuitive online apps that utilize complex algorithms such as Random Forest. This will make it simpler for healthcare providers and patients to access and utilize predictive models. Second, there is cause for worry over the completeness and representativeness of the characteristics taken into account during model training due to the lack of precise clinical data in several datasets utilized for research predicting heart failure [Table 1]. Thirdly, there needs to be a greater effort to analyze and overcome biases in data collecting and analysis because of the lack of focus on possible bias in datasets, particularly in repositories like the UCI HD Cleveland dataset. For ensuring the accuracy and practicality of predictive models in clinical practice, strong validation procedures and methods are crucial, especially in light of worries about overfitting and the transfer of models to real-world data. Finally, to improve the accuracy and credibility of clinical prediction models, it is crucial to identify and eliminate possible causes of mistake in medical test results, which have so far received little attention. Clinical decision-making and patient care may benefit greatly from more accurate, dependable, and broadly applicable prediction models for heart failure if these gaps and constraints were to be filled.

CONCLUSION

Deaths caused by cardiovascular disease have recently emerged as a major public health concern. Heart disease is caused by fatty plaques in the arteries. This review examines many recent research that addressed the prediction of heart disease using ML and data mining methods. The prediction of cardiac disease employs a variety of data mining and ML techniques. Experiments use varying datasets of people with heart disease. Participants in this survey gained knowledge about several ML methods for forecasting an occurrence of heart attacks.

Looking forward, they will keep looking for new methods that might be the best for forecasting heart failure. More ML classification algorithms and data pretreatment approaches may be used in subsequent works to provide outcomes.

REFERENCES

1. N. Bora, S. Gutta, and A. Hadaegh, "Using machine learning to Predict Heart Disease," *WSEAS Trans. Biol. Biomed.*, vol. 19, pp. 1–9, 2022, doi: 10.37394/23208.2022.19.1.
2. D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," *BMC Med. Inform. Decis. Mak.*, 2020, doi: 10.1186/s12911-020-1023-5.
3. H. Agrawal, J. Chandiwala, S. Agrawal, and Y. Goyal, "Heart Failure Prediction using Machine Learning with Exploratory Data Analysis," in *2021 International Conference on Intelligent Technologies, CONIT 2021*, 2021. doi: 10.1109/CONIT51480.2021.9498561.
4. Awan SE, Sohel F, Sanfilippo FM, Bennamoun M, Dwivedi G. Machine learning in heart failure: ready for prime time. Curr Opin Cardiol. 2018 Mar;33(2):190–195. doi: 10.1097/HCO.0000000000000491.
5. P. Probst, M. N. Wright, and A. L. Boulesteix, "Hyperparameters and tuning strategies for random forest," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 2019. doi: 10.1002/widm.1301.
6. V. Grgić, D. Mušić, and E. Babović, "Model for predicting heart failure using Random Forest and Logistic Regression algorithms," *IOP Conf. Ser. Mater. Sci. Eng.*, 2021, doi: 10.1088/1757-899x/1208/1/012039.
7. Taylan O, Alkabaa AS, Alqabbaa HS, Pamukçu E, Leiva V. Early Prediction in Classification of Cardiovascular Diseases with Machine Learning, Neuro-Fuzzy and Statistical Methods. Biology (Basel). 2023 Jan 11;12(1):117. doi: 10.3390/biology12010117.
8. Kaltman JR, Burns KM, Pearson GD, Goff DC, Evans F. Disparities in Congenital Heart Disease Mortality Based on Proximity to a Specialized Pediatric Cardiac Center. Circulation. 2020 Mar 24;141(12):1034-1036. doi: 10.1161/CIRCULATIONAHA.119.043392.
9. C. McCartan, R. Mason, S. R. Jayasinghe, and L. R. Griffiths, "Cardiomyopathy classification: Ongoing debate in the genomics era," *Biochemistry Research International*. 2012. doi: 10.1155/2012/796926.
10. R. L. McClelland *et al.*, "10-Year Coronary Heart Disease Risk Prediction Using Coronary Artery Calcium and Traditional Risk Factors Derivation in the MESA (Multi-Ethnic Study of Atherosclerosis) with Validation in the HNR (Heinz Nixdorf Recall) Study and the DHS (Dallas Heart Study)," J. Am. Coll. Cardiol., 2015 Oct 13;66(15):1643–53, doi: 10.1016/j.jacc.2015.08.035.
11. Ahmad, G. N., Fatima, H., Ullah, S., & Saidi, A. S. (2022). Efficient medical diagnosis of human heart diseases using machine learning techniques with and without GridSearchCV. IEEE Access, 10, 80151–80173.
12. Harrison SL, Buckley BJR, Rivera-Caravaca JM, Zhang J, Lip GYH. Cardiovascular risk factors, cardiovascular disease, and COVID-19: an umbrella review of systematic reviews. Eur Heart J Qual Care Clin Outcomes. 2021 Jul 21;7(4):330–339. doi: 10.1093/ehjqcco/qcab029.
13. Bays HE, Kulkarni A, German C, Satish P, Iluyomade A, Dudum R, Thakkar A, Rifai MA, Mehta A, Thobani A, Al-Saiegh Y, Nelson AJ, Sheth S, Toth PP. Ten things to know about ten cardiovascular disease risk factors-2022. Am J Prev Cardiol. 2022 Apr 6;10:100342. doi: 10.1016/j.ajpc.2022.100342.
14. G. D. Flora and M. K. Nayak, "A Brief Review of Cardiovascular Diseases, Associated Risk Factors and Current Treatment Regimes," *Curr. Pharm. Des.*, 2019;25(38):4063–4084doi: 10.2174/1381612825666190925163827.
15. G. A. Roth *et al.*, "Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study," *Journal of the American College of Cardiology*. 2020. doi: 10.1016/j.jacc.2020.11.010.
16. R. Grana, N. Benowitz, and S. A. Glantz, "E-cigarettes: A scientific review," *Circulation*. 2014.

13;129(19):1972–86.doi: 10.1161/CIRCULATIONAHA.114.007667.
17. S. Bösner *et al.*, "Accuracy of symptoms and signs for coronary heart disease assessed in primary care," Br. J. Gen. Pract., 2010, Jun;60(575):e246–57. doi: 10.3399/bjgp10X502137.
18. D. Deepika, "A novel heart disease prediction system using machine learning algorithms," 2022. https://shodhganga.inflibnet.ac.in/handle/10603/466923
19. P. Domingos, "A few useful things to know about machine learning," *Communications of the ACM*. 2012 vol 55Issue 10 PP 79-87. doi: 10.1145/2347736.2347755.
20. S. Aggarwal and Anmol, "Health Insurance Amount Prediction Using Supervised Learning," in *Proceedings of International Conference on Technological Advancements in Computational Sciences, ICTACS 2022*, 2022. doi: 10.1109/ICTACS56270.2022.9988256 https://ieeexplore.ieee.org/document/9988256.
21. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer..
22. J. Chen, X. Wu, L. Duan and S. Gao, "Domain Adversarial Reinforcement Learning for Partial Domain Adaptation," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 2, pp. 539–553, doi: 10.1109/TNNLS.2020.3028078.
23. S. Nashif, M. R. Raihan, M. R. Islam, and M. H. Imam, "Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System," *World J. Eng. Technol.*, Vol.6 Issue 4.Pg 854–873,2018, doi: 10.4236/wjet.2018.64057.
24. Xiaoyu Sun, Yuzhe Yin, Qiwei Yang and Tianqi Huo,Artifcial intelligence in cardiovascular diseases: diagnostic and therapeutic perspectives, European Journal of Medical Research (2023) 28:242.
25. A. A. Almazroi, E. A. Aldhahri, S. Bashir, and S. Ashfaq, "A Clinical Decision Support System for Heart Disease Prediction Using Deep Learning," *IEEE Access*, 2023, VOLUME 11, 2023: 61646–61659 doi: 10.1109/ACCESS.2023.3285247.
26. T. S. E. Reddy, S. R. Sripathi, D. Akula, S. Palaniswamy and S. R, "Cardiovascular Disease Prediction using Machine Learning and Deep Learning," *2022 6th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, Bangalore, India, 2022, pp. 1-5. doi: 10.1109/CSITSS57437.2022.10026391.
27. *M. Mamun, M. M. Uddin, V. Kumar Tiwari, A. M. Islam and A. U. Ferdous, "MLHeartDis:Can Machine Learning Techniques Enable to Predict Heart Diseases?,"* 2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), *New York, NY, NY, USA, 2022, pp. 0561–0565, doi:* 10.1109/UEMCON54665.2022.9965714.
28. S. Saravanan and K. Swaminathan, "Hybrid K-Means and Support Vector Machine to Predict Heart Failure," *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*, Trichy, India, 2021, pp. 1678–1683, doi: 10.1109/ICOSEC51865.2021.9591738.. doi:
29. K. L. Kumar and B. E. Reddy, "Heart Disease Detection System Using Gradient Boosting Technique," *2021 International Conference on Computing Sciences (ICCS)*, Phagwara, India, 2021, pp. 228–233, doi: 10.1109/ICCS54944.2021.00052.
30. A. Julian, R. Deepika, B. Geetha, and V. J. Sweety, "Heart disease prediction using machine learning," *Artif. Intell. Blockchain, Comput. Secur. Vol. 2*, vol. 2, no. 04, pp. 248–253, 2023, doi: 10.1201/9781032684994-38.
31. *A. M. Qadri, A. Raza, K. Munir and M. S. Almutairi, "Effective Feature Engineering Technique for Heart Disease Prediction With Machine Learning," in* IEEE Access, *vol. 11, pp. 56214-56224, 2023, doi: 10.1109/ACCESS.2023.3281484* doi: 10.1109/ACCESS.2023.3281484.
32. *A. Saboor, M. Usman, S. Ali, A. Samad, M. F. Abrar, and N. Ullah, "A Method for Improving Prediction of Human Heart Disease Using Machine Learning Algorithms," Mob. Inf. Syst., 2022, 22, Article ID 1410169, 9 pages* doi: 10.1155/2022/1410169.
33. M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart Disease Prediction using Hybrid machine Learning Model," in *Proceedings of the 6th In*ternational Conference on Inventive Computation Technologies, ICICT 2021, 2021., 2021, pp. 1329-1333 doi: 10.1109/ICICT50816.2021.9358597.

34. R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, “Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning,” *Comput. Intell. Neurosci.*, 2021, Volume 2021 | Article ID 8387680doi: 10.1155/2021/8387680.
35. R. Atallah and A. Al-Mousa, "Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method," *2019 2nd International Conference on new Trends in Computing Sciences (ICTCS)*, Amman, Jordan, 2019, pp. 1-6, doi: 10.1109/ICTCS.2019.8923053.

