Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
March 2024 5(3), 837–845

PREDICTION OF CARDIOVASCULAR MALADIES THROUGH THE IMPLEMENTATION OF MACHINE LEARNING AND DEEP LEARNING ALGORITHMS

Dr. P. Ramya [1], M. Gayathri [2], S. Lunashree [2], K. Harshiga [2]

[1] Associate Professor, Department of Computer Science and Engineering, Mahendra Engineering College, Tamilnadu, India
[2] Ug Scholar, Department of Computer Science and Engineering, Mahendra Engineering College, Tamilnadu, India

Corresponding Author
Dr. P. Ramya, ramyapb@mahendra.info

DOI
10.29121/shodhkosh.v5.i3.2024.3336

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT
The heart stands as a cornerstone within the human body's intricate network of blood circulation. Recognized as a vital organ, its pivotal role in sustaining life cannot be overstated. Heart disease, a formidable health challenge, demands serious consideration due to its significant impact on individual well-being and public health condition with a significant chance of mortality or serious long-term effects. Nevertheless, there are no effective methods for finding hidden patterns and linkages in e-health data. In order to save lives, medical diagnosis is a challenging but essential task that must be completed promptly and accurately. Clinical testing is costly, so in order to save expenses, we necessitate a precise computer-based automated decision support system that is apt for our needs. It has been suggested that machine learning applied to health analytics will enable accurate analysis of patient data. The medical field does not engage in data mining. When data mining techniques are used to patient risk factor data sets, an intelligent model can be created in the medical area. Recent advancements in data utilization have significantly impacted the field of knowledge discovery in databases (KDD), particularly in the realm of disease diagnosis. This study investigates the application of deep learning and machine learning methodologies in this context. With the emergence of numerous data mining classifiers, there is a growing focus on enhancing the accuracy and efficacy of disease diagnosis. This paper presents a novel approach—a heart attack prediction system—that leverages deep learning techniques, specifically the Multi-Layer Perceptron (MLP). MLP stands out as a sophisticated classification method, harnessing the power of artificial neural networks with deep learning capabilities. To ensure robust and dependable outcomes, the proposed methodology integrates data mining principles with deep learning techniques.

Keywords: Cardio Vascular Disease, Data Analytics, Deep Learning, Machine Learning Algorithms, Predictive System

1. INTRODUCTION
Cardiovascular disease stands out as one of the most prevalent illnesses worldwide today, significantly impacting global health. Shockingly, studies indicate that it claims the lives of over 17.7 million individuals annually. Among these fatalities, estimates attribute 7.4 million to coronary heart disease and 6.7 million to stroke. What makes this ailment particularly alarming is its ability to strike unexpectedly, often without prior detection by doctors. The challenge is compounded by the difficulty in identifying silent heart attacks, further underscoring the urgent need for accurate prediction systems. Given the scarcity of specialists and the growing instances of misdiagnosis, there's a pressing demand for reliable tools in cardiovascular disease prediction. In response to this imperative, research endeavours have been directed towards the development of cutting-edge methods in medical data mining and machine learning. This study's primary objective is to unveil the most pertinent characteristics for silent heart attack detection. By employing classification algorithms, the aim is to discern crucial patterns and features within medical data that can aid in accurate

prediction and diagnosis. An artificial neural network will be used to produce even more accurate findings. Although the idea for a system like this is not new, the approaches used today are ineffective and do not try to detect the risk of a silent heart attack. This study aims to address these problems and recommend special components that ought to be included in order to build a more complete system. The information produced by the available heart attack prediction techniques is insufficiently precise. The accuracy is being pushed to a certain degree by the machine learning techniques being used, as seen in the literature review. The present heart attack prediction algorithm also has a difficulty with the use of characteristics. Since the traits that should be used to predict heart attacks are common, the studies typically yield inaccurate results. The study endeavours to extract pertinent properties from datasets with the aim of enhancing prediction accuracy. Furthermore, it will correctly identify the problem for clients, making it easy for them to comprehend. The study sets itself apart by fusing data mining and deep learning techniques. The system that the article offers has a robust prediction algorithm, an extensive module for creating reports, and an effective module for classifying data. According to this study, a self-learning process should be created wherein past inputs of sickness outcomes affect a user's unique future risk of heart disease. To ensure error-free classification and prediction of the dataset, the suggested method mainly depends on preprocessing techniques. The accuracy of the prediction will be raised by using a large number of training sets. The basic flow of the study is shown in fig 1.

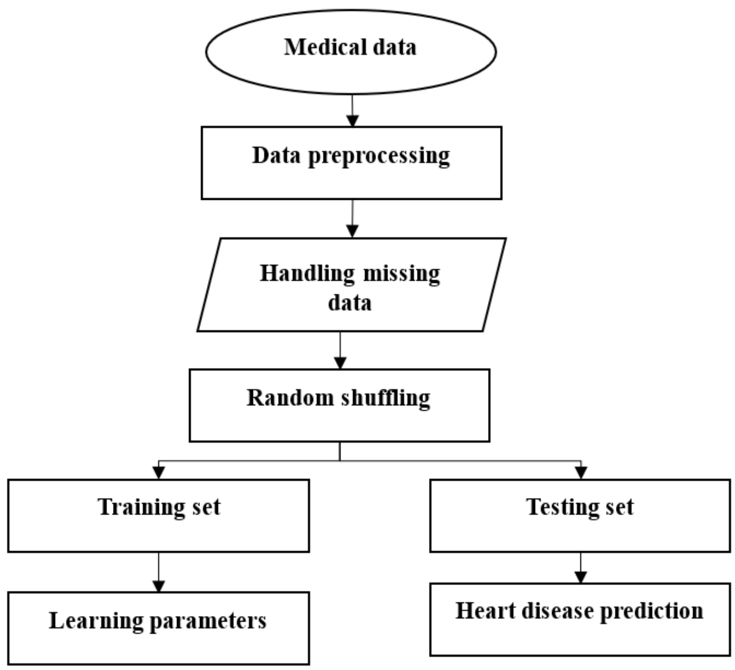

Figure 1 Flow of Heart Disease prediction

2. RELATED WORK

Aleeza Nouman, et.al,[1] developed a framework to forecast cardiac disease early on by utilising new and improved technology in a safe and effective way that helps save lives. It is challenging to safely and precisely predict cardiac disease, but cutting-edge technologies like blockchain and machine learning (ML) are transforming the current healthcare system. Blockchain and machine learning are the most effective ways to collect data while making heart disease predictions. This article provides comprehensive evaluations of various machine learning (ML) methods to discover the malady, for both hospitals and practitioners, accurate diagnosis and forecasting are crucial. When forecasting cardiac disease, several crucial factors should be considered. Thanks to advances in computing technology, a multitude of ways for more info collection and spaces are now easy to have. This results in the massive gathering of health-related data, which is excellent for clinical research.

Abhay Agrahara, et.al,[2] Describe the most recent findings on the prediction of heart disease, including comparative analysis and analytical conclusions. The study demonstrates how integrating Naive Bayes with Genetic algorithm, Decision Trees, and Artificial Neural Networks can bolster the accuracy of heart disease prediction systems across diverse scenarios. It provides a comprehensive overview of prevalent data mining and machine learning techniques, elucidating their complexities.

Furthermore, consider the application of federated learning in addressing limitations posed by small-scale medical research datasets and mitigating privacy concerns in heart disease prediction. Federated learning facilitates collaboration among healthcare institutions by allowing local model training on respective datasets while preserving data privacy. Techniques such as differential privacy further enhance privacy preservation by introducing noise during the training process. Through these methods, researchers can advance prediction accuracy while ensuring patient data confidentiality in medical research. While very sophisticated these methods, like neural network models, can be used to increase prediction accuracy, they sometimes lack the very large data sets needed for clinical or medical research. We investigate the modelling of heart disease prediction using surrogate data sets consisting of synthetic observations. We create synthetic data utilising the original observations as well as the surrogate data, depending on the features of the original observations, and we evaluate the prediction accuracy outcomes obtained from conventional machine learning models.

Ibrahim M. El-Hasnony, et.al,[3] employed quick learning methodologies to evaluate the cardio malady dataset. Numerous machine learning experiments have used this dataset to categorize and forecast heart diseases. This research aims to address the challenge of model memorization through a proposed methodology. Active learning emerges as a promising strategy, as it avoids overfitting specific data instances. Emphasizing the importance of training on samples that significantly impact system performance, the focus shifts from memorization to developing a model capable of generalizing existing data. The phenomenon of models performing well on training data but poorly on new data, often observed in complex environments, underscores the need for effective model selection strategies. To this end, five distinct selection strategies QUIRE, AUDI, MMC, Random, and Adaptive are applied to multiple-label active learning. Predictive modelling employing a grid search with a label ranking classifier is conducted for each strategy. The stopping criterion for the active learning procedure typically involves a predetermined number of iterations, with the performance of the base classifier evaluated using a test set and assessment metric. The efficacy of the entire model is assessed using a dataset on heart disease, leading to the conclusion that extrapolation could potentially enrich the learning model with additional data.

Md. Mahbubur Rahman, et.al,[4] developed an automatic prediction system that is both web-based and manual, and it can provide a conceptual report with a clear warning about a patient's cardiac status. Using a few health metrics, the suggested prediction system makes a heart disease prediction. Thirteen health parameters are used by the system, including blood pressure, ECG, kind of chest discomfort, age, and sex. To accurately identify cardiac illness, eight different algorithms are used: Ada Boost, KNN, XG Boost, and Random Forest (RF), Decision tree (DT), Naïve Bayes, Support Vector Machine (SVM), and Logistic Regression (LR). Decision Tree and Random Forest are the two methods that outperform the others. This research also led to the creation of a website that makes it simple to check one's heart health in real time from home. The recommended method identifies heart disease more quickly and affordably. In this paper, machine learning algorithms were trained and tested on pre-processed data. In the initial stage, the pre-processed data undergo division into two distinct categories. Eighty percent of these data are allocated for the training phase, while the remaining twenty percent are reserved for testing purposes. Throughout both training and testing, the proposed system utilizes various machine learning techniques, including Decision Trees, XG Boost, KNN (K-Nearest Neighbours), Support Vector Machine, Naïve Bayes, Logistic Regression, AdaBoost, and Random Forest, to train our dataset.

Raniya R. Sarra, et.al,[5] A superior model changed into devised to relieve computational burdens and refine the accuracy of coronary heart disease analysis and diagnosis. Leveraging a type model built upon the Support Vector Machine (SVM) set of rules, efforts have been directed in the direction of improving coronary heart ailment diagnosis. Two widely identified heart disorder datasets had been hired for validation purposes. Results tested a awesome increase in accuracy rates for each the Cleveland and Stat log datasets, growing from 84.21% to 89. Forty-seven% and 85.29% to 89.7%, respectively. Moreover, the device's features have been decreased from 14 to 6, ensuing in a sizable decrease in computational load from one hundred% to about forty-two%. This study endeavours to make a contribution to the advancement and utilization of structures dedicated to coronary heart disorder analysis and diagnosis inside the destiny. Recognizing the imperative of accurate cardiac ailment prediction as a critical global fitness difficulty, the observe proposes a singular heart disorder type model centred at the Support Vector Machine (SVM) approach. Employing statistical most appropriate feature choice methods to enhance prediction accuracy, the efficacy of the proposed version was demonstrated through complete comparisons with traditional models across numerous overall performance metrics.

3. EXISTING SYSTEMS

Many attempts have been made to use machine learning for jobs linked to classification, like helping to detect cervical cancer early on. Among the most popular machine learning approaches include support vector machines, k-nearest neighbours, Random Forest trees (RFT), Classification and regression trees (CART), and others. Fig 2 shows the multiple models.

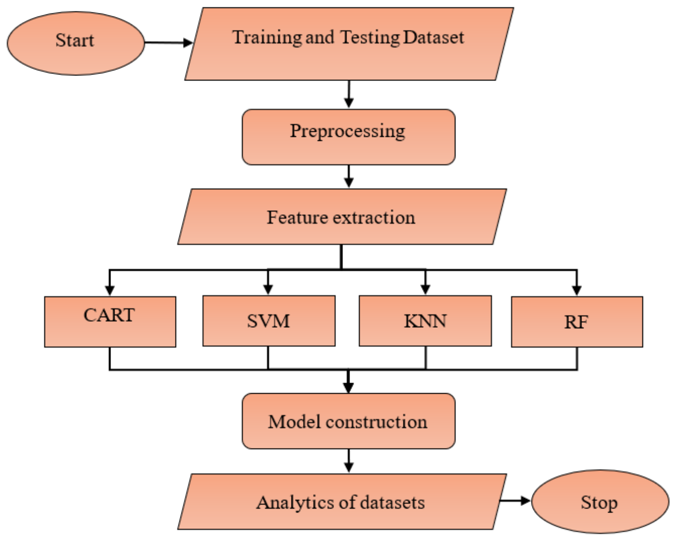

Figure 2 Machine learning model

CLASSIFICATION AND REGRESSION TREE: CART (Classification and Regression Trees) The machine learning method technique illustrates how target variables can be predicted based on other variables. At the end of each node in the decision tree, the division occurs, designating one part as a predictor variable and the other as a predictor for the target variable. Subsequently, at each node in the decision tree, sub-nodes are generated based on the threshold value of the attribute. The training set comprises root nodes, which are further divided into two groups according to the optimal feature and threshold values. This process continues iteratively, applying the same approach to segregate the subgroups until the tree either reaches its maximal potential by yielding all possible leaves or achieves its ultimate, refined subset.

SUPPORT VECTOR MACHINE (SVM): Support vector machines address three main problem cases: regression, classifying data into one or more classes, and identifying outliers The main applications of SVM include predicting hyperplanes and quadratic solving equations. This process was initially carried out using well-established quadratic optimization methods.

RANDOM FOREST: Random Forest (RF) Classification, a form of data analysis, involves constructing models from data classes. These models predict the existence of classes through a categorization model, while continuous-valued functions are represented through numerical prediction. An algorithm combining numerical prediction and classification is employed to refine accuracy estimates. Random forests, an ensemble learning method, are utilized for regression and classification tasks. During the training phase, multiple decision trees are built, and the resultant class represents either the mean of the classes (classification) or the mean/average prediction (regression). A bootstrap sample is generated, yielding numerous categorization trees, each treated with techniques based on the training set of data. The model is then trained using random forest classification on the training set. Estimating the test set's outcomes and uses a confusion matrix to gauge accuracy.

K-NEAREST NEIGHBOR ALGORITHM: Both regression and classification utilize the non-parametric classifier known as K-Nearest Neighbour (KNN). As the data distribution is parametric, no assumptions can be made about it. Unlike other classifiers, KNN does not require an explicit training phase. Instead, the data is split into training and test sets for KNN classification. To classify a row in the test set, the algorithm observes the k nearest neighbour points based on their Euclidean distance from the corresponding point in the training set.

4. PROPOSED METHODOLOGIES

Heart disease continues to claim the lives of a large number of people around the world. Heart disease or cardiovascular disease is the most dangerous disease affecting developed countries. In addition to causing sudden death for a significant portion of the population, cardiovascular diseases also leave a large proportion of victims with extreme suffering and permanent disability. Although many cardiovascular diseases are preventable, their incidence is increasing due to inadequate prevention strategies. Diagnosing heart disease has become more difficult in the medical field.

A complete and accurate review of clinical test results and the patient's medical history is the basis for this diagnosis. The main goal of advances in machine learning is to develop intelligent automated systems that help doctors diagnose patients and make decisions. An automated medical diagnostic system like this can significantly increase the chances of saving lives by improving timely medical treatment and appropriate follow-up care. These intelligent systems use triage algorithms to provide accurate diagnoses. Popular classification techniques include neural networks. In this work, a multilayer perceptron neural network with backpropagation was used as the training algorithm. This study suggests a more accurate diagnostic method to predict heart disease. The propagation process is repeated until the lowest error rate is achieved. The data in the previous section clearly shows that the accuracy rate is maximized. To solve the problem, the brain connects large groups of biological neurons through axons; A neural network is a computer system based on a large number of neural units. Each brain cell has many connections with other brain cells. The connections have the ability to support or weaken the functional state of other connected brain units. The input values of each brain unit can be added using the sum function. Before transmitting to neighbouring neurons, each link and the unit itself must pass a threshold function, also known as a limit function. Because these systems are taught rather than explicitly developed, they work best in situations where the solution or feature recognition is difficult to explain in a conventional computer program. One neuron growing connections with another neuron creates a neural network, which can have one or more layers. A multilayer artificial neural network consists of an input layer, an output layer, and one or more hidden layers. The hidden layer can be used to perform intermediate calculations before passing the input to the output layer. Once a model is created for a specific application, it is trained using inputs and targets until it masters the art of associating certain inputs with specific outputs. The network is trained until it reaches the lowest weight change value during the training cycle. After training, a model is validated by seeing whether it produces consistent results. Multilayer networks can retain information because they have access to many different synaptic weights. A neuron in the output layer can sometimes determine its output using a threshold function. Even though artificial neurons are relevant to our situation, we still call them neurons. Synapses between neurons are represented by connections, which are edges of an oriented network with nodes representing artificial neurons. The model building process has five consecutive steps:

1) Select the data that will be used as input and output according to the supervised learning method.
2) Standardize both input and output data.
3) The normalized data is trained using a neural network.
4) Confirm model fit.
5) Assess the gap between expected results and expected results.

Layered feedback neural networks consist of layers or subgroups of processing modules. The data is processed by a processing component layer, which then sends the results to a higher layer. The next layer can then perform independent calculations and pass the results to a lower layer. Ultimately, a subset of one or more processing units determines the output of the network. Each processing unit uses a weighted average of the input to perform calculations. The input layer appears first and the output layer appears last. The layers between these two are hidden layers. Because processing elements are thought to function similarly to neurons in the brain, they are called artificial cells, neuromas, or neurons. The steps of the neural network algorithm are as follows:

Step 1: Randomly initialize the weights and biases.

Step 2 involves providing the machine with training samples.

Step 3: Compute the net input and output of each unit in the hidden layer and output layer after further using the input.

Step 4: Invert the defect on the hidden layer in the fourth step. Update weights and biases to account for error propagation.

Step 5: Training and learning functions are mathematical procedures that automatically change the weights and biases of the network.

Step 6: The neural network method outperforms conventional machine learning methods based on these steps.

The proposed framework is shown in fig 3.

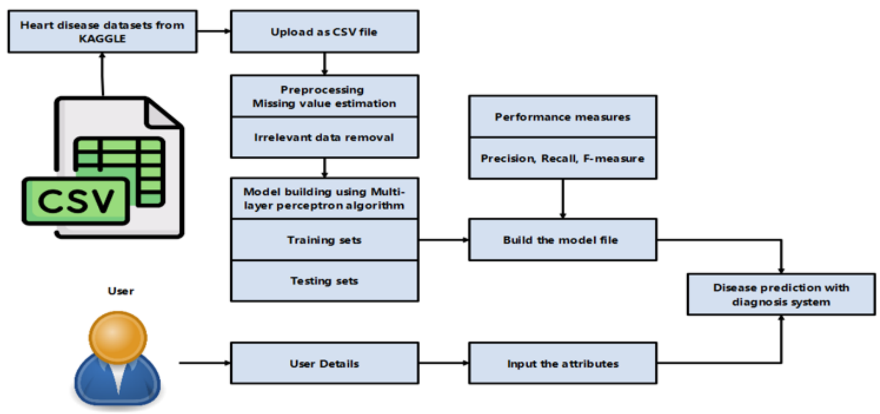

Figure 3 Proposed architecture

5. EXPERIMENTAL RESULTS

This system forecasts heart illness using a back end (MYSQL) and a front end (Python framework) deep learning method. For this study, we can upload data in the form of a CSV file with a various item. The simple aspect description is:

Table 2 Attribute Details

Attribute	Description
Id	Number
Age	In days
Gender	1 = Women 2 = Men
Height	cm
Weight	Kg
Ap_hi	Systolic BP
Ap_lo	Diastolic BP
Cholesterol	1: Normal 2: Above normal 3: Well above normal
Glucose	1: Average 2: Above average 3: Well above average
Smoke	Yes or No
Alco	Binary features
Active	Binary features
Cardio	Target variable

In last stage, use a trained system to predict various cardio problems like arrhythmia, hypertension, cardiac arrest, and coronary heart malady. A new user's details can be submitted as testing data, and categorization can be done to more accurately predict and diagnose different heart conditions. We are able to submit the datasets for 200 patients and obtain the samples from the UCI repository. And classifying the illnesses according to eleven criteria, including cardiac datasets. The dataset contains variables such as age, sex, height, weight, systolic and diastolic blood pressure, fat, sugar contain, drinking habits, smoking, and activity level.

We are able to calculate the execution of one and all method also contrast them based on the accuracy parameter. The system's performance can be examined using the F-measure parameter. The system's execution is calculated using F-measure, Precision, and Recall.

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{F measure} = 2* \frac{Precision*Recall}{Precision+Recall}$$

The performance evaluation result is shown in fig 4.

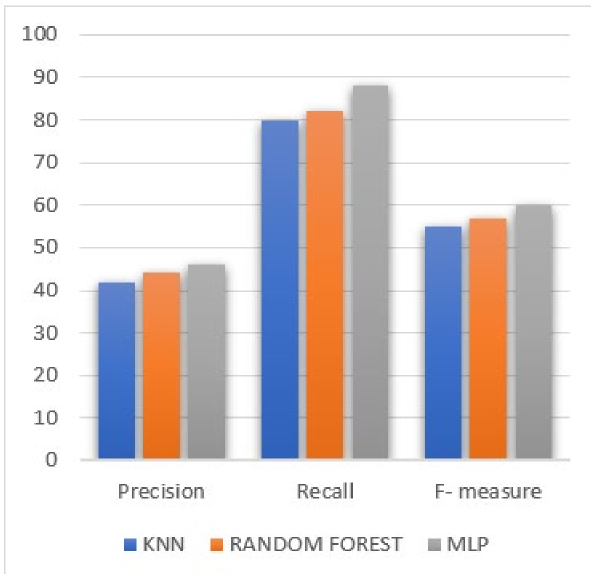

Figure 4 Performance Chart

The suggested neural network technique, like in Figure 4, produces higher level F-measure values more than the Random Forest also KNN algorithms that are currently in use.

6. CONCLUSION

This study looks at how challenging it is to limit and condense various data mining methodologies can be handy for medical detections. Many algorithms and target criteria combinations are integrated to effectively and intelligently forecast heart disease via data mining. Data mining can be used to extract meaningful medical rules from medical data, which is necessary for clinical diagnosis and sickness prognosis. Classification plays a crucial role in identifying the malady status. To establish classification, a sizable sample of hospitalized patients was used in the current study. The classification method is highly sensitive to noise in the data. Noisy data poses significant obstacles to the capability of categorization processing. It not only makes the procedure take longer, but it also makes the categorization system perform worse. Consequently, all attributes that could later serve as inconsistent variables need to be eliminated from datasets prior to applying a classification technique. Using preprocessing techniques and classification rule algorithms such as the MLP, we are able to categorize datasets that users have provided for this research project. The results of the tests show that the MLP technology yields superior results than other approaches. It is likely that in the future, data mining algorithms and approaches will be employed to improve performance efficiency.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Nouman, Aleeza, and Salman Muneer. "A systematic literature review on heart disease prediction using blockchain and machine learning techniques." International Journal of Computational and Innovative Sciences 1.4 (2022): 1-6.

Thilagavathi, G., et al. "heart disease prediction using machine learning algorithms." 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC). IEEE, 2022.

El-Hasnony, Ibrahim M., et al. "Multi-label active learning-based machine learning model for heart disease prediction." Sensors 22.3 (2022): 1184.

Rahman, Md Mahbubur. "A web-based heart disease prediction system using machine learning algorithms." Network Biology 12.2 (2022): 64.

Sarra, Raniya R., et al. "Enhanced heart disease prediction based on machine learning and χ2 statistical optimal feature selection model." Designs 6.5 (2022): 87.

Saboor, Abdul, et al. "A method for improving prediction of human heart disease using machine learning algorithms." Mobile Information Systems 2022 (2022).

Javeed, Ashir, et al. "Machine learning-based automated diagnostic systems developed for heart failure prediction using different types of data modalities: A systematic review and future directions." Computational and Mathematical Methods in Medicine 2022 (2022).

Gupta, Chiradeep, et al. "Cardiac Disease Prediction using Supervised Machine Learning Techniques." Journal of Physics: Conference Series. Vol. 2161. No. 1. IOP Publishing, 2022.

Ramesh, T. R., et al. "Predictive analysis of heart diseases with machine learning approaches." Malaysian Journal of Computer Science (2022): 132-148.

Suresh, Tamilarasi, et al. "A hybrid approach to medical decision-making: diagnosis of heart disease with machine-learning model." Int J Elec Comp Eng 12.2 (2022): 1831-1838.

[11] Nadakinamani, Rajkumar Gangappa, et al. "Clinical data analysis for prediction of cardiovascular disease using machine learning techniques." Computational intelligence and neuroscience 2022 (2022).

K. Shah, K. Sharma, and D. Saxena, ''Editorial: Health technology assessment in cardiovascular diseases,'' Frontiers Cardiovascular Med., vol. 10, Jan. 2023, Art. no. 1108503

J. A. W. Gold, F. B. Ahmad, J. A. Cisewski, L. M. Rossen, A. J. Montero, K. Benedict, B. R. Jackson, and M. Toda, ''Increased deaths from fungal infections during the coronavirus disease 2019 pandemic—National vital statistics system, United States, January 2020–December 2021,'' Clin. Infectious Diseases, vol. 76, no. 3, pp. e255–e262, Feb. 2023.

Z. Du, Y. Yang, J. Zheng, Q. Li, D. Lin, Y. Li, J. Fan, W. Cheng, X.-H. Chen, and Y. Cai, ''Accurate prediction of coronary heart disease for patients with hypertension from electronic health records with big data and Machine learning methods: Model development and performance evaluation,'' JMIR Med. Unformat., vol. 8, no. 7, Jul. 2020, Art. no. e17257.

J. Amin, M. Sharif, M. Yasmin, T. Saba, and M. Raza, ''Use of machine intelligence to conduct analysis of human brain data for detection of abnormalities in its cognitive functions,'' Multimedia Tools Appl., vol. 79, nos. 15–16, pp. 10955–10973, Apr. 2020.

A. Srivastava and A. k. Singh, ‘‘heart disease prediction using machine learning,’’ in Proc. 2nd Int. Conf. Advance Comput. Innov. Technol. Eng. (ICACITE), Apr. 2022, pp. 2633–2635.

M. Pal and S. Parija, ‘‘Prediction of heart diseases using random forest,’’ J. Phys., Conf. Ser., vol. 1817, no. 1, Mar. 2021, Art. no. 012009.

H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, ‘‘heart disease prediction using machine learning algorithms,’’ in Proc. IOP Conf. Ser., Mater. Sci. Eng., vol. 1022, 2021, Art. no. 012072.

M. Pal, S. Parija, G. Panda, K. Dhama, and R. K. Mohapatra, ‘‘Risk prediction of cardiovascular disease using machine learning classifiers,’’ Open Med., vol. 17, no. 1, pp. 1100–1113, Jun. 2022.

A. Garg, B. Sharma, and R. Khan, ‘‘heart disease prediction using machine learning techniques,’’ in Proc. IOP Conf. Ser., Mater. Sci. Eng., vol. 1022, 2021, Art. no. 012046.