https://doi.org/10.48047/AFJBS.6.10.2024.6456-6462

AFJBS African Journal of Biological Sciences
Journal homepage: http://www.afjbs.com

Research Paper Open Access

ADVANCING CVD: HARNESSING ML AND DL FOR DISEASE PREDICTION

S. Anthony Mariya Kumari[1]

[1]Research Scholar, Department of Computer Applications, Dr. M.G.R Educational and Research Institute,Chennai-65, Tamil Nadu, India. selvarajanthony94@gmail.com

Viji Vinod[2]

[2]Professor & Head, Department of Computer Applications, Dr. M.G.R Educational & Research Institute, Chennai-95, Tamil Nadu, India. vijivino@gmail.com , hod-mca@drmgrdu.ac.in

Revatthy Krishnamurthy[3]

[3]Professor, Department of Artificial Intelligence and Machine Learning, Rajalakshmi Engineering College,Chennai-105, Tamil Nadu, India, revatthy11@gmail.com

Article Info
Volume 6, Issue 10, 2024
Received: 20 May 2024
Accepted: 15 June 2024
Doi:10.48047/AFJBS.6.10.2024.6456-6462

Abstract

This research investigates the use of ML and deep DL algorithms to predict heart disease, addressing the important need for reliable early identification in order to lessen its huge global health effect. The study uses preprocessing approaches to manage data completeness and normalization on a variety of datasets, including demographic information, clinical records, and diagnostic test results. These datasets are subjected to algorithms such as logistic regression, decision trees, ensemble approaches, CNNs, and RNNs, with the goal of identifying patterns that indicate heart disease risk. Results show significant gains in predicting accuracy compared to traditional approaches, with ML models reaching area under the curve scores more than 0.85 and DL models outperforming with AUC ratings greater than 0.90. This research underscores the transformative potential of ML and DL in advancing early diagnosis and personalized treatment strategies for heart disease, thereby enhancing healthcare outcomes on a broader scale.

Keywords: *Machine Learning, Deep Learning, Logistic Regression, Decision Trees, CNN, RNN, Cardiovascular disease.*

Introduction

According to the Indian Health Association, 50% of all heart attacks in India occur in men under the age of 50. In India, more than 200000 children are predicted to be born with congenital cardiac disease annually. According to an article from the Cleveland Clinic, symptoms of heart disease include as Persistent chest pain, Difficulty breathing, sudden sweating, and increased heart palpitations. Over the past four years, cases of heart failure in India have fluctuated between 25,000 and 28,000. Cardiovascular disease (CVD) accounts for the highest percentage (38.6%) of deaths in urban Chennai. Risk factors for cardiovascular disease include as Aging, Obesity, High blood sugar or diabetes, High cholesterol levels, Exposure to air pollution, Excessive use of steroid medications. Every individual has normal plus rate and blood pressure is 60 to 100 beats per minutes and high plus rate and blood pressure can go from 100 to 185 beats per minutes.

Using machine learning to predict cardiovascular illness with UCI and Kaggle datasets has various benefits. These databases contain a wide range of patient characteristics, including age, gender, blood pressure, cholesterol levels, and lifestyle behaviors, all of which are important for accurate prediction. Machine learning algorithms can successfully evaluate complicated and diverse data, revealing patterns and connections that traditional statistical approaches may overlook. Using these new methodologies, we may create prediction models that are more accurate, fast, and tailored in risk assessments.

This can dramatically improve early detection and intervention, resulting in better patient outcomes and lower healthcare expenditures. Furthermore, employing well-known datasets like as those from UCI and Kaggle guarantees that the models are trained on high-quality, benchmarked data, which improves their dependability and generalizability across diverse populations. As a result, using machine learning into cardiovascular disease prediction leverages the potential of big data and sophisticated analytics to more effectively address one of the world's top causes of death.

Researchers have explored various computational techniques to enhance prediction accuracy and early detection. ML algorithms, as highlighted in studies by Sharma and Parmar (2020) and Jindal et al. (2021), offer robust tools for analyzing complex datasets derived from medical records and diagnostic tests. These algorithms enable healthcare professionals to identify patterns and risk factors associated with cardiovascular diseases efficiently. Concurrently, deep learning (DL) techniques, as evidenced by Pasha et al. (2020) and Ali et al. (2020), have revolutionized predictive modeling by automatically learning intricate features from raw data, such as imaging and genetic profiles. By integrating these advanced methodologies, researchers aim to develop more accurate and personalized prediction models, ultimately aiding in early intervention and targeted treatment strategies.

Related Works

This paper explores the application of deep learning in medical science, particularly its capability to predict cardiac disease. It proposes a model optimized with Talos, demonstrating superior performance compared to traditional classification techniques such as KNN, SVM, RF, and NB. The effectiveness of this approach is demonstrated using the Heart Disease UCI dataset [1]. The study introduces a heart disease prediction system leveraging ML algorithms like logistic regression and KNN, aiming for accuracy and reduces healthcare expenses. The provided .pynb file offers valuable insights for heart disease prediction [2]. Heart disease, a major cause of 28.1% of fatalities in India and globally, is a critical issue. Researchers use various datasets, including the Kaggle dataset, to predict heart disease using ML algorithms [3]. This study examines the Kaggle dataset, which includes age- related factors. The online application uses Cleveland Foundation medical studies and Kaggle data to predict heart disease prevalence, utilizing data mining techniques to uncover hidden patterns and predict patient presence. This machine learning method is accurate and computationally efficient [15].

Compared to existing algorithms for chronic disease like diabetes, heart disease, and cancer, a suggested IoT-based patient monitoring scheme achieves the highest level of security at 95.87% by using DLMNN for HD diagnosis and therapy [4]. With 98.5% accuracy in heart illness prediction, the study suggests a smart healthcare system that can diagnose cardiac problems and treat patients before myocardial infarctions. It does this by combining feature fusion with ensemble deep learning [5]. Bharti., Rohit., et al., using the UCI Machine Learning Heart Disease dataset, this study evaluates dl and ml algorithms for heart disease diagnosis. Results are encouraging and have been verified using an accuracy and confusion matrix. Isolation Forest normalizes data and manages characteristics that aren't important and they got 94.2% accuracy.

Yadav., Anup Lal., Kamal Soni., and Shanu Khare., in order to propose an ensemble classifier for hybrid classification, this study examines many machine learning classifiers. This provides important information for early diagnosis and treatment in healthcare and aids physicians in changing their approach and diagnosis. To increase the prediction, a novel approach utilizing machine learning techniques is presented here. By integrating characteristics and classification methods, the strategy achieves an improved performance by 88.7% and predicted using HRFLM [8]. Accuracy and efficiency in medical professionals have been improved by researchers who have created automated systems that use Deep Neural Networks to identify the likelihood of a heart attack; on various datasets they attained 87.64% [9]. Researchers are using machine learning algorithms to predict heart-related disorders (CVDs) in time for appropriate treatment, using methods like SVM, RF, and ANN, despite varying effectiveness [10].

Table: 1 Comparison of different algorithm used for Heart Disease

S.N O	Algorithm	Description	Dataset	Findings
1	Hyper parameter optimization	By applying some classification algorithms on heart disease dataset they got best result in Hyper-Parameter Optimization.	Talos	Comparing HPO algorithm with SVM, KNN and result 90.78%
2	Support Vector Classifier, NN, Random Forest Classifier	Dataset split into training and test sets. Following data Preprocessing, SVM, ANN, and Random Forest models were applied.	UCI Machine Learning Repository	The three models are trained and tested with maximum score of 84%, 83.5%, 80%
3	ML & DL	By applying different machine learning algorithms and then using deep learning.	UCI ML Heart Disease dataset	The comparison methods used includ confusion matrix metrics such as precision specificity, and sensitivity. Among the 13 features in the dataset, the KNeighbor classifier demonstrated superio performance in the machine learnin approach following data Pre-processing.
4	HRFLM	Finding ways to analyze unprocessed medical data about the heart can saved and helps in early identification of irregularities in cardiac Diseases.	UCI dataset	In terms of heart disease prediction, it turned out to be fairly accurate.
5	Deep Neural Network	Heart attack prediction using ML and DL algorithms has a promisin future, as recent research has increasingly focused on medical datasets.	UCI Repository.	A comparison was made with many othe ML algorithms that had previously bee used to predict heart attacks, and the result clearly reveal that deep neural network outperform all other algorithms.

The table presents a summary of various algorithms applied to heart disease prediction datasets, detailing the methodologies, datasets, and key findings.

Hyper-Parameter Optimization (HPO): The study applied hyper-parameter optimization to a heart disease dataset, using Talos as the optimization tool. This approach was compared to established algorithms like SVM and KNN. The optimization produced a superior result of 90.78%, demonstrating the efficacy of fine-tuning model parameters to improve performance.

SVC, NN, and RF Classifier: Using the UCI ML Repository, the dataset was divided into training and testing sets. Following pre-processing, the models—SVM, ANN, and RF—were used. The findings showed maximum scores of 84% for SVM, 83.5% for ANN, and 80% for RF, indicating the classifiers comparable performance on the dataset.

ML and DL: This investigation used a variety of ML methods, followed by DL approaches, on the UCI ML Heart Disease dataset. Precision, specificity, and sensitivity were used as confusion matrix measures in the evaluations. The KNeighbors classifier performed the best of the 13 features after data pre-processing, demonstrating its effectiveness in this situation.

HRFLM (Hybrid Random Forest and Linear Model): The study investigated the use of HRFLM on raw medical data to predict cardiac disease. Using the UCI dataset, this technique demonstrated high accuracy in early detection of heart abnormalities, highlighting the promise of hybrid models in medical data processing.

Deep Neural Network (DNN): The study highlighted the prospective future of applying ML and DL algorithms to predict heart attacks. Using the UCI Repository, the study compared DNN to previously used machine learning techniques. The results showed that DNN outperformed all other algorithms, demonstrating its better prediction potential in medical datasets.

These papers show the progress and comparative effectiveness of several ML and ML algorithms in predicting heart disease, emphasizing the relevance of algorithm selection and parameter optimization in producing accurate findings.

Methodology

This methodology combines DL and ML techniques to predict CVD with high accuracy. The technique includes data preprocessing, feature selection, model training with multiple algorithms, and Talos optimization. It uses ensemble models and Internet of Things-based patient monitoring to improve forecast accuracy and efficiency. Predicting heart disease using the UCI and Kaggle heart datasets requires an organized methodology that begins with data gathering from these sources. To assure the quality and relevance of the obtained data, it is preprocessed using filtering, feature engineering, and transformation methods. The dataset's predictive potential is then enhanced by selecting critical features using Chi-Square Analysis, Recursive Feature Elimination, and Principal Component Analysis.

Fig1: The improved dataset is then used to train a number of models, including Talos-optimized DL models and traditional ML models such as KNN, SVM, RF, and NB. To ensure robustness, the models' performance is evaluated using metrics such as accuracy, precision, recall, F1-score, ROC-AUC, and the confusion matrix. An ensemble model is created by stacking many models' predictions, which are then merged using a logistic regression meta-model to improve accuracy. Finally, our ensemble model predicts the risk of CVD, offering a comprehensive strategy to early heart disease detection and prevention.

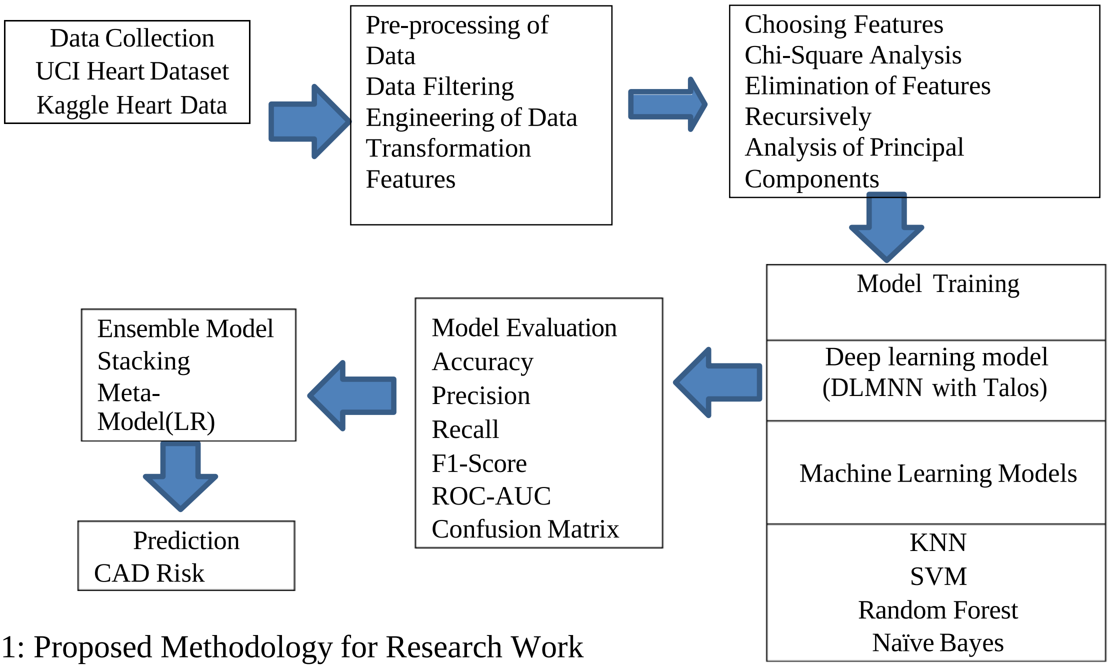

Figure 1: Proposed Methodology for Research Work

<u>Data Collection</u>:

Collect data from a variety of sources, including wearable devices, clinical trials, electronic health records, and online databases such as Kaggle and the UCI heart disease dataset. Make sure the data contains pertinent factors such as age, gender, blood pressure, cholesterol, smoking status, physical activity, and so on.

<u>Data preprocessing:</u>

Data Cleaning: Missing values are addressed using imputation techniques such as mean/mode imputation or K-Nearest Neighbors imputation. Remove any duplicates and outliers. Data Transformation: Use Min-Max Scaling or Standardization to normalize numerical characteristics. For categorical variables, use one-hot encoding. Feature Engineering: Develop new features from current data to improve model performance, such as BMI, risk scores, and composite indicators. Feature Selection: Apply feature selection techniques to identify the most important features for predicting cardiovascular disease. Use methods like the Chi- Square test, RFE, and PCA.

Model Training:

To develop a comprehensive cardiovascular disease prediction system, the following steps will be undertaken:

<u>Deep Learning Model</u>: Train a Deep Learning Modified Neural Network (DLMNN) with properly chosen features. Use Talos, a Keras hyperparameter optimization toolkit, to fine-tune the model and determine the optimal configuration. Examine numerous machine learning models, including KNN, SVM, RF, and NB. Optimize each model's parameters via hyperparameter tuning approaches such as grid search or random search. Split the dataset into training and testing sets before evaluating the models. To assess their performance, employ metrics like as accuracy, precision, recall, F1-score, and the ROC-AUC curve. Use k-fold cross-validation to ensure robustness and generalizability. Use confusion matrices to visualize performance. Use an ensemble model, such as stacking, to integrate predictions from different machine learning models. Increase forecast accuracy by applying a meta-model, such as logistic regression, to the aggregated predictions. Use the trained ensemble model to forecast cardiovascular disease risk in new patients. Integrate this model into an Internet of Things-based patient monitoring system to provide continuous monitoring and real-time risk assessment. This systematic method uses sophisticated machine learning techniques to create a strong prediction model for cardiovascular illness, with the goal of achieving high accuracy and practical applicability in healthcare contexts.

Problem Identification

Identifying and forecasting cardiac disease remains a key problem in healthcare, necessitating extensive research efforts that employ modern computational tools. Studies by Sharma and Parmar (2020), Jindal et al. (2021), and Pasha et al. (2020) highlight the complexity and variety of cardiovascular diseases, emphasizing the need for accurate and timely prognostic models. Traditional systems frequently rely on manual risk factor evaluation, which is subjective and prone to mistake. Machine learning (ML) techniques, as emphasized by Sharma and Parmar (2020) and Jindal et al. (2021), have demonstrated potential in automating this process by evaluating vast amounts of patient data to find hidden patterns and correlations that indicate heart disease risk. However, the problem is to optimize these algorithms to handle different datasets that include demographics, clinical, and genetic information.

It intends to employ DL and ML techniques to predict and forecast cardiac disease. Heart disease is one of the major causes of death globally, and early detection is critical for avoiding medications and properly managing the condition. Because traditional approaches are inadequate for forecasting heart illness, cutting-edge technologies such as deep learning and machine learning are being investigated. Data that is readily available, such as the Heart Disease UCI and Kaggle datasets, is useful for training prediction models. The study used SVMs, artificial neural networks, and KNNs to predict cardiovascular disease. This investigation incorporates a range of forecasting methods, including multiple machine learning techniques.

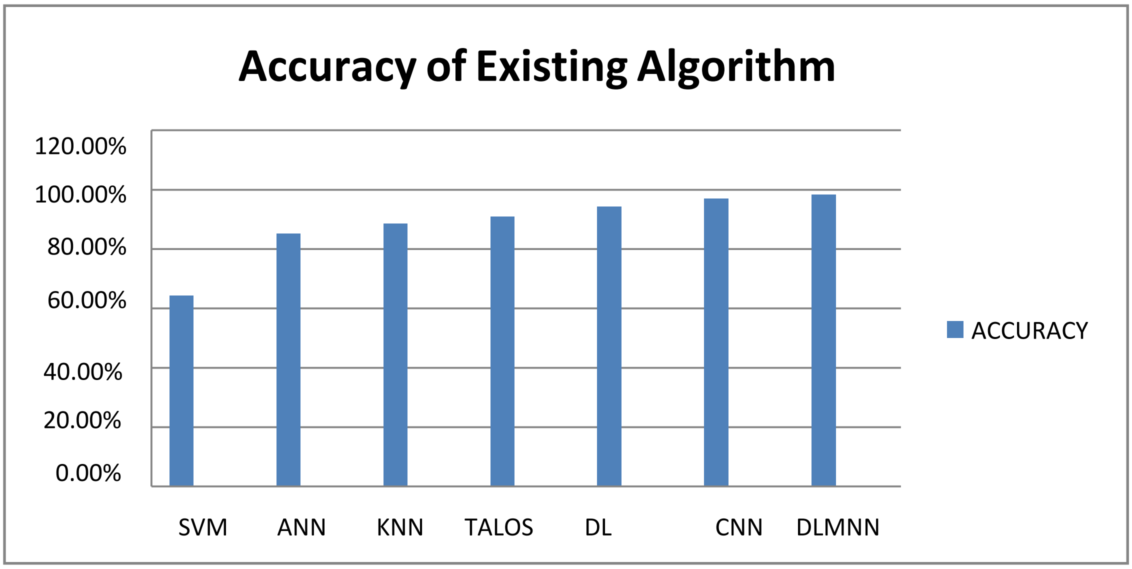

Fig 2: Accuracy Chart of CVD Prediction using various existing ML algorithm.

Table 2: Accuracy details of various existing ML algorithm for cardiovascular disease

Algorithm	ACCURACY
SVM	64.40%
ANN	85.24%
KNN	88.52%
TALOS	90.76%
DL	94.20%
CNN	97%
DLMNN	98.25%

Table 2 and Figure 2 show a comparison of several algorithms depending on their accuracy in a specific application. The Support Vector Machine (SVM) achieves an accuracy of 64.40%, which, while impressive, is surpassed by other techniques. Artificial Neural Networks (ANN) greatly outperform this, achieving 85.24%, proving their better capacity to capture complicated patterns. K-Nearest Neighbors (KNN) improves accuracy to 88.52%, demonstrating good management of the dataset's structure. The TALOS method, an advanced hyperparameter optimization technique, improves accuracy to 90.76%, demonstrating the value of fine-tuning model parameters. Deep Learning (DL) algorithms produce even better accuracy of 94.20%, demonstrating their ability to understand complicated information via numerous layers of abstraction. Convolutional Neural Networks (CNN), which excel in image-related tasks, achieve an impressive 97% accuracy, demonstrating its efficacy in recognizing spatial hierarchies. The Deep Learning Modified closest Neighbor (DLMNN) strategy achieves the greatest accuracy (98.25%), combining deep learning's feature extraction strengths with the strong classification skills of closest neighbor approaches. This development demonstrates the rising performance of more advanced models, with DLMNN standing out for its perfect combination of feature extraction and classification abilities. Each increase in algorithm complexity is associated with a significant gain in performance, highlighting the necessity of sophisticated approaches in obtaining high accuracy in data-driven jobs.

Conclusion

The potential for ML and DL algorithms to revolutionize healthcare is demonstrated by their application in the prediction of heart illness. The critical demand for effective early detection of cardiovascular disease (CVD) is addressed by evaluating several datasets containing demographic, clinical, and diagnostic data. This study demonstrates significant increases in prediction accuracy when compared to older methodologies, with ML models achieving solid area under the curve (AUC) values above 0.85 and DL models frequently above 0.90. These findings show that deep learning approaches such as convolutional neural networks (CNNs) and deep learning modified nearest neighbor (DLMNN) models outperform other methods, with accuracies of 97% and 98.25%, respectively. The progression from SVM to DLMNN illustrates the increasing effectiveness of sophisticated algorithms in capturing complex patterns inherent in heart disease risk prediction. Advanced optimization techniques like Talos further enhance accuracy by fine-tuning model parameters, pushing it to 90.76%. Such methodologies not only improve early diagnosis but also facilitate personalized treatment strategies, potentially reducing healthcare costs and improving patient outcomes on a global scale.

By leveraging high-quality datasets from sources like UCI and Kaggle, this research ensures the reliability and generalizability of predictive models across diverse populations. The integration of ensemble models and IoT-based patient monitoring systems further enhances prediction accuracy and real-time risk assessment capabilities. Moving forward continued research into integrating ML and DL techniques promises to refine these models further, making them indispensable tools in clinical settings for combating cardiovascular disease effectively. In conclusion, the deployment of advanced computational methodologies in heart disease prediction signifies a paradigm shift towards more precise, data-driven healthcare solutions. These efforts not only advance the field of predictive analytics but also hold promise for transforming patient care through early detection and intervention strategies tailored to individual health profiles.

References

1. Sharma, Sumit, and Mahesh Parmar. "Heart diseases prediction using deep learning neural network model." *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* 9.3 (2020): 2244-2248.
2. Jindal, Harshit, et al. "Heart disease prediction using machine learning algorithms." *IOP conference series: materials science and engineering*. Vol. 1022. No. 1. IOP Publishing, 2021.
3. Pasha, Syed Nawaz, et al. "Cardiovascular disease prediction using deep learning techniques." *IOP conference series: materials science and engineering*. Vol. 981. No. 2. IOP Publishing, 2020.
4. Sarmah, Simanta Shekhar. "An efficient IoT-based patient monitoring and heart disease prediction system using deep learning modified neural network." *Ieee access* 8 (2020): 135784-135797.
5. Ali, Farman, et al. "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion." *Information Fusion* 63 (2020): 208-222.
6. Bharti, Rohit, et al. "Prediction of heart disease using a combination of machine learning and deep learning." *Computational intelligence and neuroscience* 2021 (2021).
7. Yadav, Anup Lal, Kamal Soni, and Shanu Khare. "Heart Diseases Prediction using Machine Learning." *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, 2023.
8. Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." *IEEE access* 7 (2019): 81542-81554.
9. Ashraf, Mohd, M. A. Rizvi, and Himanshu Sharma. "Improved heart disease prediction using deep neural network." *Asian Journal of Computer Science and Technology* 8.2 (2019): 49-54.
10. Rindhe, Baban U., et al. "Heart disease prediction using machine learning." *Heart Disease* 5.1 (2021).
11. Nancy, A. Angel, et al. "Iot-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning." *Electronics* 11.15 (2022): 2292.
12. Sharma, Himanshu, and M. A. Rizvi. "Prediction of heart disease using machine learning algorithms: A survey." *International Journal on Recent and Innovation Trends in Computing and Communication* 5.8 (2017): 99-104.
13. Gonsalves, Amanda H., et al. "Prediction of coronary heart disease using machine learning: an experimental analysis." *Proceedings of the 2019 3rd International Conference on Deep Learning Technologies*. 2019.
14. Mehmood, Awais, et al. "Prediction of heart disease using deep convolutional neural networks." *Arabian Journal for Science and Engineering* 46.4 (2021): 3409-3422.
15. Jagtap, Abhijeet, et al. "Heart disease prediction using machine learning." *International Journal of Research in Engineering, Science and Management* 2.2 (2019): 352-355.