ISSN 2320 - 2602
Volume 13 No. 1, January 2024
International Journal of Advances in Computer Science and Technology
Available Online at http://www.warse.org/IJACST/static/pdf/file/ijacst011312024.pdf
https://doi.org/10.30534/ijacst/2024/011312024

An Overview of Cardiac Disease Diagnosis using Machine Learning Algorithms

Shreyas R Kale, Shridhar S, Siddarth Y K, Siddharth S Sonavane
Alvas Institute of Engineering and Technology, India, shreyaskale907@gmail.com
Alvas Institute of Engineering and Technology, India, shridharsshridhar2003@gmail.com
Alvas Institute of Engineering and Technology, India, siddarthilk@gmail.com
Alvas Institute of Engineering and Technology, India, siddharth.sonavane2002@gmail.com@gmail.com

Received Date : December 02 , 2023 Accepted Date : December 28, 2023 Published Date : January 07, 2024

ABSTRACT

This abstract investigates the use of machine learning algorithms in the detection of cardiac illness, namely Support Vector Machines (SVM), Logistic Regression, K-Nearest Neighbors (KNN), and Decision Trees. Preprocessing and gathering patient data, including demographic information, medical history, and other health markers, is part of the study. Features are selected based on their relevance to heart disease diagnosis, and labeled datasets are employed for training and validation. SVM, with its capacity to find optimal hyperplanes, is employed to discern patterns in the data. Logistic Regression, known for its simplicity and interpretability, aids in probability estimation. KNN is a flexible instance-based algorithm that makes predictions by utilizing nearby data points. Decision trees are used because they may represent intricate linkages and provide clarity in decision-making. The abstract explores how

Comprehensible these algorithms are and how that affects the precision with which heart disease is diagnosed. Robust generalization is ensured by model validation approaches like cross-validation. The study also explores continuous monitoring applications, providing ongoing risk assessments and contributing to personalized treatment plans. The choice of algorithm depends on dataset characteristics and the interpretability requirements of healthcare professionals.

Key words: Cardiovascular, Continuous Monitoring Effectiveness, Machine Learning

1. INTRODUCTION

Cardiovascular diseases, encompassing a spectrum of conditions affecting the heart and blood vessels, stand as a formidable and pervasive global health challenge. The impact of these diseases extends far beyond mere physiological implications, infiltrating societal structures and posing substantial economic burdens. Between the early 1900s and the present, the incidence and fatality rate of cardiovascular illnesses have shown an unsettling rising trend, reaching a point in the modern age when they constitute a substantial share of the world's morbidity and death[1].

In the historical context, heart diseases emerged as a relatively inconspicuous contributor to mortality, constituting 10% of total deaths in the early twentieth century. However, this unassuming profile underwent a dramatic transformation as the late twentieth century witnessed a staggering 25% increase in death rates attributable to cardiovascular diseases. What was once perceived as an affliction primarily affecting individuals aged 65 and older has evolved into a pervasive and indiscriminate threat, eclipsing infectious diseases as the predominant cause of death worldwide.

The World Health Organization (WHO), which states that cardiovascular illnesses are the world's top cause of death, emphasizes how serious the issue is.[1] In 2004 alone, an estimated 17.1 million lives succumbed to these diseases, marking a daunting 29% of total global deaths. This escalating toll is not confined to developed nations; instead, it infiltrates the heart of developing countries, where inadequate healthcare infrastructures struggle to cope with the rising tide of cardiovascular challenges.

Within this global tapestry of health disparities, Iran, for instance, grapples with heart diseases contributing to a staggering 46% of total deaths [3]. The exponential growth of cardiovascular diseases brings with it not only physiological challenges but also profound socio-economic implications. Effective prevention, early detection, and treatment options are necessary in light of the mounting financial and physical pressures placed on societies.

Amid these challenges, the quest for early diagnosis emerges as a pivotal frontier in the battle against cardiovascular diseases. Traditional diagnostic methods, such as echocardiograms, while offering non-invasive insights into cardiac health, are not without their challenges. Result interpretation is often a laborious and time-intensive task, demanding expertise that may not be readily available [3].

To navigate these complexities, the convergence of healthcare and cutting-edge technology becomes imperative. In this context, the exploration of machine learning techniques, particularly data mining, represents a promising avenue. These techniques hold the potential to sift through vast medical datasets intelligently, discern hidden patterns, and predict the onset of heart diseases with enhanced accuracy and efficiency.

Beyond the realm of diagnostics, the economic implications of cardiovascular diseases loom large, particularly in developed nations grappling with rising healthcare costs. The imperative for early detection emerges not only as a medical necessity but as an economic strategy to mitigate the exorbitant financial toll associated with advanced-stage treatments [5].

This talk explores the rapidly changing field of machine learning applications in the medical domain and explores the complexities involved in creating AI systems that support medical decision-making. The toolbox of machine learning models—which includes everything from logistic regression and support vector machines to artificial neural networks—becomes crucial to the investigation of improved diagnostic abilities[2]. To navigate these complexities, the convergence of healthcare and cutting-edge technology becomes imperative. In this context, the exploration of machine learning techniques, particularly data mining, represents a promising avenue [8]. These techniques hold the potential to sift through vast medical datasets intelligently, discern hidden patterns, and predict the onset of heart diseases with enhanced accuracy and efficiency.

The journey into the machine learning realm is augmented by considerations of data pre-processing, the delicate balance of datasets, and the strategic selection of features. These components serve as the cornerstone for optimizing machine learning models to achieve higher performance. The proposed method, a fusion of an imperialist competitive algorithm and K-nearest neighbor, represents a novel foray into synergistic techniques that promise not only reduced feature counts but heightened classification accuracy.

The focus shifts to include the wider uses of machine learning in healthcare as the conversation progresses. The focus of the conversation shifts to how early diagnosis may significantly save healthcare expenses and give medical practitioners insightful information. The significance of careful training data, strong cross-validation methods, and sensible performance assessment measures is emphasized.

In summation, this intricate tapestry of exploration navigates through the historical nuances, global impact, economic ramifications, and technological frontiers associated with cardiovascular diseases. The intersection of healthcare and machine learning emerges not merely as a scientific pursuit but as a beacon of hope in the quest for early detection, improved diagnostics, and ultimately, the mitigation of the profound challenges posed by cardiovascular diseases on a global scale.

2. MACHINE LEARNING

Modern healthcare relies heavily on machine learning, especially when it comes to diagnosing heart disease. Machine learning approaches help to identify cardiovascular diseases more accurately and quickly by utilizing sophisticated algorithms and processing capacity [4]. Here is a quick summary of the ways in which machine learning may be used to diagnose heart disease.

1. Data Collection and Pre-processing: The first step in using machine learning to diagnose cardiac disease is gathering a variety of patient data, such as imaging data from electrocardiograms (ECG) [3], echocardiograms, and angiograms, as well as medical histories, lifestyle details, and test results. Preparing this material for analysis entails cleaning and arranging it.
2. Feature Extraction: Variables or features that might be predictive in detecting heart disease are taken out of the data. For machine learning models to produce precise predictions, these properties are essential. Features may include blood pressure measurements, cholesterol levels, demographic data, and other pertinent health markers.
3. Algorithm Training: Labelled datasets, in which the input data is connected to matching outputs (the presence or absence of heart disease), are used to train supervised learning algorithms.[6] Frequently used machine learning techniques for diagnosing cardiac disease include neural networks, logistic regression, decision trees, and support vector machines.
4. Model Validation: The model is verified using different datasets after training to see how well it performs on fresh, untested data. In order to make sure the model is not over fitting to the training set and that it generalizes properly, cross-validation techniques are used.
5. Prediction and Risk Assessment: The model may be used to forecast fresh patient data after it has been trained and verified. Machine learning algorithms can help healthcare providers make judgments regarding additional diagnostic tests and treatment strategies by estimating the likelihood of cardiac disease and categorizing patients into risk groups.
6. Integration with Imaging Data: Medical imaging data analysis is one area where machine learning excels. When interpreting complicated pictures from imaging modalities like angiograms and echocardiograms, algorithms can help detect structural abnormalities, blockages, or other heart problems.
7. Continuous Monitoring and Adaptive Models: It is possible to create machine learning models that will continuously monitor patients and adjust over time to their changing health situations. Proactive

intervention and customized treatment plans are made possible by this.

8. Ethical Considerations: Ethics and privacy issues must be carefully considered when using machine learning in healthcare, especially when diagnosing heart disease. It is crucial to guarantee the privacy and security of patient data.

In conclusion, machine learning improves assessment efficiency and accuracy for heart disease diagnosis, helping medical practitioners identify and treat the condition early. It enhances conventional diagnostic techniques and advances data-driven, individualized patient care strategies.

3. MACHINE LEARNING ALGORITHMS

3.1 Logistic Regression

A statistical technique for binary classification issues is logistic regression, which is appropriate in situations where the objective is to determine whether an instance falls into one of two groups [10]. When predicting the presence or absence of heart illness based on a set of input variables, logistic regression can be a useful tool in the diagnosis of heart disease. This is how logistic regression aids in the diagnosis of heart disease:

- Binary Classification: Determining if a patient has heart disease (positive class) or not (negative class) is a common step in the diagnosis of heart disease. Such binary classification problems are a good fit for logistic regression [12].
- Probability Estimation: A probability estimate of the chance that a certain patient has heart disease is provided by logistic regression. After that, a binary forecast is made by comparing this likelihood to a predetermined threshold.
- Interpretability: Logistic regression models are easily comprehensible and straightforward. The direction and magnitude of each input feature's effect on the prediction are indicated by the coefficients that are assigned to it. In the medical field, interpretability is critical to winning over healthcare experts, therefore this transparency is beneficial [12].
- Risk Assessment: By using logistic regression, odds ratios—which represent the probability of an event, like heart disease, occurring—can be computed. This assists in determining the risk connected to various circumstances and supports the decision-making process for medical professionals.
- Handling Linear Relationships: A linear relationship between the dependent variable's log-odds and the independent variables is the underlying assumption of logistic regression. This is suitable in cases where the characteristics and the risk of heart disease have a linear or nearly linear relationship [13].
- Feature Importance: The effect of each input feature on the likelihood of getting heart disease is indicated by the coefficients in a logistic regression. Higher absolute coefficient features make up a larger portion of the forecast.
- Low Complexity: Because it is less likely to over fit, logistic regression is appropriate in scenarios when there is a shortage of data. It achieves a compromise between performance and model complexity.
- Quick Training and Prediction: Logistic regression models are well-suited for real-time or near-real-time applications because to their computational efficiency and rapid training.

It's crucial to remember that, despite its benefits, logistic regression might not be as good at capturing intricate nonlinear correlations in the data as more sophisticated models like neural networks. The features of the dataset, the intricacy of the relationships, and the needs for interpretability of the medical experts engaged in the diagnostic process all influence the choice of algorithm. In comprehensive cardiac disease diagnostic systems, logistic regression is frequently employed as a baseline model and may be supplemented by more sophisticated methods.

3.2 Support Vector Machine

Support Vector Machines (SVM) are an effective machine learning technique for classification tasks that can be used in the detection of cardiac disease. When it's necessary to divide data into two groups according to a set of characteristics, SVM comes in handy. The following is how SVM may be used to diagnose heart disease:

- Data Preparation: Get pertinent patient data, such as demographics, medical history, and readings from an electrocardiogram (ECG) and other health markers like blood pressure and cholesterol [2].
- Feature Selection: Gather pertinent patient data, such as medical history and demographics. Choose and identify key elements from the dataset that will probably help with the proper categorization of heart disease. Risk factors and diagnostic assessments are a couple examples of these aspects. or, as well as a number of health markers including blood pressure, cholesterol, and ECG readings [3].
- Data Preprocessing: Preprocess the data by dealing with any outliers, scaling or normalizing features, and managing missing values. Normalization is frequently essential since SVM is sensitive to the size of the features.
- Labelling: Give the data labels that reflect whether or not cardiac disease is present. The SVM model will be trained and validated using this labeled dataset.
- Training the SVM Model: The goal of the SVM method is to identify the hyperplane that divides the data into distinct groups the best. By altering settings

to optimize the margin between classes, the SVM learns the ideal hyperplane during the training phase. The classifications would normally represent people with and without heart disease in the context of a diagnosis of heart disease.

- Kernel Function: Select the proper kernel function. Different kernel functions, including radial basis function (RBF), polynomial, and linear, can be used by SVM to translate data into a higher-dimensional space, making it simpler to identify a hyperplane that divides classes.[7]
- Model Validation: Examine the SVM model using a different dataset that wasn't used for training. In order to evaluate the model's generalization performance and make sure it functions well on fresh, untested data, methods like cross-validation are used.
- Prediction: The SVM model may be used to predict if a fresh collection of patient characteristics correlates to a likelihood of heart disease once it has been trained and validated. Based on the data point's location in relation to the hyper plane, SVM allocates it to one of the classes.
- Interpretability: Because the margin and support vectors—data points that affect the hyper plane's position—can be examined, support vector machine learning (SVM) yields findings that are easy to understand. This can help medical practitioners comprehend the elements that went into the categorization.
- Continuous Monitoring: SVM models may be modified to provide continuing risk assessments, enable individualized treatment programs, and monitor patients continuously.

It's crucial to remember that the features of the dataset and the particular needs of the task of diagnosing heart disease determine which machine learning algorithm—including SVM—is best[9]. SVM is only one of many classification tools available, and the type of data it is used with will determine how effective it is.

3.3 Decision Tree

Heart disease detection can be aided by the flexible and comprehensible machine learning approach known as decision trees. Here's a detailed breakdown of how Decision Trees may be applied in this situation:

- Data Collection: Compile a dataset with pertinent patient data, including demographics, medical history, and other health indicators, such as heart disease risk factors.
- Data Pre-processing: Take care of missing values, deal with outliers, and scale or normalize features as necessary to clean up and preprocess the data.
- Feature Selection: Determine the key elements that are most likely to play a role in the diagnosis of heart disease. These characteristics might be blood pressure, cholesterol, age, gender, and the outcomes of diagnostic testing.
- Labeling: Label the data such that it is clear whether or not each patient has cardiac disease. The Decision Tree model will be trained and validated using this labeled dataset.
- Training the Decision Tree: Through recursive segmentation of the data according to the chosen characteristics, the Decision Tree algorithm acquires decision-making skills. The algorithm selects the feature that offers the best class separation at each node.
- Splitting Criteria: Choose the standards by which nodes will be divided. The optimal feature and threshold for dividing the data at each node are determined by the algorithm by assessing several splitting criteria, such as information gain or Gini impurity.[3]
- Tree Structure: The method builds a tree structure as it goes along, with each internal node denoting a choice based on a particular attribute and each leaf node representing a class label (such as the presence or absence of heart disease) [5].
- Model Interpretability: Decision trees' interpretability is one of its main benefits. The decision-making process may be understood by analyzing the resulting tree structure. The characteristics that most influence the diagnosis can be readily interpreted by medical practitioners.
- Model Validation: Analyze the Decision Tree model with a different dataset that wasn't used for training. To evaluate the model's generalization performance and make sure it works well on fresh, untested data, methods such as cross validation are used.
- Prediction: The Decision Tree may be used to determine if a fresh set of patient characteristics is associated with an increased risk of heart disease after it has been trained and verified. When the patient's characteristics reach a leaf node with the expected class, they continue to move through the tree, according to the decision criteria at each node.
- Continuous Monitoring: Decision trees may be modified to provide continuing risk assessments, help with individualized treatment regimens, and monitor patients continuously.

It's important to remember that Decision Trees, particularly complicated ones, can occasionally be prone to over fitting. Reducing over fitting and enhancing generalization performance may be accomplished by utilizing strategies like pruning or group approaches like Random Forests [13]. The particulars of the dataset and the requirements for healthcare professionals' interpretability determine which machine learning method is best.

3.4 K Nearest Neighbors

A machine learning approach called K-Nearest Neighbors (KNN) is utilized for regression and classification problems. KNN may be used to find patterns in patient data and forecast the presence or absence of heart illness in the context of diagnosing heart disease [6]. This is how KNN may be applied to the diagnosis of heart disease.

- Data Collection: Compile a dataset with pertinent patient data, such as demographics, medical history, and other heart disease-related health factors.
- Data Preprocessing: Preprocess the data by dealing with any outliers, scaling or normalizing features, and managing missing values. Because KNN is sensitive to feature scale, normalization is frequently crucial [7].
- Feature Selection: Determine the critical elements that are most likely to help with a precise diagnosis of heart disease. These characteristics might include blood pressure, cholesterol, age, gender, and the outcomes of diagnostic testing.
- Labeling: Label the data to reflect whether or not each patient has cardiac disease. The KNN model will be trained and validated using this labeled dataset.
- Training the KNN Model: KNN is an algorithm for instance-based learning. Without building an explicit model, the model learns by storing the whole training dataset in memory. The associations between instances in the feature space are committed to memory by the algorithm.
- Choosing 'k': Choose 'k', the number of closest neighbors to be taken into account while forecasting. A critical decision that might affect the model's performance is the selection of 'k'. The best 'k' value may be found with the use of cross-validation procedures [6].
- Distance Metric: Select the distance metric that will be used to calculate how similar two data points are. Manhattan distance, Euclidean distance, and other bespoke metrics based on the data's properties are examples of common distance measures [4].
- Prediction: KNN finds the 'k' nearest neighbors in the feature space to anticipate a new, unseen data item. The anticipated class for the new data point is determined by the majority class among these neighbors.
- Model Interpretability: Because KNN is based on the majority class of adjacent examples, it offers an easy way to understand predictions. In contrast to other models like logistic regression or decision trees, it could be more difficult to comprehend how each factor affects the prediction.
- Model Validation: Utilize an alternative dataset that was not used for training to assess the KNN model. In order to evaluate the model's generalization performance and make sure it functions well on fresh, untested data, cross-validation techniques are used.
- Continuous Monitoring: KNN may be modified to provide continuing risk assessments, help with individualized treatment regimens, and monitor patients continuously.

When the decision boundary is complicated and nonlinear, KNN is especially helpful. It is flexible enough to adjust to different class distributions and responsive to subtle local trends in the data. However, the curse of dimensionality may have an impact on its performance, and for big datasets, it may be computationally costly [7][12]. The particulars of the dataset and the interpretability standards of medical practitioners choose the machine learning algorithm KNN included to use.

4. RESULT

1. Logistic Regression: Logistic Regression is often used as a baseline model due to its simplicity and interpretability. It performs well when the relationship between features and the likelihood of heart disease is approximately linear.
2. Support Vector Machine: SVMs, with their ability to find optimal hyperplanes, can handle complex decision boundaries. They are effective when the relationship between features and the presence of heart disease is non-linear.
3. Decision Tree: Decision Trees can capture complex relationships in the data. They are interpretable and can be useful when understanding the contribution of individual features to the diagnosis is important.
4. K Nearest Neighbors: KNN is sensitive to local patterns in the data and can adapt well to varying distributions of classes. It can be effective when the decision boundary is complex and nonlinear[8].

The effectiveness of these algorithms can vary based on factors such as the size and quality of the dataset, the nature of the features, and the specific characteristics of the heart disease being diagnosed [7][9]. In practice, the choice of algorithm might involve experimenting with multiple models and selecting the one that performs best on a given dataset. Additionally, ensemble methods or hybrid approaches combining multiple algorithms may further improve predictive accuracy.

For the most up-to-date and specific results, it's recommended to refer to recent research studies or clinical applications that have conducted experiments on heart disease diagnosis using machine learning algorithms.

5. CONCLUSION

In conclusion, the accurate prediction and diagnosis of heart diseases are paramount given the vital role of the heart in the human body. Looking forward, the future scope involves integrating advanced machine learning approaches for a more nuanced analysis, with an emphasis on early prediction to minimize mortality rates through increased disease awareness. The synergy of computer technologies and communication tools has significantly elevated medical practices. While Artificial Neural Network ensembles are potent, their complexity hinders widespread acceptance among practitioners [11].

We report on a decision support system for diagnosing cardiac illness that uses a Support Vector Machine (SVM) and radial basis function network topology [10]. Findings show that SVM has excellent classification accuracy, sensitivity, and specificity, confirming its standing as a reliable method for the detection of heart disease.

In summary, ongoing technological evolution and advancements in machine learning methodologies promise to improved healthcare outcomes.

REFERENCES

1. Machine Learning based Heart Disease Diagnosis using Non-Invasive Methods: A Review |Narender Kumar and Dharmender Kumar 2021 J. Phys.: Conf. Ser. 1950 012081.
2. Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare JIAN PING LI, AMIN UL HAQ, SALAH UD DIN, JALALUDDIN KHAN , ASIF KHAN 1 , AND ABDUS SABOOR.
3. Efficient machine learning based detection of heart disease | Ricardo Buettner Aalen University, Germany, Marc Schunter Aalen University, Germany | 2019 IEEE International Conference on E-health Networking, Application & Services (HealthCom)
4. Heart Disease Prediction Using Machine Learning Algorithms | 2020 International Conference on Electrical and Electronics Engineering (ICE3-2020).
5. Santhana Krishnan J and Geetha S, “Prediction of Heart Disease using Machine Learning Algorithms” ICIICT, 2019.
6. Aditi Gavhane, Gouthami Kokkula, Isha Panday, Prof. Kailash Devadkar, “Prediction of Heart Disease using Machine Learning”, Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology(ICECA), 2018.
7. Senthil kumar mohan, chandrasegar thirumalai and Gautam Srivastva, “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques” IEEE Access 2019.
8. M. Nikhil Kumar, K. V. S. Koushik, K. Deepak, “Prediction of Heart Diseases Using Data Mining and Machine Learning Algorithms and Tools” International Journal of Scientific Research in Computer Science, Engineering and Information Technology ,IJSRCSEIT 2019.
9. Amandeep Kaur and Jyoti Arora,“Heart Diseases Prediction using Data Mining Techniques: A survey” International Journal of Advanced Research in Computer Science , IJARCS 2015-2019.
10. Pahulpreet Singh Kohli and Shriya Arora, “Application of Machine Learning in Diseases Prediction”, 4th International Conference on Computing Communication And Automation(ICCCA), 2018.
11. Hazra, A., Mandal, S., Gupta, A. and Mukherjee, “ A Heart DiseaseDiagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review” Advances in Computational Sciences and Technology , 2017.
12. Heart Disease Diagnosis Using Machine Learning Algorithm Shashikant U. Ghumbre1 and Ashok A. Ghatol.
13. A Study of Heart Disease Diagnosis using Machine Learning and Data Mining Intisar Ahmed | California State University, San Bernardino | 12-2022