JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

HEART DISEASE IDENTIFICATION METHOD USING MACHINE LEARNING CLASSIFICATION IN E-HEALTHCARE

***Ms.Krushna G Jagtap*[1], *Dr. Anita Mahajan*[2]**

[1]*Student, of Computer Engineering, Ajeenkya D. Y. Patil School of Engineering, Lohegaon, Savitribai Phule Pune University, Pune, Maharashtra, India*

[2]*Assistant Professor, Dept. of Computer Engineering, Ajeenkya D. Y. Patil School of Engineering, Lohegaon, Savitribai Phule Pune University, Pune, Maharashtra, India*

Abstract

Heart disease is a life-threatening disorder that needs prompt and correct diagnosis to enhance patient outcomes. Traditional diagnostic approaches often have limits in both accuracy and efficiency. As a result, numerous machine learning and deep learning algorithms have been used to automate the prediction of cardiac disease, delivering a greater degree of accuracy. This study uses a variety of algorithms, including Support Vector Machine (SVM), Naive Bayes, Decision Tree, Random Forest, and Artificial Neural Networks (ANN), to predict the presence or absence of heart disease based on patient characteristics such as age, gender, cholesterol levels, blood pressure, and other health indicators. Each method has distinct capabilities; for example, the Random Forest algorithm excels at managing complicated datasets by generating an ensemble of decision trees, while SVM is well-known for its capacity to handle high-dimensional data. Naive Bayes is simple and fast, making it ideal for real-time predictions. Meanwhile, Decision Trees are easily interpretable. Furthermore, ANNs are used to capture complicated, nonlinear connections in data, whereas Convolutional Neural Networks (CNNs) are investigated for deep learning-based feature extraction and classification. The findings show that integrating these models greatly improves the prediction accuracy and reliability of heart disease diagnosis.

Keywords: Heart disease, Diagnosis, Machine learning, Deep learning, Predictive modelling, Random Forest classifier, Feature extraction, Normalization

1.Introduction:

Heart disease encompasses a variety of disorders that impact the structure and function of the heart, such as arrhythmias, heart failure, and coronary artery disease. Effective cardiac disease prediction and timely diagnosis are essential for enhancing patient care, lowering medical expenses, and saving lives. While clinical signs and risk factor analysis are the foundation of traditional risk assessment approaches, machine learning techniques may improve prediction accuracy and help healthcare make well-informed choices.

This study compares and evaluates the effectiveness of several machine learning algorithms for the prediction of heart disease. Based on a set of clinical and demographic characteristics, the research seeks to determine the best model for correctly classifying individuals as having or not having heart disease. K-Nearest Neighbors, Random Forest, Neural Network, Decision Tree, and Logistic Regression are among the models that have been assessed. Accuracy, precision, recall, and F1-score are among the relevant performance measures used in the training and assessment of the machine learning models.

The study compares and contrasts the various machine learning models, looking at their prediction power, shortcomings, and strengths. The research also looks at how model performance is affected by data pre-processing methods such feature scaling and class balancing. The results reveal the most accurate model for predicting heart disease and provide insights into the efficacy of each algorithm. The results of this study may help academics and healthcare professionals choose the best machine learning strategy for predicting heart disease. Accurately identifying those who are at risk allows for the implementation of preventive interventions, which enhance patient outcomes and the distribution of resources within healthcare systems.

Figure 1

2.Related work:

JIAN PING L et.al (2020) One of the most complicated illnesses that affects a large number of individuals worldwide is heart disease. In the medical world, especially in cardiology, early and effective detection of cardiac disease is crucial. In this work, we presented a machine learning-based approach to cardiac problem identification that is both accurate and efficient. The categorization methods used in the system's development include Standard features selection algorithms like Relief, Minimal redundancy maximal relevance, Least absolute shrinkage selection operator, and Local learning have been used to eliminate redundant and irrelevant features, while support vector machines, logistic regression, artificial neural networks, K-nearest neighbor, Naïve bays, and decision trees have also been employed [1].

Dwarakanath B et.al (2022) In the context of e-healthcare, online illness detection services have become more popular and have increased in quality due to developments in data mining, wearable technology, and cloud computing. Through early illness detection, e-healthcare services contribute to a lower mortality rate. At the same time, heart disease (HD) is a fatal condition, and early HD diagnosis is essential for patient survival. Clinical data analysis relies heavily on early HD diagnosis and classification [2].

Santhosh V et.al (2022) Using information gathered from prior patients as well as information entered by the user at that specific moment, this project started with the early identification of all likely symptoms and indicators that might eventually lead to the discovery of cardiac illnesses. Health care data utilized for surveillance in the modern day is more than just a time-building sequence of daily counts. Rather, a multitude of suggested demographic and symptom data, both regional and temporal, are accessible at the moment of execution. All of this information is included into our suggested strategy, which is a classification approach that contrasts data from that specific baseline distribution with current healthcare data [3].

Mohan Raja. Pulicharla et.al (2023) The pursuit of artificial intelligence gave rise to the science of machine learning. In the early days of artificial intelligence as a subject of study, several scholars were fascinated by the concept of allowing computers to learn from data. In addition to what were then called "neural networks," which were essentially perceptrons and other models that were later shown to be reimaginings of the generalized linear models of statistics, they attempted to address the problem using a number of symbolic approaches. Additionally, probabilistic reasoning was used, especially for automated medical diagnosis. In a more formal definition that is often used, Tom M.

Mitchell said that a computer program is said to learn from experience E with regard to a class of tasks T and performance measure P if its performance at tasks in T, as measured by P, rises with experience [4].

Amin Ul Haq et.al (2020) accurate diabetes detection has received a lot of attention. Developing a diagnostic system to successfully identify diabetes in the context of e-healthcare is a significant problem for the scientific community. By providing a framework to evaluate medical data for illness detection, machine-learning techniques are playing an increasingly important role in healthcare services. Among the shortcomings of the current diagnostic systems are their lengthy calculation times and poor forecast accuracy. We have put up a machine learning-based diabetes diagnostic system to address these problems [5].

Deepika Tenepalli et.al (2024) Since its use is necessary for day-to-day living, the Internet of Things (IoT) has been embraced in numerous applications in recent years. Additionally, the healthcare system is using this emerging technology to provide patients efficient emergency treatments. The number of illnesses and medical issues among individuals is rapidly increasing in the present situation. Because clinics and hospitals have limited space and medical resources, it is getting increasingly difficult to accommodate and offer healthcare services for more arriving patients. As a result, the healthcare industry began integrating IoT and assistive technologies to provide effective wireless healthcare services and to continuously monitor patients[6].

Sumit Kumar et.al (2022) to develop creative solutions to clinical problems, medical technology inventors have recently concentrated on a variety of clinical medicines. However, several disadvantages still surface, such as high processing costs, higher mistake rates, reduced training capacity, large storage space requirements, and decreased precision. The suggested research study offers a novel cascaded extreme learning machine for efficient heart disease (HD) prediction in order to overcome these limitations. Data pre-processing involves the use of normalization and missing data screening techniques. The Framingham risk factor extraction module is used to extract features from the pre-processed data, and these features are then fused to create a feature vector[7].

S Siamala Devi et.al (2021) One of the most unexpected illnesses is heart disease, which has affected many people worldwide. Accurate and timely evidence of coronary heart disease is essential to medical treatment, particularly in the field of cardiology. Based on AI processes, a useful and accurate paradigm for identifying cardiac disease is put forward. Medical data with 335 features and 26 features is used to test this method. The greatest notable expectation precision of 88.4% is achieved by MLR. This approach is also compared to data on coronary heart disease in Cleveland. Furthermore, in this case, MLR outperforms other techniques [8].

Deepak Kumar et.al (2021) both the cardiovascular burden the overall number of deaths and the frequency of cardiovascular diseases are rising rapidly on a global scale. Heart failure (HF) ultimately develops in patients with heart disease. Any additional help for the current medical support systems is crucial so that the doctor can forecast the patients' chances of survival. In today's sophisticated healthcare systems, the cooperative use of IoT devices with machine learning has become crucial. The Cardiac Diagnostic Feature and Demographic Identification (CDF-DI) systems are presented in this paper. They are an Internet of Things (IoT)-enabled framework that is secured by Public Key Infrastructure (PKI) and has significant models that detect certain heart disease features linked to heart failure. We analyzed the Cardiac secondary dataset using statistical and machine learning methods in order to do this. It is well recognized that people with heart failure who have increased serum creatinine and serum sodium levels may have renal problems [9].

Arokia Jesu Prabhu L et.al (2020) While other industries progress with the help of cognitive computing, the health care industry is still developing and providing more benefits to all customers. The aging population leads to poor decision-making, which has a negative influence on treatment quality and increases treatment costs, adding to the already complicated healthcare system. However, there are a number of obstacles that hinder progress in this area, including query inconsistencies, user domain information sets, and gaps between the knowledge base and user queries. From 1-D cardiovascular beatings to automated discovery utilizing multi-dimensional clinical data, the fast advancement of machine learning and artificial intelligence for medical applications has already been shown in recent years[10].

Khalid Mahmood Aamir et.al (2021) This study's primary focus is on identifying aberrant signals and classifying arrhythmias into two groups: premature ventricular contraction and ventricular tachycardia. Determining whether a signal was obtained from a healthy or ill individual is the only goal of signal detection. The proposed study approach presents a mathematical model for the signal detector based on instantaneous frequency (IF) computation. Following the detection of a patient's signal, the classifier uses the signal as input and predicts the class label to classify the target illness. Templates for ventricular tachycardia and premature ventricular contraction are created independently for the classifier's application. In the spectral domain, the similarities between a given signal and both templates are calculated[11].

Huru Hasanova et.al (2022) Globally, cardiovascular disease is a leading killer of both sexes. It is possible to detect early signs of cardiac disease by monitoring a number of measures, including blood pressure, cholesterol, blood sugar, and body weight. The current healthcare system is being completely transformed by the technologies. We can now remotely monitor patients, save their data, and analyze it for further study thanks to the Internet of Things (IoT). For quick processing and effective event detection, new and sophisticated secure algorithms must be proposed. In this research, we provide SCA_WKNN, a Sine Cosine Weighted K-Nearest Neighbor (SCA) approach for cardiac illness prediction that uses blockchain data to train. Due to the immutability of its data, the blockchain provides a secure setting for storing patient information and an authentic source for learning data [12].

Abeer Thamara et.al (2021) Machine learning has gained popularity in recent years for usage in a broad range of applications and studies. In many domains, including medical research and the healthcare system, it is essential. In these situations, machine learning is utilized to forecast illnesses or discover significant trends in medical data. The benefits of machine learning approaches that aid in the development of effective support infrastructure for medical disciplines and enhance healthcare services are shown in this study, which also provides an overview of various machine learning algorithms and their applications across several domains. The primary goal of this survey is to highlight earlier efforts using machine learning algorithms in the healthcare system and to provide academics who want to investigate machine learning in the healthcare system all the information they need [13].

Sarah A. Alzakari et.al (2024) remote healthcare monitoring has changed dramatically as a result of medical technology advancements, which are essential for continued observation and personalized therapy. This is of utmost importance for addressing chronic health conditions like hypertension, which significantly increases the likelihood of cardiovascular disease, particularly in older individuals. This method increases accuracy by merging physical data from patients' regular medical monitoring with ECD from thorough medical records. Improved cardiac sickness prediction is a result of this novel strategy. answer that use state-of-the-art methods for machine learning and IoT to fulfill this need. To boost prediction accuracy, we use the potent Extreme Gradient Boosting (XGBoost) approach to efficiently search through large datasets for important features. With our method, it achieved a higher prediction accuracy of 99.4%, outperforming random forests, decision trees, and naive Bayes. This study presents a unique method for remote healthcare monitoring that combines advanced machine learning models with Internet of Things technology [14].

S. Punitha et.al (2024) An accurate and timely diagnosis might lower the COVID-19 death rate worldwide. Radiologists can more quickly remove and identify the aberrant spots using computer-aided diagnostic (CAD). This work suggests a paradigm that uses the Internet of Medical Things (IoMT) to facilitate early COVID-19 diagnosis by combining machine learning with intelligence-based e-Health service platforms. In the suggested system, WOA optimizes a WNN's learning parameters, momentum constant, hidden nodes, input characteristics, and starting weights. The suggested method uses a WNN classifier to identify the troublesome areas after extracting the Laws 16 Texture Energy Measures (LTEM) from the preprocessed CT lung images. The proposed method is evaluated using a publicly available COVID-19 dataset that contains both COVID-19 and non-COVID-19 cases. The findings show that the recommended approach has an 82% sensitivity, a 73.3% specificity, and an accuracy of 84.8% [15]

Subasish Mohapatra et.al (2023) one of the main causes of mortality worldwide is heart disease, often known as cardiovascular diseases (CVDs). Heart diseases, which are generally categorized as several kinds of faulty cardiac defects, account for around one in four fatalities. However, the identification of CVDs is a laborious procedure that requires human analysis of data gathered from various clinical testing. New techniques for automating the detection of such abnormalities in human cardiac problems should be developed in order to provide medical practitioners faster analysis by reducing the time it takes to get a diagnosis and improving results. Electronic Health Records (EHRs) are often utilized to identify useful data trends that improve the predictive power of machine-learning algorithms. In a variety of sectors, including healthcare, machine learning in particular greatly helps with problems like forecasts. For the benefit of humanity, the wealth of clinical data that is now accessible must be used. The predictive model for heart disease prediction presented in this research is based on stacking several classifiers at the base and meta levels. Combining several diverse learners yields the strong model output. The model obtained 92% prediction accuracy with a precision score of 92.6%, sensitivity of 92.6%, and specificity of 91%. The model's performance was evaluated using a variety of metrics, including accuracy, precision, recall, F1-scores, and area under the ROC curve values [16]

3. Methods details:

The proposed system aims to develop an accurate and reliable heart disease prediction model using machine learning techniques.

The ultimate goal is to facilitate early detection and risk assessment, enabling timely interventions and personalized treatment plans for better patient outcomes.

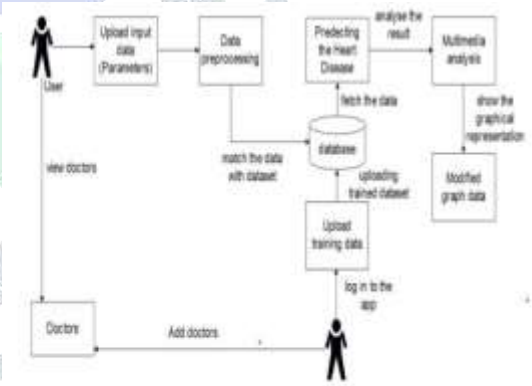


Figure 2: The Flow

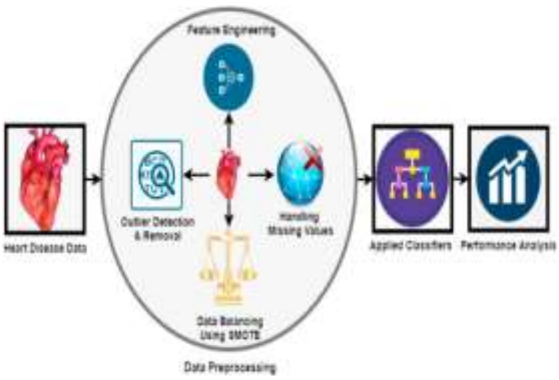


Diagram of Proposed Work

i. Data Collection: Collect relevant and accurate data from various sources like hospitals, clinics, healthcare datasets, or patient records. In the context of heart disease identification, the dataset should include variables like age, gender, cholesterol levels, blood pressure, ECG readings, smoking habits, exercise patterns, and other relevant clinical data. The dataset may be sourced from publicly available datasets like the **Heart Disease UCI Dataset**, electronic health records (EHR), or real-time health monitoring systems

Data Preprocessing: Clean the collected data to remove inaccuracies, deal with missing values, and prepare the data for the next stages in the process. Use imputation techniques (mean, median, or mode imputation) or remove rows/columns with too many missing values.

Data Normalization: Scale numerical data, especially for models that are sensitive to feature scale like Support Vector Machines (SVM) or Neural Networks. Min-Max scaling or Z-score standardization can be applied.

Data Encoding: Convert categorical variables (e.g., smoking status, chest pain type) into numerical format using one-hot encoding or label encoding.

Outlier Detection: Identify and remove outliers which may distort model performance.

ii. Feature Extraction: Derive new features from the original dataset that can better represent the underlying patterns. This is especially useful when raw data is not directly usable.

Principal Component Analysis (PCA): Reduces the dimensionality by creating new uncorrelated features from the existing ones.

Feature Engineering: Create new features or transform existing ones to improve model accuracy.

Example: Combining age and cholesterol levels into a “risk score” or creating binary variables like "high cholesterol" or "high blood pressure" based on predefined thresholds.

Domain Knowledge: Using insights from healthcare experts to engineer meaningful features that are relevant to heart disease prediction.

iii. Model Selection: Choose an appropriate machine learning model based on the dataset's nature, the problem's complexity, and the goal (classification of heart disease).

Algorithms:

Logistic Regression: Useful for binary classification, predicting the likelihood of heart disease.

Decision Trees: Can easily capture non-linear relationships and interpretability in the context of healthcare.

Random Forest: An ensemble method that can improve accuracy by averaging the predictions of multiple decision trees.

Support Vector Machines (SVM): Suitable for small to medium-sized datasets with a clear margin of separation.

K-Nearest Neighbors (KNN): A simple, instance-based method that classifies based on the proximity to other instances.

Neural Networks: Can be used for more complex datasets, potentially yielding higher accuracy for large and intricate dataset

Model Training: Train the selected model using the training data.

Hyperparameter Tuning: Adjust the hyperparameters (like tree depth, number of neighbors, regularization) to improve model performance using techniques like Grid Search or Randomized Search.

Cross-validation: Apply techniques like K-fold cross-validation to assess model stability and prevent overfitting by splitting the data into training and testing sets multiple times.

iv. Model Evaluation and Performance Metrics: Evaluate the model's performance using relevant metrics to understand its prediction accuracy and generalizability.

Confusion Matrix: A table that describes the performance of a classification model, showing the True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN)

5.Real-world Testing: Apply the trained model to real-world data, potentially from an e-Healthcare system or clinical environment.

Deployment: Integrate the model into an eHealth application for real-time heart disease prediction based on new patient data.

Continuous Learning: Ensure the model adapts over time as new data becomes available. Techniques like online learning can help the model stay up-to-date.

User Feedback: Incorporate feedback from medical professionals and patients to assess the effectiveness and practical utility of the model.

4. Results and Discussion:

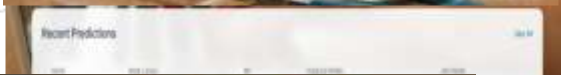
The experimental results demonstrated the performance of each model in predicting heart disease.

The random forest model achieved the highest accuracy of 87.5%, followed by SVM (86.3%), and logistic regression (81.9%).

Furthermore, it achieves a precision of 80%, recall of 90%, and an F1-score of 85%. These metrics provide a comprehensive assessment of the model's predictive performance.

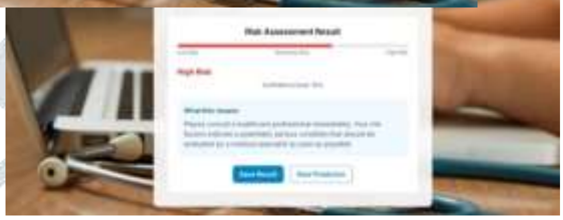
Login page:

Patient's Details:



Risk

Assessment Result:



5.

Conclusion:

In conclusion, the heart disease prediction model developed using various machine learning algorithms, including Logistic Regression, Random Forest, Neural Networks, and others, offers valuable insights into the early detection and risk assessment of heart disease. By systematically collecting and preprocessing comprehensive clinical and demographic data, the model can effectively analyze various risk factors and provide accurate predictions regarding the presence or absence of heart disease.

The developed heart disease prediction model has significant implications for healthcare practice. It enables early detection, risk stratification, and personalized patient care by providing accurate predictions regarding the presence or absence of heart disease. This model empowers healthcare professionals to make informed decisions, develop tailored treatment plans, and allocate appropriate resources based on individual patient profiles. The accurate identification of heart disease helps in reducing morbidity and mortality rates associated with cardiovascular conditions.

Future research should focus on incorporating more sophisticated feature engineering techniques and exploring other advanced machine learning algorithms to improve prediction accuracy.

6.Future Scope:

1.Integration of Advanced Machine Learning Models
2.Personalized Medicine and Predictive Analytics
3.Federated Learning for Data Privacy Edge Computing for Real-Time Diagnosis
4.Regulatory Compliance and Clinical Validation

7.References :

[1] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," *IEEE Access*, vol. 8, pp. 107562–107582, 2020, doi: 10.1109/ACCESS.2020.3001149.

[2] D. B., L. M., A. R., J. S. Kallimani, R. Walia, and B. Belete, "A Novel Feature Selection with Hybrid Deep Learning Based Heart Disease Detection and Classification in the e-Healthcare Environment," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–12, Sep. 2022, doi: 10.1155/2022/1167494.

[3] M. R. Pulicharla and D. A. Singhal, "Techniques for Machine Learning: Identifying Heart Disease within E-Healthcare through Implementation: Logistic Regression Model".

[4] S. Siamala Devi, G. Harini Karthika, and M. Deepika, "Machine Learning based Classification for Heart Disease Identification," *J. Phys. Conf. Ser.*, vol. 1916, no. 1, p. 012174, May 2021, doi: 10.1088/1742-6596/1916/1/012174.

[5] A. J. P. L *et al.*, "Medical information retrieval systems for e-Health care records using fuzzy based machine learning model," *Microprocess. Microsyst.*, p. 103344, Oct. 2020, doi: 10.1016/j.micpro.2020.103344.

[6] A. M. Alajlan, "Automatic Heart Disease Detection by Classification of Ventricular Arrhythmias on ECG Using Machine Learning," *Comput. Mater. Contin.*, vol. 71, no. 1, pp. 17–33, 2022, doi: 10.32604/cmc.2022.018613.

[7] H. Hasanova, M. Tufail, U.-J. Baek, J.-T. Park, and M.-S. Kim, "A novel blockchain-enabled heart disease prediction mechanism using machine learning," *Comput. Electr. Eng.*, vol. 101, p. 108086, Jul. 2022, doi: 10.1016/j.compeleceng.2022.108086.

[8] A. Thamara, M. Elsersy, A. Sherif, H. Hassan, O. Abdelsalam, and K. H. Almotairi, "A Novel Classification of Machine Learning Applications in Healthcare," in *2021 3rd IEEE Middle East and North AfricCOMMunications Conference(MENACOMM)*, Agadir, Morocco: IEEE,doi:10.1109/MENACOMM50742.2021.9678232.

[9] S. A. Alzakari *et al.*, "Enhanced heart disease prediction in remote healthcare monitoring using IoT-enabled cloud-based XGBoost and Bi-LSTM," *Alex. Eng. J.*, vol. 105, pp. 280–291, Oct.2024,doi:10.1016/j.aej.2024.06.036.

[10] I. M. El-Hasnony, O. M. Elzeki, A. Alshehri, and H. Salem, “Multi-Label Active Learning-Based Machine Learning Model for Heart Disease Prediction,” *Sensors*, vol. 22, no. 3, p. 1184, Feb.2022, doi:10.3390/s22031184.