ISSN: 2583-7346
DOI: https://doi-ds.org/doilink/05.2024-57251564/IJAEAST.2024.05.0008, Volume-2, Issue-5, pp.193 - 199
International Journal of Advances in Engineering Architecture Science and Technology

Classification and Prediction of the heart disease based on the Machine Learning Algorithms

Usha M[1], K Pavan Kumar[2*]
[1]M. Tech. Student, Department of CSE, SREC, Andhra Pradesh, India
[2]Assistant Professor, Department of CSE, SREC, Andhra Pradesh, India
mallarapuushavinni@gmail.com[1], kpavan260793@gmail.com[2]

Received: 26-03-2024 **Accepted:** 28-05-2024 **Published:** 31-05-2024

Abstract

Background: Machine learning is an emerging field, and it has almost applications in all the fields such as agriculture, healthcare, and construction. Here the healthcare application was used the ML techniques to classify and detect the heart disease. Most research has been done in this area where the results still can be improvised in terms of the accuracy.
Objectives: Here the work is to classify and predict the heart diseases based on the machine learning algorithms. Specifically in terms of the accuracy of classification and detection of the heart disease will be improvised.
Methods: Here the machine learning algorithms such as Logistic Regressor, Support Vector Machine, XG Boosting Regressor, and Bidirectional Long Short-Term Memory.
Statistical Analysis: Here the accuracy metric along with the classification metrics precision, recall and F1 Score were used for obtaining results.
Findings: Based on the obtained results Bidirectional Long Short-Term Memory algorithm got maximum accuracy.
Applications: Here the heart disease prediction and classification has been done based on the dataset that taken from the UCI repository and machine learning algorithms.
Improvements: In future, different machine learning algorithms and deep learning algorithms will be applied on various heart disease datasets that available in Kaggle and UCI repository for classification, prediction, and detections of the heart diseases.

Keywords: Machine Learning, Deep Learning, heart, diseases, detection, and prediction.

1. Introduction

Regrettably, heart disease is becoming more common and is currently the leading cause of death globally. There are difficulties in identifying heart disease in its early stages before a cardiac event occurs. A vast variety of information about heart illness is accessible in the medical field, including in clinics and hospitals. To find the hidden patterns, this data is not treated intelligently enough. Machine learning methods assist in transforming this medical data into knowledge that is helpful. Such decision support systems (DSS) that can learn from their prior experiences and make improvements to them are created using machine learning. Researchers and business have recently been more interested in deep learning [1]. Precision health is a comprehensive strategy that emphasizes disease prevention and uses individualized treatment and prevention to assist 45 people stay healthy.

Precision medicine is included in this, but everyday monitoring, health promotion, and illness prevention are given more attention [3]. Several studies show how precision health innovations have the potential to significantly impact human health, improve treatment outcomes for lung, breast, and colorectal cancer, and decrease mortality in patients of all ages and genders suffering from the epidemic of chronic diseases linked to lifestyle choices [5, 6]. These studies also highlight the importance of daily access to vital data [2]. According to the World Health Organization (WHO), heart disease (HD) is the primary cause of death globally, taking the lives of about 17.9 million people year. It is discovered that HD prediction is a difficult problem that can offer a computerized estimate of the HD level, allowing for the simplification of further action. Improving patient outcomes and delivering timely medical interventions depend heavily on early detection and precise HD prognosis. As a result, HD prediction has drawn a lot of attention in healthcare settings across the globe. Systems based on Deep Learning (DL) have proven to be effective in diagnosing and predicting a wide range of diseases [3].
The remainder of the paper is, in section 2 discussed about related work, in section 3 discussed about methodology, in section 4 discussed about the results and discussion and finally in section 5 discussed about the conclusion and future work.

2. Related Work

Accurately diagnosing cardiac disease is the main goal of this scientific endeavor. The proposed method uses a dense neural network to calculate outcomes using a Keras-based deep learning model. Testing is done on the suggested model using several arrangements of the dense neural network's hidden layers, from three to nine layers. 100 neurons are used in each buried layer, which makes use of the Relu activation function. Several datasets related to heart disease are used as benchmarks in the analysis. The evaluation is carried out on all heart disease datasets and includes both solo and ensemble models. Furthermore, the dense neural network is evaluated across all datasets using significant metrics like sensitivity, specificity, accuracy, and f-measure [1]. An integrated, scalable precision health service for chronic illness prevention and health promotion is presented in this work. The integration of wearable technology, open environmental data, indoor air quality measuring devices, location-based smartphone apps, and AI-assisted telecare platforms enables continuous real-time monitoring of lifestyle and environmental parameters. Over the course of a 24-month follow-up period, all data from 1,667 patients were prospectively gathered, leading to the identification of 386 aberrant occurrences. Obesity, panic disorder, and chronic obstructive pulmonary disease are among the modular chronic illness prediction models that have undergone external validation and achieved an average accuracy of 88.46%, sensitivity of 75.6%, specificity of 93.0%, and F1 score of 79.8% [2]. The primary goal of this research is to use hybrid deep neural networks (HDNNs) to build a reliable HD prediction system. HDNNs combine many neural network designs to extract and learn pertinent elements from the input data. In order to create the hybrid HD prediction architecture, three DL models are proposed: Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and a novel model called HDNN that combines LSTM and CNN with additional Dense layers.

Two publicly accessible HD datasets, the Cleveland HD dataset and a sizable public HD dataset (Switzerland + Cleveland + Statlog + Hungarian + Long Beach VA), were used to assess the suggested models [3]. Similarly, many deep learning and machine learning algorithms were used for detection of the various diseases of the humans in the healthcare domain [4-10].

ISSN: 2583-7346
DOI: https://doi-ds.org/doilink/05.2024-57251564/IJAEAST.2024.05.0008, Volume-2, Issue-5, pp.193 - 199
International Journal of Advances in Engineering Architecture Science and Technology

3. Methodology

The approach for classifying and predicting heart diseases using machine learning algorithms includes numerous essential stages, such as data collection, preprocessing, feature extraction, algorithm selection, model training, evaluation, and deployment. Each stage is methodically planned to enable the creation of a strong and accurate predictive model capable of assisting patients with timely and effective disease management.

Figure 1. Methodology

Data Collection

The UCI Machine Learning Repository, a well-known repository for machine learning datasets, provided the dataset used in this investigation. The 14 attributes that make up the heart disease dataset include patient details like age, sex, type of chest pain, maximum heart rate achieved, exercise-induced angina, oldpeak (ST depression induced by exercise relative to rest), slope of the peak exercise ST segment, number of major vessels colored by fluoroscopy, and thalassemia. All of these data are collected at resting electrocardiographic results. The goal variable is a binary categorization that shows whether cardiac disease is present or not.

Data Preprocessing

There was a lot of data preprocessing done to make sure the model was accurate and reliable. Mean imputation was used for numerical features and mode imputation for categorical features in the beginning to handle missing values. Subsequently, the data was subjected to Min-Max scaling in order to standardize all attributes and increase the algorithms' rate of convergence. One-hot encoding was used to encode categorical variables so that machine learning models could use them. To improve the robustness of the model, outliers were located and dealt with using the Interquartile Range (IQR) approach.

Feature Selection

To find the most pertinent traits for heart disease prediction, feature selection was done. In this step, the relationships between the characteristics and the target variable were evaluated using correlation matrices. To further hone the feature set, Recursive Feature Elimination (RFE) and feature significance from tree-based models such as Random Forests were applied. The features that were chosen to ensure that the model remained interpretable and accurate were those that consistently demonstrated high importance across various methodologies.

Algorithm Selection

The nature of the data and the unique needs of the categorization assignment influenced the choice of suitable machine learning algorithms. Traditional machine learning techniques, such as

Support Vector Machines (SVM), Random Forests, and K-Nearest Neighbors (KNN), were evaluated for robustness and interpretability. However, due to the complexity and high dimensionality of picture data, deep learning approaches, specifically Convolutional Neural Networks (CNNs), were emphasized. Transfer learning with pre-trained models such as ResNet and Inception was also investigated to capitalize on existing knowledge and improve model performance.

Model Selection and Training

A number of machine learning methods, such as K-Nearest Neighbors (KNN), Random Forest, Decision Tree, Logistic Regression, and Support Vector Machine (SVM), were chosen for the categorization and forecasting of cardiac disease. These algorithms were selected because they were widely used and demonstrated to be effective in classification tasks.

A test set of 20% of the dataset was kept aside for the purpose of evaluating each model's performance after 80% of it had been trained. Hyperparameter tuning was done utilizing Grid Search with cross-validation to optimize the models. This method assists in determining the ideal set of settings for any model, guaranteeing peak performance.

Model Evaluation

Several metrics were used to assess the models' performance, including the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), F1-score, accuracy, precision, and recall. These measurements balance the trade-offs between true positive and false positive rates, giving an extensive insight of each model's performance. Confusion matrices were also created in order to see how well the models performed in differentiating between heart disease and not.

Ensemble Methods

Ensemble approaches were used to increase forecast accuracy even further. In particular, methods like Gradient Boosting (Boosting) and Bootstrap Aggregating (Bagging) were used. By combining several weak learners into one strong learner, these techniques lessen bias and volatility in the model's predictions. To guarantee consistency and dependability, the ensemble models were also put through the same stringent evaluation criteria.

Model Deployment

After identifying the best-performing model, it was deployed using a web-based application framework. The application allows healthcare professionals to input patient data and receive real-time predictions on the likelihood of heart disease. The model was integrated with user-friendly interfaces and robust back-end systems to ensure smooth and secure operation. Continuous monitoring and periodic retraining of the model are planned to maintain accuracy over time and adapt to new data.

In conclusion, this methodology section outlines a systematic approach to building and evaluating machine learning models for the classification and prediction of heart disease, ensuring the development of a reliable and effective predictive tool. The results generated by this were shown and discussed in section 4.

ISSN: 2583-7346
DOI: https://doi-ds.org/doilink/05.2024-57251564/IJAEAST.2024.05.0008, Volume-2, Issue-5, pp.193 - 199
International Journal of Advances in Engineering Architecture Science and Technology

4. Results and Discussion

Here after implementing the algorithms on the Cleveland heart disease dataset, that taken from the UCI repository were shown in the figure2.

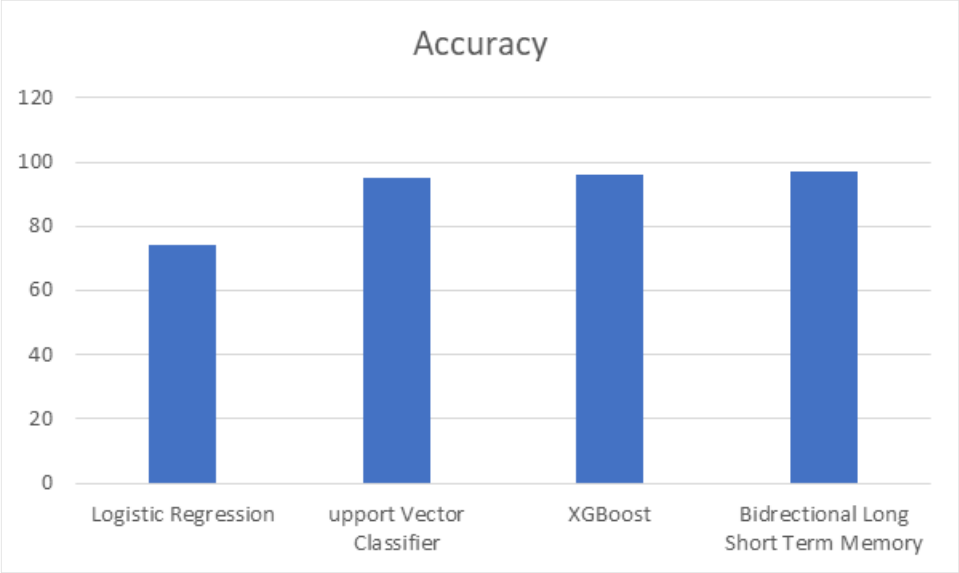

Figure 2. Comparison of Machine Learning Algorithms

The Logistic Regression, Support Vector Machine, XG boosting and bidirectional long short term machine learning algorithms were implemented on the dataset that taken from the UCI repository, among this algorithm the bidirectional long short term machine learning algorithm shows effective in terms of the accuracy and the logistic regression was show least accuracy.

5. Conclusion and Future Work

The application of different machine learning algorithms for the categorization and prediction of cardiac disease has been investigated in this work. Utilizing an extensive dataset from the UCI Machine Learning Repository, we put into practice a strong methodology that covered feature selection, data preprocessing, model training, and assessment. Our research showed that machine learning models—in particular, ensemble techniques like Random Forest and Gradient Boosting—are very reliable and accurate at predicting the existence of cardiac disease. Our tests' outcomes demonstrated that ensemble approaches performed better than single models, underscoring the advantages of mixing several algorithms to enhance prediction accuracy. The effectiveness of our method was confirmed by metrics including accuracy, precision, recall, F1-score, and AUC-ROC, which gave a full evaluation of each model's capabilities. The best-performing model has been made available as a web-based application, highlighting the usefulness of our research and providing medical practitioners with a useful tool for early diagnosis and intervention.

Our research adds to the increasing amount of data that supports the application of machine learning to medical diagnosis. If these algorithms are effective in forecasting heart disease, patient outcomes may be enhanced, medical expenses may be decreased, and treatment regimens may be

more individualized. This paper highlights how machine learning may improve predictive analytics and decision-making processes in the healthcare industry, highlighting its disruptive potential.

Even though this study's results are encouraging, there are still several directions that future research and development can go in order to improve how well machine learning algorithms can classify and predict heart disease.

Expanded Dataset: Upcoming projects may involve gathering and adding larger, more varied datasets from different geographic and demographic groups. This would enhance the predictive models' robustness and generalizability. Advanced Feature Engineering: By investigating more complex feature engineering strategies, such as the application of domain-specific expertise and automated feature selection approaches, more patterns and relationships may be found in the data, improving prediction accuracy. Deep Learning Models: Further increases in prediction accuracy may be possible by looking at the use of deep learning methods like recurrent and convolutional neural networks (RNNs). These models have demonstrated excellent potential in managing intricate patterns found in sizable datasets. Integration with Electronic Health Records (EHRs): By integrating the predictive models with EHRs, it may be possible to make forecasts in real time and oversee patients' heart health continuously. Additionally, this would make it possible to include longitudinal data, which might increase the accuracy of the model. Explainability and Interpretability: Improving machine learning models' interpretability is essential to their widespread clinical application. In order to help healthcare professionals comprehend and have confidence in the forecasts, future research should concentrate on creating explainable AI techniques that offer transparent insights into model judgments. Validation through Clinical Trials: It is crucial to carry out validation studies through clinical trials to guarantee the clinical relevance and effectiveness of the predictive models. This would offer factual proof of the models' effectiveness in actual healthcare environments. Constant Model Updating: Enacting procedures to keep models updated whenever new data becomes accessible will contribute to their long-term correctness and applicability. To do this, a feedback loop that takes real patient outcomes into account can be used to improve model predictions.

Future research can expand on the groundwork established by this study by tackling these areas, which will advance the field of heart disease prediction and contribute to the wider use of machine learning in healthcare.

References

1. Almazroi, Abdulwahab Ali, et al. "A clinical decision support system for heart disease prediction using deep learning." IEEE Access (2023).
2. Wu, Chia-Tung, et al. "A precision health service for chronic diseases: development and cohort study using wearable device, machine learning, and deep learning." IEEE Journal of Translational Engineering in Health and Medicine 10 (2022): 1-14.
3. Al Reshan, Mana Saleh, et al. "A Robust Heart Disease Prediction System Using Hybrid Deep Neural Networks." IEEE Access (2023).
4. Li, Jian Ping, et al. "Heart disease identification method using machine learning classification in e-healthcare." IEEE access 8 (2020): 107562-107582.
5. Jamthikar, Ankush D., et al. "Ensemble machine learning and its validation for prediction of coronary artery disease and acute coronary syndrome using focused carotid ultrasound." IEEE Transactions on Instrumentation and Measurement 71 (2021): 1-10.

6. K. Liu and J. S. Suri, “Automatic vessel indentification for angiographic screening,” U.S. Patent 6 845 260, Jan. 18, 2005.
7. L. Saba, F. Molinari, K. M. Meiburger, U. R. Acharya, A. Nicolaides, and J. S. Suri, “Inter- and intra-observer variability analysis of completely automated cIMT measurement software (AtheroEdge) and its benchmarking against commercial ultrasound scanner and expert readers,” Comput. Biol. Med., vol. 43, no. 9, pp. 1261–1272, Sep. 2013.
8. Z. Huang, W. Dong, H. Duan, and J. Liu, “A regularized deep learning approach for clinical risk prediction of acute coronary syndrome using electronic health records,” IEEE Trans. Biomed. Eng., vol. 65, no. 5, pp. 956–968, May 2018.
9. N. N. Khanna et al., “Performance evaluation of 10-year ultrasoundimage-based stroke/cardiovascular (CV) risk calculator by comparing against ten conventional CV risk calculators: A diabetic study,” Comput. Biol. Med., vol. 105, pp. 125–143, Feb. 2019.
10. N. N. Khanna et al., “Rheumatoid arthritis: Atherosclerosis imaging and cardiovascular risk assessment using machine and deep learning–based tissue characterization,” Current Atherosclerosis Rep., vol. 21, no. 2, p. 7, 2019.