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 Blockage Detection Using Machine Learning

[1]Amarjeet Kumar, [2]Ujjwal Rai

[1]Kushwant Kaur

Department of Computer Science and Engineering, Department of Computer Science and Engineering
Chandigarh University, Mohali, Punjab-140301

Abstract: This research paper presents a machine learning-based predictive model for the detection of heart blockage. Heart blockage is a significant cardiovascular condition that requires timely and accurate diagnosis for effective management. Leveraging machine learning techniques, specifically Logistic Regression, this study aims to contribute to the advancement of heart blockage detection. The trained logistic regression model exhibits promising performance, achieving an 85% accuracy on the training dataset and approximately 82% accuracy on the test dataset. By utilizing a diverse dataset of cardiovascular imaging data, this research demonstrates the potential of machine learning in augmenting traditional diagnostic methods for heart blockage detection. The findings underscore the importance of integrating machine learning into cardiovascular diagnostics, thereby enhancing patient care and healthcare system efficiency. As medical data availability continues to grow and machine learning techniques evolve, this research contributes to ongoing efforts to improve heart blockage diagnosis and management.

Keywords- Logistic Regression, Machine Learning, Heart Disease, Accuracy

I. INTRODUCTION

Heart blockage, a prevalent cardiovascular condition with potentially severe consequences, remains a significant concern in public health, posing challenges to healthcare systems worldwide. Timely and accurate detection of heart blockage is crucial for effective management and improving patient outcomes. Traditionally, the diagnosis of heart blockage has heavily relied on clinical assessments and imaging modalities interpreted by medical professionals. However, these methods are susceptible to variability among observers and can consume considerable time. This research endeavors to explore the application of machine learning, specifically Logistic Regression, in predictive modeling for heart blockage detection. While machine learning techniques have shown promise in various medical domains, their application in cardiovascular diagnosis is relatively novel. The logistic regression model trained in this study exhibits an 85% accuracy on the training dataset and approximately 82% accuracy on the test dataset, indicating its potential utility in heart blockage detection. In recent years, the intersection of machine learning and healthcare has witnessed remarkable advancements, offering opportunities to enhance diagnostic accuracy and efficiency. By leveraging machine learning algorithms, this study aims to develop a robust and reliable system for heart blockage detection, contributing to improved patient care and healthcare system effectiveness. The primary objectives of this research are as follows: To develop a machine learning-based heart blockage detection system using Logistic Regression. To assess the performance of the developed model on a diverse dataset of cardiovascular imaging data. To compare the diagnostic accuracy of the model with traditional diagnostic methods and expert interpretations. To demonstrate the potential of machine learning techniques in augmenting healthcare professionals' diagnostic capabilities for heart blockage detection. This research paper is structured as follows: Section II provides a comprehensive review of existing literature on machine learning applications in cardiovascular diagnosis, with a focus on heart blockage detection. Section III outlines the methodology, including dataset description, model architecture, training process, and evaluation metrics. In Section IV, the results of model experiments and comparisons with baseline methods are presented. Section V discusses the implications of the findings and outlines future research directions. Finally, Section VI concludes the paper, by summarizing key insights and highlighting the significance of machine learning in advancing cardiovascular healthcare. With the ever-increasing availability of medical data and the rapid advancement of machine learning techniques, this research aims to contribute to the ongoing efforts to improve heart blockage detection, ultimately enhancing patient care and healthcare system efficiency.

II. LITERATURE REVIEW

Li, J. P., Haq, A. U., Din, S. U., Khan, J., Khan, A., & Saboor, A. (2020) [1]: The article presents an advanced heart disease diagnosis system using machine learning and novel feature selection methods. Through algorithms like SVM and FCMIM, it enhances accuracy and reduces execution time, showing promise for efficient implementation in healthcare for timely disease identification.

Nagavelli, U., Samanta, D., & Chakraborty, P. (2022) [2]: This paper addresses heart failure disease detection using various machine-learning approaches. It explores Naïve Bayes for predicting heart disease, analyzes ischemic heart disease localization with SVM and XGBoost, introduces an improved SVM for heart failure identification, and presents a comprehensive heart disease prediction model utilizing DBSCAN, SMOTE-ENN, and XGBoost. The study aims to provide clinicians with an effective tool for early heart problem diagnosis.

Taylor, O. E., Ezekiel, P. S., & Deedam-Okuchaba, F. B. (2019) [3]: This paper outlines a heart disease detection model employing machine learning algorithms and Agile Methodology. Four classifiers were trained on a Heart Dataset, with the Decision Tree Classifier yielding the best accuracy (98.83%). The model, implemented in Python with Flask, takes user inputs for predictions, demonstrating promising results for early heart disease detection.

Miao, K. H., Miao, J. H., & Miao, G. J. (2016) [4]: This research focuses on enhancing coronary heart disease diagnosis through an advanced ensemble machine learning approach, employing an adaptive Boosting algorithm. The models, tested on datasets from various sources, demonstrated high accuracies, surpassing previous research results. The developed ensemble learning models offer reliable and clinically valuable diagnoses, particularly beneficial for patients in developing regions with limited access to heart disease specialists, potentially saving numerous lives globally.

Gonsalves, A. H., Thabtah, F., Mohammad, R. M. A., & Singh, G. (2019, July) [5]: This paper presents the development of eight statistical and machine-learning models for predicting the mortality of hospital patients with pneumonia based on their initial presentation. These models were created using data from 9847 patient cases and tested on 4352 additional cases. The primary evaluation metric assessed each model's error in predicting survival when considering different fractions of patients surviving (between 0.1 and 0.6). Results showed that all models had similar error rates, particularly when predicting around 30% survival. Differences in the models were more related to their complexity than their performance, indicating potential for future implementation in clinical guidelines.

Ed-Daoudy, A., & Maalmi, K. (2019, April) [6]: This paper addresses the pressing need for early heart disease detection by proposing a real-time prediction system using Apache Spark. Leveraging streaming big data analytics and machine learning, the system combines Spark MLlib and Apache Cassandra for efficient data processing, classification, storage, and visualization, offering a powerful and cost effective solution.

Lutimath, N. M., Chethan, C., & Pol, B. S. (2019) [7]: This paper explores the application of machine learning, particularly Naïve Bayes Classification and support vector machines, in detecting heart diseases using the UCI machine learning repository dataset. Focusing on coronary heart disorder, the study utilizes R language for implementation and aims to predict the classification accuracy of patients suffering from heart disease, showcasing the significance of machine learning in healthcare.

Kukar, M., Kononenko, I., Grošelj, C., Kralj, K., & Fettich, J. (1999) [8]: This study addresses the importance of improving diagnostic procedures for Ischaemic heart disease. Utilizing machine learning methods, experiments with various algorithms achieved performance comparable to clinicians, with enhanced sensitivity and specificity demonstrated through ROC analysis. The research highlights the potential of machine learning in increasing diagnostic accuracy for heart disease.

Choudhary, G., & Singh, S. N. (2020, October) [9]: This study addresses the challenging task of heart disease diagnosis using machine learning algorithms. Analyzing a vast dataset, the proposed work focuses on identifying crucial features for an effective diagnostic system. Employing decision trees and Ada-Boost algorithms, the research aims to assist doctors in diagnosing heart patients accurately, emphasizing the importance of feature reduction for enhanced classifier performance.

Adhikari, N. C. D. (2018). [10]: Researchers analyze Indian patient data to build a predictive model for heart attack probability. This model aims to aid doctors in treatment decisions, fostering transparency with patients. Metrics like True Positive Rate, False-Negative Rate, and AUC-ROC are crucial for validation, besides accuracy, in the model's development.

Kaur, B., & Kaur, G. (2022, September).[11]: With the rising prevalence of heart disease worldwide, early diagnosis is crucial to prevent fatalities. Technology-based software leveraging data mining and machine learning techniques aids in efficient heart disease diagnosis. This research employs modified machine learning algorithms, including random forest, K-means, genetic algorithm, and logistic regression, to predict coronary heart disorder with up to 95% accuracy, enhancing diagnostic capabilities in the medical field.

Polat, K., Şahan, S., & Güneş, S. (2007) [12]: This study utilizes machine learning for heart disease diagnosis, employing a new weighting scheme based on the k-nearest neighbor method for preprocessing. The Artificial Immune Recognition System (AIRS) with fuzzy resource allocation serves as the classifier, achieving a promising 87% classification accuracy using the UCI Machine Learning Database.

Loh, B. C., & Then, P. H. (2017). [13]: This review addresses the challenge of diagnosing and treating cardiovascular diseases in developing nations and rural areas. It highlights the potential of telemedicine and mHealth, combined with machine and deep learning, to enhance heart disease management, especially in underserved regions.

Aggarwal, R., & Kumar, S. (2022, November).[14]: Detecting heart disease early is challenging for physicians due to various patient attributes. Machine learning aids in early detection by analyzing factors like cholesterol levels and heart rate. This study employs multiple classifiers on a dataset from the UCI ML repository, achieving a high accuracy of 98.08% to 98.25% using MLPPCA and CHI square methods, facilitating early disease detection in medical science.

Kora, P., Abraham, A., & Meenakshi, K. (2020). [15]: The integration of Bacterial-foraging-optimization (BFO) and Particle Swarm Optimization (PSO) improves the detection of abnormal cardiac beats. This hybrid approach, BFPSO, combined with support vector machines (SVMs) and Wavelet Transform, enhances accuracy and speed in identifying cardiac arrhythmia. Results indicate significant improvements, achieving 98.9% to 99.3% accuracy in detecting specific cardiac conditions.

III. RESEARCH METHODOLOGY

Decision Tree Classifier:

A decision tree classifier is a flowchart-like structure where each internal node represents a decision based on a feature, and each leaf node represents a class label. The algorithm recursively splits the data based on different features, aiming to create partitions that separate the classes as much as possible. The splits are made based on criteria like entropy or Gini impurity to minimize uncertainty and maximize information gain. Decision trees are easy to understand and interpret, and they can handle both categorical and numerical data.

Naive Bayes:

Naive Bayes is a probabilistic algorithm utilized in machine learning to predict the likelihood of heart blockage based on various features associated with a patient's medical data. This algorithm leverages Bayes' Theorem, a fundamental concept in probability theory, to calculate the probability of heart blockage given a set of observed features.
Bayes' Theorem is expressed as:

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)}$$

Logistic regression:

Logistic Regression is a binary classification algorithm in machine learning. It predicts the probability of an instance belonging to a certain class (e.g. Healthy Heart or Defective Heart) based on input features. It uses the logistic function to map output values between 0 and 1, making it suitable for binary classification tasks.

Random classifier:

A Random Classifier, also known as a Random Baseline, is the simplest form of classification model in machine learning. It randomly assigns class labels to instances without any learning or pattern recognition. It serves as a baseline to compare the performance of more sophisticated models.

IV. DATA AND SOURCE OF DATA

For this study, secondary data has been collected. From the website of Kaggle from the heart blockage dedication. The dataset used in this study comprises 303 instances with 14 features each. Each instance represents a patient, and the features include various clinical and demographic variables. The target variable indicates the presence or absence of heart blockage, where a value of 1 signifies the presence of heart blockage (defective) and 0 indicates normal heart function

V. PROCESS FLOW:

1. Start
2. Data Collection / Data Preprocessing
3. Classification Data
   a. Training Data
   b. Test Data
4. Classification Techniques
   - Decision Tree classifier
   - Naïve Bayes
   - Logistic Regression
   - Random Classifier

5. Test The Model
6. Display Result
7. End

VI. RESULT

The logistic regression model developed for heart blockage detection demonstrated promising performance, achieving an accuracy rate of 85% on the training dataset and approximately 82% on the test dataset. This indicates the model's ability to effectively distinguish between instances of heart blockage and normal heart function. While accuracy serves as a fundamental metric for evaluating model performance, additional evaluation metrics such as precision, recall, and area under the receiver operating characteristic curve (AUCROC) were also considered to provide a comprehensive assessment. Precision measures the proportion of true positive predictions among all positive predictions made by the model. Recall, on the other hand, quantifies the proportion of true positive predictions identified correctly out of all actual positive instances in the dataset. These metrics help gauge the model's ability to accurately identify instances of heart blockage while minimizing false positives. Furthermore, the area under the receiver operating characteristic curve (AUCROC) provides insight into the model's ability to discriminate between positive and negative instances across various threshold values. A higher AUCROC value indicates superior discrimination ability, suggesting better overall model performance. The logistic regression model's performance on these evaluation metrics contributes to its robustness and reliability for heart blockage detection. These findings support the potential application of machine learning techniques in clinical settings to assist healthcare professionals in accurately diagnosing heart blockage, thereby improving patient care and healthcare system efficiency.

VII. CONCLUSION

Following the implementation and evaluation of four distinct machine learning algorithms—Logistic Regression, Decision Tree, Naïve Bayes, and Random Classifier for heart blockage detection, it was discerned that Logistic Regression exhibited the highest efficiency among them. With an accuracy rate of 82.89% on the test dataset, Logistic Regression outperformed its counterparts, including Decision Tree and Naïve Bayes, which achieved accuracies of 80.43% each, while Random Classifier attained an accuracy of 81.57%.

Subsequently, the selected Logistic Regression model, attaining an accuracy rate of 85% on the training dataset and approximately 82% on the test dataset, was further developed into an interactive web application using HTML, CSS, and the Django framework of Python. This user-friendly interface enables individuals to obtain preliminary predictions regarding their heart health. Given the alarming prevalence of heart diseases in India and globally, the integration of machine learning technology for early heart disease prediction holds immense potential to significantly impact society.

The achieved accuracy of the Logistic Regression model underscores its efficacy in distinguishing between instances of heart blockage and normal heart function. This model, integrated into a robust web application, serves as a promising tool for preemptive heart disease assessment, potentially mitigating the detrimental effects of heart-related ailments on public health.

VIII. FUTURE WORK

In conclusion, the utilization of machine learning, specifically Logistic Regression, for heart blockage detection presents promising outcomes. The model, trained on a diverse dataset of cardiovascular imaging data, exhibits commendable accuracy rates, indicating its potential to assist in early disease identification. Future work in this area involves several avenues for improvement and refinement. Firstly, continuous optimization of the model is essential to enhance its performance further. This includes fine-tuning hyperparameters, exploring different feature representations, and incorporating advanced techniques to boost accuracy and robustness. Additionally, the inclusion of additional evaluation metrics beyond accuracy, such as precision, recall, and area under the receiver operating characteristic curve (AUCROC), would provide a more comprehensive assessment of the model's performance. These metrics can offer insights into the model's sensitivity and specificity, aiding in minimizing diagnostic errors and improving reliability. Collaboration with healthcare professionals is paramount for validating the model's efficacy in real-world clinical settings. Their feedback and input can help refine the model and ensure its alignment with clinical practices and standards. Moreover, ethical considerations and adherence to regulatory standards are imperative throughout the development process. Ensuring patient privacy, data security, and compliance with healthcare regulations are essential aspects that require careful attention. Continuous monitoring and adaptation to advancements in the field are crucial for maintaining and enhancing the model's accuracy and relevance over time. Staying abreast of emerging technologies, research findings, and clinical insights will facilitate ongoing improvements in heart blockage detection accuracy and patient care.

REFERENCES

[1] Li, Jian Ping, et al. "Heart disease identification method using machine learning classification in e-healthcare." IEEE access 8 (2020): 107562-107582.

[2] Nagavelli, Umarani, Debabrata Samanta, and Partha Chakraborty. "Machine learning technology-based heart disease detection models." Journal of Healthcare Engineering 2022 (2022).

[3] Taylor, O. E., P. S. Ezekiel, and F. B. Deedam-Okuchaba. "A model to detect heart disease using machine learning algorithm." International Journal of Computer Sciences and Engineering 7.11 (2019): 1-5.

[4] Miao, Kathleen H., Julia H. Miao, and George J. Miao. "Diagnosing coronary heart disease using ensemble machine learning." International Journal of Advanced Computer Science and Applications 7.10 (2016).

[5] 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.

[6] Ed-Daoudy, Abderrahmane, and Khalil Maalmi. "Real-time machine learning for early detection of heart disease using big data approach." 2019 international conference on wireless technologies, embedded and intelligent systems (WITS). IEEE, 2019.

[7] Lutimath, Nagaraj M., C. Chethan, and Basavaraj S. Pol. "Prediction of heart disease using machine learning." International journal Of Recent Technology and Engineering 8.2S10 (2019): 474-477.

[8] Kukar, Matjaž, et al. "Analysing and improving the diagnosis of ischaemic heart disease with machine learning." Artificial intelligence in medicine 16.1 (1999): 25-50.

[9] Choudhary, Garima, and Shailendra Narayan Singh. "Prediction of heart disease using machine learning algorithms." 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE). IEEE, 2020.

[10] Adhikari, N. C. D. (2018). Prevention of heart problem using artificial intelligence. International Journal of Artificial Intelligence and Applications (IJAIA), 9(2).

[11] Kaur, B., & Kaur, G. (2022, September). Heart disease prediction using modified machine learning algorithm. In International Conference on Innovative Computing and Communications: Proceedings of ICICC 2022, Volume 1 (pp. 189-201). Singapore: Springer Nature Singapore.

[12] Polat, K., Şahan, S., & Güneş, S. (2007). Automatic detection of heart disease using an artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism and k-nn (nearest neighbour) based weighting preprocessing. Expert Systems with Applications, 32(2), 625-631.

[13] Loh, B. C., & Then, P. H. (2017). Deep learning for cardiac computer-aided diagnosis: benefits, issues & solutions. Mhealth, 3.

[14] Aggarwal, R., & Kumar, S. (2022, November). MLPPCA: Heart Disease Detection using Machine learning. In 2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC) (pp. 457-461). IEEE.

[15] Kora, P., Abraham, A., & Meenakshi, K. (2020). Heart disease detection using hybrid of bacterial foraging and particle swarm optimization. Evolving Systems, 11(1), 15-28.