International Journal of
Biomedical Innovations and Engineering

Volume 1, Issue 1, 2023
January–June
DOI (Journal): 10.37591/IJBIE

https://journals.stmjournals.com/ijbie

Research IJBIE

Heart Attack Prediction Using Machine Learning

Sumit Halder[1],*, Amit Chakraborty[1], Saswati Naskar[2], Saibal Majumder[2]

Abstract

Heart attacks have become a prevalent and serious condition in recent years due to a variety of causes. Numerous variables, including age, sex, fat, and others, can be used to predict it. In the current study, it was found that a data set with 13 parameters and 302 distinct data values, collected from a Kaggle dataset to assess patient condition, was covered. This article delves into the application of machine learning and artificial intelligence algorithms for the prevention of heart disease. The primary objective of this study is to enhance the accuracy and precision of identifying cardiac ailments by utilizing advanced algorithms. These algorithms play a pivotal role in discerning the presence or absence of cardiac issues within a given case. By harnessing the power of these technologies, the article aims to revolutionize cardiac complaint detection, ultimately contributing to early intervention and improved patient outcomes in the realm of heart disease management.

Keywords: Artificial intelligence, machine learning, regression, support vector machines (SVM), logistic regression

INTRODUCTION

Today, a significant proportion of the global population finds itself grappling with the daunting challenges posed by heart disease, cholesterol-related problems, and arterial blockages. The gravity of this issue is becoming increasingly evident, as human beings are unable to accurately anticipate such threats on their own. However, amidst these concerns, the emergence of machine learning (ML) as a transformative technology has paved the way for potential solutions. ML has already proven its prowess across various industries, and it can undoubtedly lend its power to tackle this formidable health issue as well. By harnessing the capabilities of ML, we can develop a predictive model capable of discerning the likelihood of an individual experiencing a heart attack. This endeavor involves training the ML model using pertinent data, enabling it to learn patterns and correlations within the dataset. Notably, algorithms such as linear regression and support vector machines (SVM) can be employed to train these models effectively. By leveraging the predictive capabilities of ML, we can potentially revolutionize the field of cardiovascular health and enhance our ability to proactively address the risk factors associated with heart disease.

LITERATURE REVIEW

Heart disease prediction systems have made significant progress in recent years by incorporating various advanced machine learning algorithms. These algorithms, such as simple regression systems, SVM, random forest techniques, and Naïve Bayesian methods, have been employed to combat heart disease effectively [1]. Among these methods, the Heart Disease Prediction System (HDPS) has shown remarkable

Author for Correspondence
Sumit Halder
E-mail: sumithalderr2021@gmail.com

[1]Student, Department of Computer Science and Engineering, Greater Kolkata College of Engineering and Management (GKCEM), South 24 Parganas, West Bengal, India
[2]Assistant Professor, Department of Computer Science and Engineering, Greater Kolkata College of Engineering and Management, South 24 Parganas, West Bengal, India

Received Date: July 13, 2023
Accepted Date: July 30, 2023
Published Date: August 10, 2023

Citation: Sumit Halder, Amit Chakraborty, Saswati Naskar, Saibal Majumder. Heart Attack Prediction Using Machine Learning. International Journal of Biomedical Innovations and Engineering. 2023; 1(1): 8–13p.

accuracy when compared to representative documents and SVM, setting a new standard for predictive models in this domain [2].

In the pursuit of more sophisticated cardiac complaint predictions, researchers have explored modern data mining techniques. A comprehensive study compared the performance of several algorithms, including c4.5, K-means, decision trees, SVM, and Naïve Bayes, along with other contemporary machine learning methods [3]. These techniques have played a crucial role in enhancing the accuracy and efficiency of cardiac illness outcome predictions.

Another research endeavor focused on developing intelligent systems for reflections on cardiovascular disease using the K-Star algorithm [4]. This study utilized the Learning Vector Quantification Neural System Calculation to forecast heart infections. The neural system in this structure employed color-chasing techniques to recognize 13 clinical characteristics, enabling accurate predictions even in cases with proximal coronary symptoms.

Overall, the integration of diverse machine learning algorithms and innovative approaches has revolutionized heart disease prediction systems. These advancements have not only improved the accuracy of diagnoses but also elevated the sophistication of decision tree literacy systems, surpassing previous methodologies by an impressive 98.85 percent [2]. Consequently, these developments hold great promise for the future of cardiac health monitoring and treatment.

METHODOLOGY AND ANALYSIS

The methodology for this heart attack prediction and analysis project involved the utilization of a dataset obtained from Kaggle. This dataset consisted of records containing information about various patients, including 13 different characteristics. The availability of such a dataset proved to be a valuable resource, offering essential information and tools for conducting thorough research or designing effective predictive models.

To begin the data preprocessing phase, the dataset was divided into training and testing data. This division is crucial for training machine learning estimators and evaluating their performance accurately. Once the model was trained, the values in the test data were scaled to ensure uniformity and compatibility with the trained model. Scaling involved transforming the values to a range between 0 and 1, aiding in the interpretation and comparison of different features.

To gain insights into the relationships between various features, such as age, sex, and fat, a histogram was plotted to visualize the distribution of these variables Figure 1. This graphical representation helps in understanding the distribution patterns and identifying any potential correlations among the features. Additionally, a correlation matrix was generated, providing a quantitative measure of the relationships between different features Figure 2. The correlation matrix aids in identifying the strength and direction of correlations, highlighting variables that might have a significant impact on the prediction of heart attacks.

Overall, this project employed a systematic methodology, starting from data collection from Kaggle, followed by data preprocessing and exploratory data analysis techniques such as histograms and correlation matrices. These steps were essential in preparing the dataset for further analysis and building predictive models for heart attack prediction.

RESULT AND DISCUSSIONS

Logistic Regression

Logistic regression (LR) is a statistical technique used to estimate the parameters of a logistic model. It is commonly employed when dealing with a dependent variable that exhibits categorical characteristics, such as binary outcomes or multiple discrete classes. By analyzing the relationship between the independent variables and the probability of belonging to a particular category, LR aids in predicting the categorical outcome [5–10].

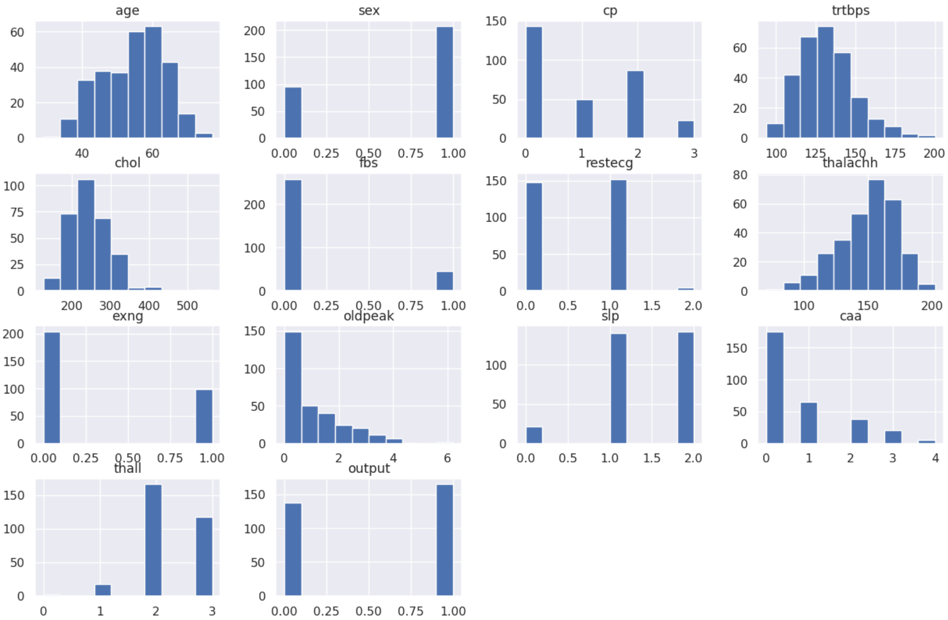

Figure 1. Histogram of the features.

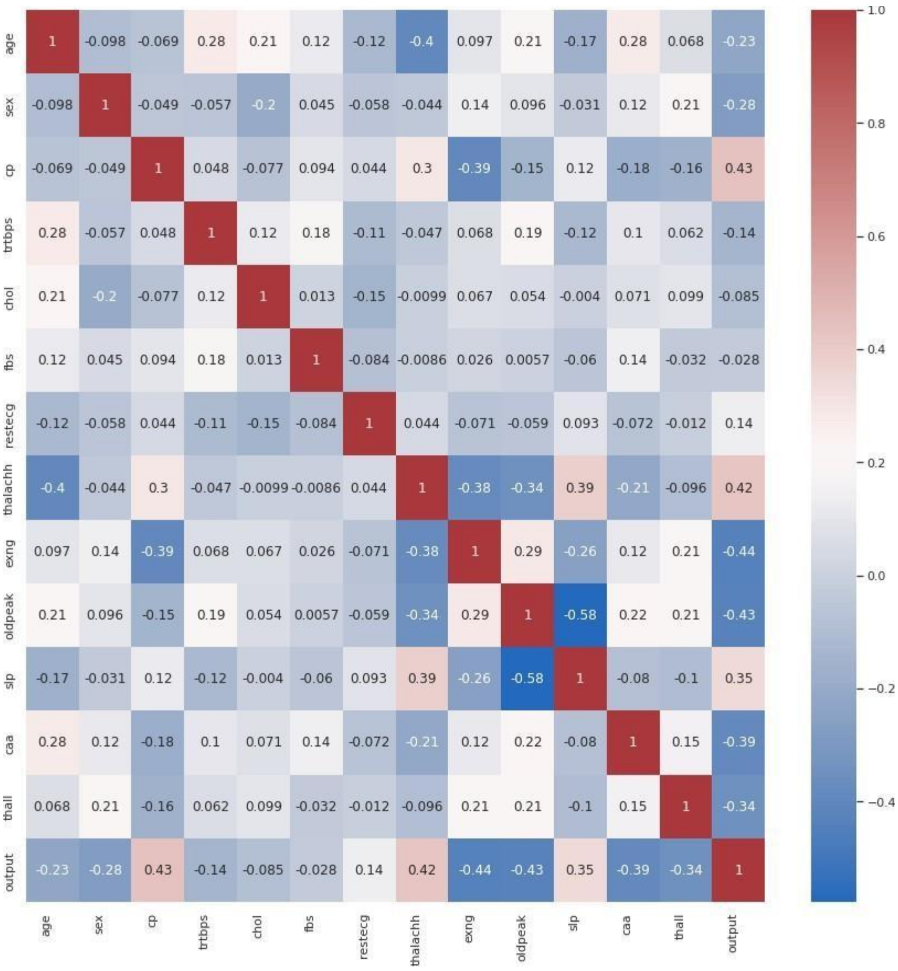

Figure 2. Heatmap of the correlation matrix.

In the case of LR, the model determines the optimal values for the parameters that maximize the likelihood of the observed data. These parameters represent the coefficients assigned to the independent variables, indicating the direction and strength of their influence on the outcome. By applying an appropriate algorithm, LR computes the probability of a given data point falling into a specific category Figure 3.

After training the logistic model, its accuracy can be assessed by comparing the predicted outcomes with the actual observed outcomes. In this instance, LR achieved an accuracy rate of 77%, indicating that the model correctly classified 77% of the instances [11].

Support Vector Machine

On the other hand, SVM is a powerful algorithm used for classification and regression tasks. SVM seeks to create a hyperplane in a high-dimensional feature space that effectively separates different classes of data points Figure 4. The goal is to maximize the margin, i.e., the distance between the hyperplane and the nearest data points of each class, which facilitates accurate classification.

By employing a kernel function, SVM can transform the input data into a higher-dimensional space, allowing for the creation of a hyperplane that effectively separates the classes. The accuracy of SVM is typically determined by evaluating its performance on a test dataset. In this context, SVM achieved an accuracy rate of 81%, indicating that it correctly classified 81% of the instances.

In summary, LR and SVM are both popular machine learning techniques used for classification tasks. While LR estimates the parameters of a logistic model to predict categorical outcomes, SVM constructs a hyperplane to classify different data points. In the specific examples mentioned, LR achieved an accuracy of 77%, whereas SVM attained a higher accuracy rate of 81%.

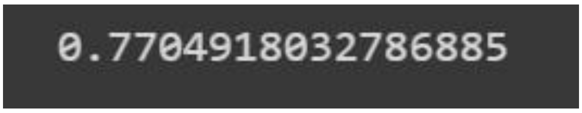

Figure 3. Result of logistic regression.

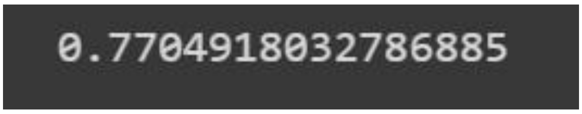

Figure 4. Result of support vector machine.

CONCLUSIONS

In the future, there is a growing concern that the number of patients suffering from heart problems will increase significantly due to unhealthy lifestyles. With such a rise in heart-related issues, the risk of cardiac arrest also escalates, emphasizing the need for early detection and intervention. This is where the application of machine learning can prove to be immensely valuable, potentially saving countless innocent lives. By leveraging machine learning techniques, it becomes possible to predict the likelihood of developing heart disease at an early stage. Such predictions can enable healthcare professionals to take proactive measures and provide appropriate treatments to individuals identified as being at high risk. This preventive approach holds great potential for reducing the occurrence of life-threatening cardiac events. Recent research endeavors have explored the effectiveness of various machine learning algorithms in predicting heart disease. Among these algorithms, the LR algorithm has emerged as one of the most efficient tools, exhibiting an impressive accuracy score of 77%. However, it is important to note that the accuracy of these algorithms is heavily reliant on the quality and representativeness of the dataset used for training and testing purposes. The success of machine learning models in accurately predicting heart disease hinges on the availability of comprehensive and diverse datasets that encompass a wide range of patient profiles, medical histories, and risk factors.

Such datasets enable the algorithms to learn and identify patterns, correlations, and indicators that may be indicative of impending heart problems. With access to reliable and extensive data, machine learning algorithms can be fine-tuned to achieve higher accuracy levels, allowing for more effective identification and early intervention. In conclusion, the future of healthcare and the battle against heart disease lies in the hands of machine learning. By leveraging powerful algorithms such as LR, healthcare professionals can obtain valuable insights and predictions, leading to proactive measures that can potentially save lives. However, the accuracy of these algorithms is intrinsically tied to the quality and diversity of the datasets used for training and testing, underscoring the importance of comprehensive data collection and analysis. Through continued advancements in machine learning and data-driven healthcare, we have the potential to make significant strides in preventing and mitigating heart problems in the years to come.

Acknowledgments

The authors would like to express their heartfelt appreciation to the esteemed organizers of the NCEATS 2023 for their valuable consideration of their manuscript for potential publication in the renowned journal. This acknowledgment is a sincere recognition of the organizers' meticulous efforts in reviewing and assessing the numerous submissions received for this prestigious conference. The authors extend their gratitude for the opportunity to present their research findings and contribute to the intellectual discourse within their field of study. The acknowledgment also serves as an acknowledgment of the organizers' commitment to promoting excellence in scientific research and fostering an environment that encourages scholarly exchange and collaboration. The authors are deeply grateful for the platform provided by NCEATS 2023, which has allowed them to share their insights and ideas with a distinguished audience. This acknowledgment reflects the authors' genuine admiration for the organizers' dedication to advancing knowledge and fostering innovation in the academic community.

REFERENCES

1. Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques IEEE/ACS International Conference on Computer Systems and Applications; 2008, Doha, Qatar. p. 108–15. doi: 10.1109/AICCSA.2008.4493524.
2. Nikan S, Gwadry-Sridhar F, Bauer M. Machine learning application to predict the risk of coronary artery atherosclerosis International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, USA, 2016; 2016. p. 34–9. doi: 10.1109/CSCI.2016.0014.
3. Bashir S, Qamar U, Younus Javed M. An ensemble based decision support framework for intelligent heart disease diagnosis. International Conference on Information Society (i- Society 2014), London, UK, 2014. p. 259–64. doi: 10.1109/i-Society.2014.7009056.
4. Ramalingam VV, Dandapath A, Karthik Raja MK. Heart disease prediction using machine learning techniques: a survey. Int J Eng Technol. 2018;7(3):684–7p. doi: 10.14419/ijet.v7i2.8.10557.
5. Chapman P, Clinton J, Kerber R, Khabeza T, Reinartz T, Shearer C, et al. CRISP-DM 1.0 Step-by-step data mining guide. SPSS, Copenhagen. 2000. p. 1–78.
6. Charly K. Data mining for the enterprise. Proceedings of the Thirty-First Hawaii International Conference on System Sciences, Kohala Coast, HI, USA, 1998, vol. 7, pp. 295–304. doi: 10.1109/HICSS.1998.649224.
7. Fayyad U. Data mining and knowledge discovery in databases: implications for scientific databases. Proceedings of the Ninth International Conference on Scientific and Statistical Database Management; August 11–13. 1997; Olympia, Washington, USA. IEEE; 1997. pp. 2–11. DOI: 10.1109/SSDM.1997.621141.
8. Giudici P. Applied Data Mining: Statistical Methods for Business and Industry. John Wiley & Sons; 2011.

9. Han J, Pei J, Tong H. Data Mining: Concepts and Techniques. Fourth edition. Cambridge MA: Morgan Kaufmann; 2023.
10. Mohd D, Awang R. Intelligent heart disease prediction system using data mining techniques. 2008 IEEE/ACS International Conference on Computer Systems and Applications, Doha, Qatar, 2008. pp. 108-115. doi: 10.1109/AICCSA.2008.4493524.
11. Kaur H, Wasan SK. Empirical study on applications of data mining techniques in healthcare. J Comput Sci. 2006;2(2):194–200. doi: 10.3844/jcssp.2006.194.200.

