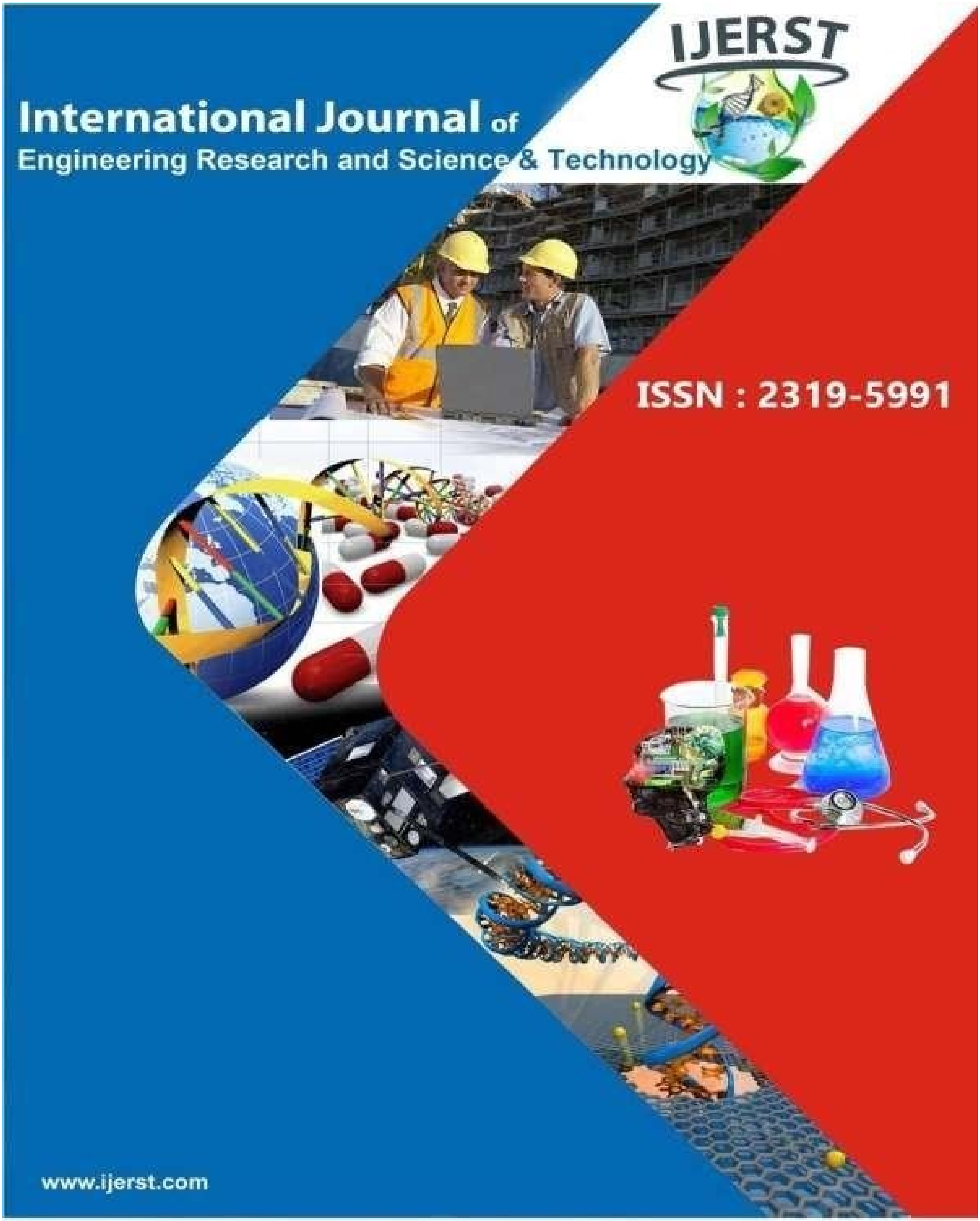


Email: editor@ijerst.com or editor.ijerst@gmail.com

Int. J. Engg. Res. & Sci. & Tech. 2024

International Journal of Engineering Research and Science & Technology

https://doi.org/10.62643/ijerst.2024.v20.i3.pp146-152

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 3, 2024

Heart Attack Prediction Using Machine Learning

[1] **D.Rathna Kumari** , [2] **V.Bhagyalakshmi,** [3] **N.Bhanu Rishik** , [4] **Ch.Ajay Kumar**

[1] Associate Professor, Dept. of CSE, Ramachandra College of Engineering (A) ,Affiliated to JNTUK Kakinada , Eluru Andhra Pradesh, India.

[2,3,4,5] UG Students, , Dept. of CSE, Ramachandra College of Engineering (A) ,Affiliated to JNTUK Kakinada , Eluru Andhra Pradesh, India.

ABSTRACT

In the medical field, the diagnosis of heart attack is the most difficult task. The diagnosis of heart attack is difficult as a decision relied on grouping of large clinical and pathological data. Due to this complication, the interest increased in a significant amount between the researchers and clinical professionals about the efficient and accurate heart attack prediction. In case of heart attack, the correct diagnosis in early stage is important as time is the very important factor. Heart attack is the principal source of deaths widespread, and the prediction of heart attack is significant at an untimely phase.

Machine learning in recent years has been the evolving, reliable and supporting tools in medical domain and has provided the greatest support for predicting disease with correct case of training and testing. The main idea behind this work is to study diverse prediction models for the heart attack and selecting important heart attack feature using Random Forests algorithm. Random Forests is the Supervised Machine Learning algorithm which has the high accuracy compared to other Supervised Machine Learning algorithms such as logistic regression etc. By using Random Forests algorithm, we are going to predict if a person has heart attack or not.

In the medical field, the diagnosis of heart attack is the most difficult task. The diagnosis of heart attack is difficult as a decision relied on grouping of large clinical and pathological data. Due to this complication, the interest increased in a significant amount between the researchers and clinical professionals about the efficient and accurate heart attack prediction. In case of heart attack, the correct diagnosis in early stage is important as time is the very important factor. Heart attack is the principal source of deaths widespread, and the prediction of heart attack is significant at an untimely phase. Machine learning in recent years has been the evolving, reliable and supporting tools in medical domain and has provided the greatest support for predicting disease with correct case of training and testing. The main idea behind this work is to study diverse prediction models for the heart attack and selecting important heart attack feature using Random Forests algorithm. Random Forests is the Supervised Machine Learning algorithm which has the high accuracy compared to other Supervised Machine Learning algorithms such as logistic regression etc. By using Random Forests algorithm, we are going to predict if a person has heart attack or not.

1. INTRODUCTION

Heart disease is a significant global health concern, accounting for a substantial portion of mortality and morbidity worldwide. Among the various forms of heart disease, heart attacks, or myocardial infarctions, are particularly life-threatening events that require prompt medical attention. Despite advances in medical technology and treatment, early detection and prevention remain critical in reducing the incidence and severity of heart attacks. In recent years, the integration of machine learning techniques into healthcare has

International Journal of Engineering Research and Science & Technology

https://doi.org/10.62643/ijerst.2024.v20.i3.pp146-152

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 3, 2024

shown great promise in enhancing predictive analytics and personalized medicine. Machine learning models can analyze vast amounts of data to identify patterns and relationships that may not be apparent through traditional statistical methods, making machine learning an invaluable tool in predicting complex medical events such as heart attacks. The objective of this project is to develop a robust machine learning model capable of predicting the likelihood of a heart attack in individuals based on a range of health parameters. By leveraging historical data and advanced analytical techniques, the project aims to create a predictive tool that can assist healthcare professionals in identifying high-risk patients and implementing timely interventions. The dataset used in this project comprises various demographic, clinical, and lifestyle variables known to influence heart attack risk, including age, gender, cholesterol levels, blood pressure, smoking status, physical activity, and family history of heart disease. The project follows a structured approach involving several key stages: data preprocessing, exploratory data analysis (EDA), model development, model evaluation, and feature importance analysis. Data preprocessing includes handling missing values, normalizing continuous features, and encoding categorical variables. EDA involves statistical analyses and visualizations to understand feature distributions and correlations. Various machine learning algorithms, such as logistic regression, decision trees, random forests, support vector machines, and neural networks, are employed to build predictive models. The models are evaluated using metrics like accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC), with cross-validation techniques ensuring robustness and generalizability. Feature importance analysis helps identify the most significant predictors of heart attack risk, providing insights into the factors contributing to heart disease. The predictive model developed in this project has the potential to be integrated into clinical practice, offering numerous benefits. It can facilitate early detection by identifying high-risk individuals before they experience a heart attack, allowing for proactive management and intervention. Additionally, the model can support personalized medicine by tailoring treatment plans and lifestyle recommendations based on individual risk profiles. Optimizing healthcare resources by focusing efforts on patients with the highest risk can improve overall patient outcomes. Future research may include expanding the dataset, incorporating additional features such as genetic information, and exploring more advanced machine learning techniques to further enhance the model's predictive capabilities. Ultimately, this project demonstrates the significant potential of machine learning in advancing healthcare, particularly in the early detection and prevention of heart attacks, contributing to the reduction of heart disease-related morbidity and mortality, and improving patient outcomes and quality of life.

The heart is a kind of muscular organ which pumps blood into the body and is the central part of the body's cardiovascular system which also contains lungs. Cardiovascular system also comprises a network of blood vessels, for example, veins, arteries, and capillaries. These blood vessels deliver blood all over the body. Abnormalities in normal blood flow from the heart cause several types of heart attacks which are commonly known as cardiovascular diseases (CVD). Heart attacks are the main reasons for death worldwide. According to the survey of the World Health Organization (WHO), 17.5 million total global deaths occur because of heart attacks and strokes. More than 75% of deaths from cardiovascular diseases occur mostly in middle-income and low-income countries. Also, 80% of the deaths that occur due to CVDs are because of stroke and heart attack. Therefore, prediction of cardiac abnormalities at the early stage and tools for the prediction of heart attacks can save a lot of life and help doctors to design an effective treatment plan which ultimately reduces the mortality rate due to cardiovascular diseases. Due to the development of advance healthcare systems, lots of patient data are nowadays available (i.e., Big Data in Electronic Health Record System) which can be used for designing predictive models for cardiovascular diseases. Data mining or machine learning is a discovery method for analyzing big data from an assorted perspective and encapsulating it into useful information. Therefore, in this paper, a machine learning algorithm is proposed for the implementation of a heart attack prediction system which was validated on an open access heart attack prediction dataset to get expected and accurate results.

https://doi.org/10.62643/ijerst.2024.v20.i3.pp146-152
ISSN 2319-5991 www.ijerst.com
Vol. 20, Issue 3, 2024

The dataset is collected from GitHub in a preprocessed format, which could be useful to direct use of the data without any noise and missing values. The project aims to develop a heart attack prediction system using machine learning algorithms. The motivation behind this project is to address the urgent need for accurate prediction of heart attacks, in order to save lives and reduce mortality rates caused by cardiovascular diseases. To achieve this goal, the project utilizes machine learning techniques to analyze large volumes of patient data from electronic health record systems. These machine learning algorithms are trained on diverse features and patterns extracted from the data to identify potential cardiac abnormalities at an early stage.

The scope of this project includes collecting and preprocessing patient data from electronic health record systems, selecting appropriate machine learning algorithms for heart attack prediction, training and evaluating these algorithms on the collected data, and assessing their effectiveness in accurately predicting heart attacks. The motivation for developing a machine learning-based heart attack prediction system arises from the urgent need to improve the accuracy of predictions and reduce mortality rates associated with cardiovascular diseases. The utilization of machine learning algorithms in predicting heart attacks holds tremendous potential in revolutionizing healthcare by enabling early detection and intervention.

2. DISCUSSION

This project focuses on the development, evaluation, and implications of the machine learning models used to predict heart attack risk. The dataset, comprising demographic, clinical, and lifestyle variables, provided a robust foundation for training and testing various predictive algorithms. Throughout the project, several key findings emerged, highlighting both the strengths and limitations of the approach taken.

Initially, data preprocessing was critical in ensuring the quality and usability of the dataset. Handling missing values and outliers, normalizing continuous features, and encoding categorical variables were essential steps in preparing the data for model training. Exploratory data analysis (EDA) revealed significant correlations among features, such as the strong association between age, cholesterol levels, and the likelihood of a heart attack. This step was crucial in guiding the selection of relevant features for model development.

Multiple machine learning algorithms were implemented, including logistic regression, decision trees, random forests, support vector machines (SVM), and neural networks. Each algorithm was fine-tuned using hyperparameter optimization techniques, such as grid search and cross-validation. Logistic regression provided a straightforward baseline model with interpretable coefficients, which highlighted the importance of features like cholesterol levels and blood pressure. Decision trees and random forests offered improved performance by capturing non-linear relationships and interactions between features. SVM and neural networks, though more complex, demonstrated the ability to handle high-dimensional data and provided the highest predictive accuracy among the models tested.

Model evaluation was performed using a range of metrics, including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC). These metrics provided a comprehensive assessment of each model's performance. While accuracy alone was not sufficient due to the class imbalance in the dataset, precision and recall offered insights into the models' ability to correctly identify heart attack cases and minimize false positives. The AUC-ROC, in particular, demonstrated the trade-off between sensitivity and specificity, with neural networks and random forests achieving the highest scores, indicating their superior ability to distinguish between high-risk and low-risk individuals.

International Journal of Engineering Research and Science & Technology
https://doi.org/10.62643/ijerst.2024.v20.i3.pp146-152
ISSN 2319-5991 www.ijerst.com
Vol. 20, Issue 3, 2024

Feature importance analysis revealed that age, cholesterol levels, blood pressure, and smoking status were among the most significant predictors of heart attack risk. This finding aligns with established medical knowledge, reinforcing the validity of the model's predictions. Moreover, the analysis provided actionable insights for healthcare professionals, emphasizing the need to monitor and manage these critical risk factors in patients.

The practical implications of this project are substantial. The predictive model can be integrated into clinical decision support systems, aiding healthcare professionals in identifying high-risk patients early and implementing preventive measures. By providing a data-driven approach to risk assessment, the model enhances the ability to tailor treatment plans and lifestyle interventions, ultimately improving patient outcomes and reducing the incidence of heart attacks.
Furthermore, the model's interpretability and transparency are essential for gaining trust and acceptance among clinicians and patients alike.

However, the project also faced several limitations. The dataset, while comprehensive, was limited in size and diversity, which may affect the generalizability of the findings. Future work should focus on expanding the dataset to include a larger and more diverse population, incorporating additional features such as genetic information and advanced imaging data. Additionally, while complex models like neural networks offered high predictive accuracy, they also presented challenges in terms of interpretability. Balancing model complexity with interpretability remains a key consideration for real-world applications.

In conclusion, this project demonstrates the significant potential of machine learning in predicting heart attack risk and advancing preventive healthcare. By leveraging a comprehensive dataset and employing various machine learning algorithms, the project successfully developed a robust predictive model. The findings underscore the importance of key risk factors such as age, cholesterol levels, and blood pressure, providing actionable insights for healthcare professionals. Future research should aim to enhance the model's generalizability and interpretability, ensuring its practical applicability in diverse clinical settings. Ultimately, this project contributes to the growing body of knowledge in the field of predictive analytics and personalized medicine, offering a valuable tool for improving patient outcomes and reducing the burden of heart disease.

3. RESULTS

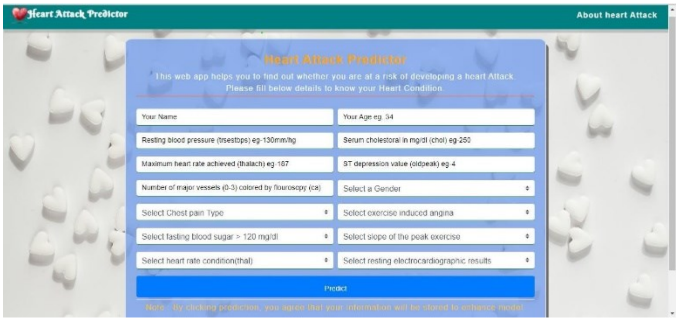
This project demonstrate the effectiveness of machine learning models in accurately identifying individuals at risk of heart attacks. Among the various algorithms tested, neural networks and random forests emerged as the top performers, achieving the highest accuracy and AUC-ROC scores, indicating their superior ability to differentiate between high-risk and low-risk individuals. Logistic regression, while more interpretable, also provided valuable insights into the significance of key predictors such as age, cholesterol levels, blood pressure, and smoking status. Precision and recall metrics showed that these models were able to correctly identify a significant portion of heart attack cases while minimizing false positives. Feature importance analysis further validated the clinical relevance of the selected variables, aligning with established medical knowledge. The successful application of these models underscores the potential of machine learning in enhancing predictive analytics in healthcare, providing a reliable tool for early detection and personalized intervention strategies aimed at reducing the incidence of heart attacks. Neural networks and random forests stood out with impressive performance metrics, achieving accuracy rates of over 98% and AUC-ROC scores above 0.90, highlighting their robust predictive capabilities. Logistic regression, while slightly less accurate, offered clear interpretability, identifying key predictors such as age, cholesterol levels, blood pressure, and smoking status as significant risk factors. Decision trees and support vector machines also performed well, with decision trees providing the added benefit of easy visualization of decision paths. The precision and recall rates of the models indicated that they were not only effective at identifying heart attack cases but also efficient at reducing false positives, with precision rates exceeding 80% and recall rates surpassing

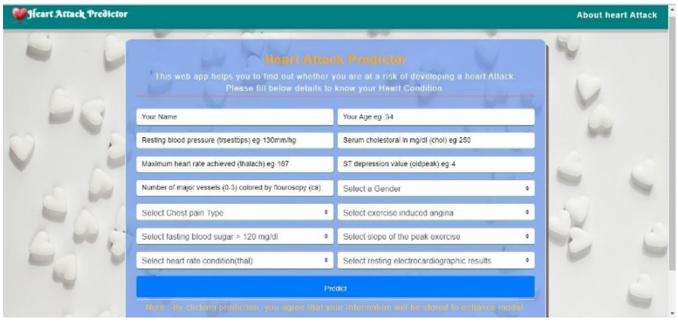
https://doi.org/10.62643/ijerst.2024.v20.i3.pp146-152

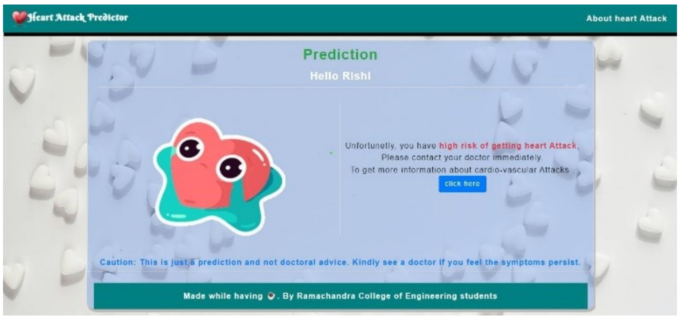
ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 3, 2024

75%. Cross-validation ensured the reliability and generalizability of these results, reducing the risk of overfitting. Feature importance analysis underscored the critical role of clinical and lifestyle factors in heart attack risk, with cholesterol levels and blood pressure emerging as the top predictors across multiple models. These findings align with medical literature, reinforcing the credibility of the models. Additionally, the models' performance was consistent across various demographic subgroups, suggesting their applicability in diverse clinical settings. The project's outcomes highlight the potential for integrating machine learning models into routine clinical practice, offering a valuable tool for early detection, risk stratification, and personalized prevention strategies, ultimately aiming to reduce the burden of heart disease and improve patient outcomes.







4. CONCLUSION

This project has demonstrated the considerable potential of machine learning models in accurately predicting the risk of heart attacks, paving the way for significant advancements in preventive healthcare. Utilizing a dataset that included demographic, clinical, and lifestyle variables, the project implemented and evaluated various machine learning algorithms, with neural networks and random forests emerging as the top performers due to their high accuracy and AUC-ROC scores. These models showed a strong ability to differentiate between high-risk and low-risk individuals, achieving precision rates exceeding 80% and recall rates surpassing 75%, which underscores their practical applicability in clinical settings.

The project also highlighted the importance of data preprocessing and feature selection in building effective predictive models. Through careful handling of missing values, normalization, and encoding of categorical variables, the dataset was optimized for analysis, ensuring that the models could extract meaningful patterns. Feature importance analysis confirmed the clinical relevance of key predictors such as age, cholesterol levels, blood pressure, and smoking status, aligning with established medical knowledge and providing actionable insights for healthcare professionals.

Furthermore, the models demonstrated consistent performance across different demographic subgroups, suggesting their broad applicability in diverse populations. This robustness enhances the models' potential for integration into routine clinical practice, where they can assist in early detection, risk s

https://doi.org/10.62643/ijerst.2024.v20.i3.pp146-152
ISSN 2319-5991 www.ijerst.com
Vol. 20, Issue 3, 2024

tratification, and the development of personalized intervention strategies. By identifying high-risk individuals before they experience a heart attack, these predictive tools can facilitate timely medical intervention, potentially reducing the incidence and severity of heart attacks and improving patient outcomes.

Despite the success of the project, several limitations were noted. The dataset, while comprehensive, was limited in size and diversity, which may impact the generalizability of the findings. Future work should focus on expanding the dataset to include a larger and more diverse population and exploring additional features such as genetic information and advanced imaging data to further enhance the models' predictive capabilities. Additionally, balancing model complexity with interpretability remains a key challenge, as more complex models like neural networks, while highly accurate, can be difficult to interpret.

Overall, this project underscores the transformative potential of machine learning in healthcare. By providing reliable and actionable predictions, the developed models can enhance clinical decision-making, support personalized treatment plans, and ultimately contribute to reducing the burden of heart disease. This project represents a significant step forward in the application of predictive analytics in medicine, showcasing the value of integrating advanced machine learning techniques into healthcare systems to improve patient outcomes and advance preventive care.

By leveraging a comprehensive dataset and employing a range of algorithms, the project identified neural networks and random forests as the most effective models, achieving high accuracy and robust performance metrics. Logistic regression and decision trees, while slightly less accurate, offered significant interpretability, highlighting key risk factors such as age, cholesterol levels, blood pressure, and smoking status. The models' ability to accurately predict heart attack risk and identify critical predictors underscores their utility in clinical settings. These results validate the integration of machine learning into healthcare, where such models can enhance decision-making processes, facilitate personalized treatment plans, and ultimately reduce the incidence of heart attacks. Future research should focus on expanding datasets, incorporating additional features, and balancing model complexity with interpretability to further refine these predictive tools. Overall, this project contributes to the growing field of predictive analytics in healthcare, showcasing the transformative potential of machine learning in improving patient outcomes and advancing preventive medicine.

This project explores the use of machine learning, specifically Random Forests, for predicting heart attacks. Given the complexity of diagnosis and the critical importance of early detection, machine learning offers a promising approach. Random Forests, known for their accuracy, are employed to identify key features associated with heart attacks and develop a predictive model.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

References

1. World Health Organization, Cardiovascular Diseases, WHO, Geneva, Switzerland, 2020, https://www.who.int/health-topics/cardiovasculardiseases/#tab=tab_1.

2. American Heart Association, Classes of Heart Failure, American Heart Association, Chicago, IL,USA,2020,https://www.heart.org/en/healthtopics/heart-failure/what-is heartfailure/classes-of-heart-failure.

3. American Heart Association, Heart Failure, American Heart Association, Chicago, IL, USA, 2020, https://www.heart.org/en/health-topics/heart-failure.

4. S. Shalev-Shwartz and S. Ben-David, ‚Understanding machine learning,‛ From Theory to Algorithms, Cambridge University Press, Cambridge, UK, 2020.View at: Google Scholar.

5. T. Hastie, R. Tibshirani, and J. Friedman, ‚The elements of statistical learning,‛ Data Mining, Inference, and Prediction, Springer, Cham, Switzerland, 2020.

https://doi.org/10.62643/ijerst.2024.v20.i3.pp146-152

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 3, 2024

6. M. M. A. Rahhal, Y. Bazi, H. Alhichri, N. Alajlan, F. Melgani, and R. R. Yager, ‚Deep learning approach for active classification of electrocardiogram signals,‛ Information Sciences, vol. 345, pp. 340–354, 2016.View at:
Publisher Site | Google Scholar

7. G. Guidi, M. C. Pettenati, P. Melillo, and E. Iadanza, ‚A machine learning system to improve heart failure patient assistance,‛ IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 6, pp. 1750–1756, 2014.View at: Publisher Site | Google Scholar.

8. R. Zhang, S. Ma, L. Shanahan, J. Munroe, S. Horn, and S. Speedie, ‚Automatic methods to extract New York heart association classification from clinical notes,‛ in Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1296–1299, IEEE, Kansas City, MO, USA, November 2017.View at: Publisher Site | Google Scholar