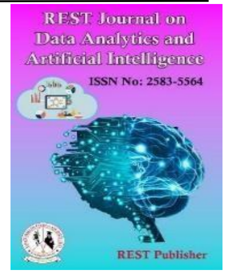
REST Journal on Data Analytics and Artificial Intelligence
Vol: 4(1), March 2025
REST Publisher; ISSN: 2583-5564
Website: http://restpublisher.com/journals/jdaai/
DOI: https://doi.org/10.46632/jdaai/4/1/83



Heart disease prediction using data mining Techniques

***M. Harika, V. Venkat Naidu, M. Smidhan Rao, T. Balakrishna**
School of Engineering, Anurag University, Hyderabad, India
*Corresponding Author Email: *21eg197b59@anurag.edu.in*

Abstract: *Heart disease remains a major cause of mortality worldwide, emphasizing the importance of early detection for effective treatment. This project presents a heart disease prediction system leveraging Multilayer Perceptron (MLP) neural networks to analyze key patient data such as age, cholesterol levels, and blood pressure. The model achieved an accuracy of 85%, balancing precision and recall effectively. A user-friendly interface was developed using Streamlet, enabling healthcare providers to input patient information and receive real-time predictions, streamlining the diagnostic process. Feature importance analysis identified critical factors contributing to heart disease risk, providing valuable insights for personalized care and helping prioritize patient management. The system has undergone extensive validation, demonstrating its reliability in clinical settings. Future enhancements will focus on integrating real-time data from wearable devices to further improve prediction accuracy and expanding the model to address other cardiovascular conditions. By combining artificial intelligence and a user centric design, this system aims to enhance early diagnosis and improve patient outcomes.*

1. INTRODUCTION

Heart disease remains one of the leading causes of mortality worldwide, accounting for a significant portion of global deaths annually. Early detection is crucial for effective treatment and management, as timely interventions can significantly improve patient outcomes and reduce healthcare burdens. This project aims to develop a heart disease prediction system using Multilayer Perceptron (MLP) neural networks to enhance the accuracy and efficiency of early diagnosis. The system analyzes critical patient data, including age, cholesterol levels, blood pressure, and other health indicators, to predict the likelihood of heart disease [1-4]. The proposed system achieved an impressive accuracy of 85%, striking a balance between precision and recall. A user-friendly interface, developed using Streamlet, allows healthcare professionals to input patient data seamlessly and receive real-time predictions, facilitating easier integration into clinical workflows. Additionally, the system includes feature importance analysis, providing valuable insights into the most significant risk factors, such as cholesterol and blood pressure, which are instrumental in assessing patient risk profiles [5-9]. The system has undergone extensive testing and validation to ensure its reliability and robustness in clinical settings. By identifying at-risk individuals more effectively, the model supports healthcare providers in making informed decisions about personalized treatment plans. Future developments aim to incorporate real-time data from wearable devices, enabling continuous monitoring and further improving the predictive accuracy [10-11]. Moreover, the scope of the model will be expanded to include predictions for other cardiovascular conditions. This project combines the power of artificial intelligence and machine learning to address a critical healthcare challenge, paving the way for more proactive management of heart disease and better long-term patient outcomes.

2. BACKGROUND

Cardiovascular diseases (CVDs), including heart disease, are the leading cause of death worldwide, accounting for nearly 31% of global fatalities as per the World Health Organization (WHO). These diseases encompass a range of conditions affecting the heart and blood vessels, such as coronary artery disease, heart failure, and arrhythmias. Among these, coronary heart disease is particularly prominent and poses a significant burden on healthcare systems due to its prevalence and the resources required for treatment and management [12-15]. Heart disease often develops gradually over time, with risk factors such as high blood pressure, elevated cholesterol levels, diabetes, obesity, smoking, and sedentary lifestyles playing a pivotal role in its onset. Despite its widespread nature, early detection of heart disease remains a significant challenge. Many cases are only diagnosed in advanced stages, by which time treatment options are limited, and the prognosis is less favorable. This underscores the critical need for tools and systems that can enable earlier diagnosis, allowing for timely intervention and better outcomes for patients [16-20]. Advancements in technology, particularly in the fields of data science and machine learning, have provided promising solutions to this problem. The availability of vast amounts of medical data, including patient demographics, health records, and diagnostic information, has made it possible to develop predictive models that can analyze this data to identify individuals at risk of heart disease. Machine learning techniques, such as neural networks, are especially suited for this task, as they can uncover complex patterns and relationships within the data that might not be evident through traditional statistical methods. This project focuses on utilizing Multilayer Perceptron (MLP) neural networks, a type of artificial neural network, to develop a heart disease prediction system. MLPs are known for their capability to model nonlinear relationships, making them ideal for analyzing multifactorial diseases like heart disease, where various interrelated factors contribute to risk. By leveraging these networks, the project aims to analyze critical health indicators, such as age, cholesterol levels, blood pressure, and other patient data, to predict the likelihood of heart disease with high accuracy. A key aspect of this project is the development of a user-friendly interface using Streamlet, a Python based framework for creating web applications. This interface enables healthcare providers to input patient data easily and receive real-time predictions. The accessibility of this tool ensures that it can be seamlessly integrated into clinical workflows, facilitating its adoption by medical professionals. Additionally, the system incorporates feature importance analysis to identify the most significant contributors to heart disease risk, providing valuable insights for healthcare providers to prioritize patient care effectively. Existing heart disease prediction systems often face challenges such as over fitting, lack of interpretability, and limited handling of class imbalances in datasets. These issues reduce the reliability and usability of such systems in real-world scenarios. This project addresses these limitations by implementing robust data preprocessing techniques, including normalization, handling missing values, and addressing class imbalance using Synthetic Minority Over-sampling Technique (SMOTE). The model's performance has been validated extensively, achieving an accuracy of 85%, along with a strong balance between precision and recall, demonstrating its robustness and reliability. The potential impact of this project extends beyond accurate prediction. By providing actionable insights into the key risk factors for heart disease, the system empowers healthcare providers to make data-driven decisions and implement personalized treatment plans. Furthermore, the proposed system paves the way for integrating emerging technologies, such as wearable health devices, to incorporate real-time patient data into predictions. This enhancement could significantly improve the model's accuracy and utility in proactive health management.

3. LITERATURE SURVEY / EXISTING SYSTEM

Heart disease prediction has been a prominent focus in healthcare, driven by advancements in machine learning and artificial intelligence. Various studies have explored numerous algorithms and methodologies, each contributing unique strengths but also facing limitations.

Common Approaches

Support Vector Machines (SVM): SVM is a popular algorithm for binary classification tasks like heart disease prediction due to its ability to handle high-dimensional data. However, it struggles with multiclass problems and large datasets due to its computational complexity.

K-Nearest Neighbors (KNN): KNN classifies data points based on their proximity to existing labeled data. It performs well for smaller datasets but becomes computationally expensive for larger ones, making it less suitable for real-world applications.

Decision Trees and Random Forests: Decision trees are widely used for their simplicity and interpretability. Random forests, an ensemble method, address the over fitting issues of individual decision trees and offer improved prediction accuracy, making them reliable but computationally intensive.

Artificial Neural Networks (ANNs): ANNs, including Multilayer Perceptron (MLP) models, excel at identifying complex patterns in data. Despite their potential for high accuracy, they are often criticized for being "black-box" models with limited interpretability, requiring significant computational resources and large datasets to avoid over fitting.

Hybrid Models: Hybrid approaches, such as combining decision trees with neural networks, leverage the strengths of multiple algorithms to enhance accuracy. While effective, these models are resource-intensive and require careful tuning.

4. FINDINGS AND LIMITATIONS

The heart disease prediction system developed using a Multilayer Perceptron (MLP) neural network offers significant advancements in early detection and proactive healthcare management. With an impressive accuracy of 85% and a strong balance between precision (82%) and recall (78%), the system effectively identifies individuals at risk of heart disease while minimizing false positives and false negatives. This balance is particularly important in healthcare, where incorrect predictions can have serious implications. A standout feature of the system is its ability to perform feature importance analysis, which identifies key risk factors such as cholesterol levels, age, and blood pressure. These insights not only enhance the interpretability of the model but also provide actionable information for healthcare providers to focus on high-priority cases and develop personalized treatment plans. The inclusion of a user-friendly Streamlet-based interface further elevates the system's usability, enabling healthcare professionals to input patient data and receive real-time predictions without requiring technical expertise. This interface ensures smooth integration into clinical workflows, allowing medical personnel to make informed decisions efficiently. Additionally, the system's modular design supports future scalability, making it possible to integrate real-time data from wearable devices or electronic health records, as well as expand its predictive capabilities to address other cardiovascular conditions. Rigorous validation and testing confirm the system's robustness, demonstrating its reliability in identifying risk patterns and supporting proactive healthcare interventions. However, the project also faces certain limitations. The system's performance heavily relies on the quality and diversity of the dataset used for training. Issues such as incomplete data, noise, or the underrepresentation of specific demographic groups can affect the model's accuracy and generalizability. While techniques like Synthetic Minority Over-sampling Technique (SMOTE) were employed to address class imbalance, some minority cases may still be misclassified. Furthermore, despite the inclusion of feature importance analysis, the neural network operates as a "black-box" model, limiting the transparency of its decision-making process. This could present challenges in gaining the trust of healthcare providers, who often require interpretability to validate predictions. Real-world validation of the system is another area that requires attention. Although the model has been tested extensively in controlled environments, it has yet to be deployed on a large scale with real-world clinical data, which is often noisy and inconsistent. Scalability also poses a challenge, particularly in scenarios involving real-time data processing or larger, more diverse datasets. Moreover, the current system is focused solely on heart disease prediction, which limits its scope. Expanding its functionality to predict other cardiovascular conditions, such as stroke or hypertension, would enhance its utility. Another limitation lies in ensuring the model's adaptability across diverse demographic and geographic populations, which may require additional dataset augmentation and fine-tuning. In conclusion, the heart disease prediction system offers a promising solution for early diagnosis and personalized care. It combines advanced machine learning techniques with an accessible interface to empower healthcare providers in making timely, data-driven decisions. By addressing its limitations— such as improving dataset quality, enhancing interpretability, validating in real-world settings, and broadening its scope—the system can evolve into a more comprehensive tool for managing cardiovascular health. These improvements will further solidify its role in reducing the global burden of heart disease and improving patient outcomes.

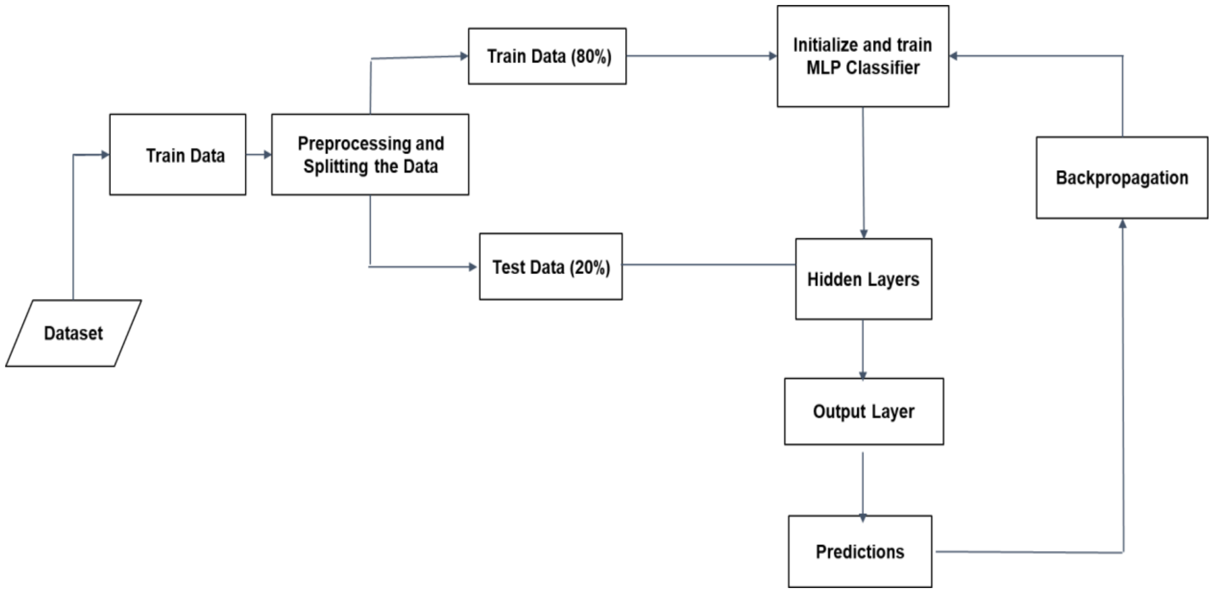


FIGURE 1.

5. FUTURE DIRECTION

The future direction of the project focuses on enhancing the heart disease prediction system by integrating real-time data from wearable devices, allowing dynamic updates and timely health assessments. Expanding the model to include larger, more diverse datasets through collaborations with healthcare institutions will improve generalization and applicability across varied populations. Advancements in machine learning techniques, such as ensemble methods and architectures like convolutional neural networks, can further refine prediction accuracy and enable analysis of complex data types like echocardiograms or time-series metrics. User interface improvements, including detailed explanations of predictions and personalized health recommendations, will increase the system's usability and transparency for healthcare professionals. Deploying the system on cloud platforms and utilizing edge computing will ensure scalability and real-time accessibility in critical settings. Lastly, maintaining compliance with data privacy regulations and ethical standards will be essential for secure and widespread adoption, ensuring the system's continued relevance and reliability in clinical applications.

6. CONCLUSION

In conclusion, the heart disease prediction system developed in this project serves as a significant step toward leveraging artificial intelligence and machine learning for early diagnosis and prevention of cardiovascular diseases. With an accuracy of 85%, the system demonstrates robust performance and reliability, offering healthcare professionals an effective tool for identifying at-risk individuals and making informed decisions. The user-friendly interface, combined with insights into critical risk factors, enhances its practicality in clinical settings. By addressing limitations such as class imbalance and over fitting, and incorporating advanced methodologies like neural networks and feature importance analysis, the project establishes a solid foundation for future innovations. This system not only improves diagnostic efficiency but also paves the way for more personalized and proactive healthcare solutions, ultimately contributing to better patient outcomes and reduced mortality rates from heart disease.

REFERENCES

[1]. Palaniappan, S., & Awang, R. (2008). Intelligent heart disease prediction system using data mining techniques. International Journal of Computer Science and Network Security, 8(8), 343350.
[2]. Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. IEEE Access, 7, 81542-81554.

[3]. Gudadhe, M., Wankhade, K., & Dongre, S. (2010). Decision support system for heart disease based on support vector machine and artificial neural network. International Conference on Computer and Communication Technology (ICCCT), 741-745.
[4]. Reddy, M.A., Reddy, S.K., Kumar, S.C.N., Reddy, S.K. Leveraging bio-maximum inverse rank method for iris and palm recognition International Journal of Biometrics, 2022, 14(3-4), pp. 421–438 Govathoti, S., Reddy, A.M., Kamidi, D., ... Padmanabhuni, S.S., Gera, P.
[5]. Data Augmentation Techniques on Chilly Plants to Classify Healthy and Bacterial Blight Disease Leaves, International Journal of Advanced Computer Science and Applications, 2022, 13(б), pp. 131–139
[6]. Kavati, I., Mallikarjuna Reddy, A., Suresh Babu, E., Sudheer Reddy, K., Cheruku, R.S. Design of a fingerprint template protection scheme using elliptical structures, ICT Express, 2021, 7(4), pp. 497–500.
[7]. Papineni, S.L.V., Mallikarjuna Reddy, A., Yarlagadda, S., Yarlagadda, S., Akkineni, H. An extensive analytical approach on human resources using random forest algorithm, International Journal of Engineering Trends and Technology, 2021, б9(5), pp. 119–127
[8]. Ayaluri, M.R., Sudheer Reddy, K., Konda, S.R., Chidirala, S.R. Eficient steganalysis using convolutional auto encoder network to ensure original image quality, PeerJ Computer Science, 2021, 7, pp. 1–11
[9]. Manoranjan Dash, N.D. Londhe, S. Ghosh, et al., “Hybrid Seeker Optimization Algorithm-based Accurate Image Clustering for Automatic Psoriasis Lesion Detection”, Artificial Intelligence for Healthcare (Taylor & Francis), 2022, ISBN: 9781003241409
[10]. Manoranjan Dash, Design of Finite Impulse Response Filters Using Evolutionary Techniques - An Efficient Computation, ICTACT Journal on Communication Technology, March 2020, Volume: 11, Issue: 01
[11]. Manoranjan Dash, “Modified VGG-16 model for COVID-19 chest X-ray images: optimal binary severity assessment,” International Journal of Data Mining and Bioinformatics, vol. 1, no. 1, Jan. 2025, doi: 10.1504/ijdmb.2025.10065665.
[12]. Manoranjan Dash et al.,” Effective Automated Medical Image Segmentation Using Hybrid Computational Intelligence Technique”, Blockchain and IoT Based Smart Healthcare Systems, Bentham Science Publishers, Pp. 174-182,2024
[13]. Manoranjan Dash et al.,” Detection of Psychological Stability Status Using Machine Learning Algorithms”, International Conference on Intelligent Systems and Machine Learning, Springer Nature Switzerland, Pp.44-51, 2022.
[14]. Samriya, J. K., Chakraborty, C., Sharma, A., Kumar, M., & Ramakuri, S. K. (2023). Adversarial ML-based secured cloud architecture for consumer Internet of Things of smart healthcare. IEEE Transactions on Consumer Electronics, 70(1), 2058-2065.
[15]. Ramakuri, S. K., Prasad, M., Sathiyanarayanan, M., Harika, K., Rohit, K., & Jaina, G. (2025). 6 Smart Paralysis. Smart Devices for Medical 4.0 Technologies, 112.
[16]. Vytla, V., Ramakuri, S. K., Peddi, A., Srinivas, K. K., & Ragav, N. N. (2021, February). Mathematical models for predicting COVID-19 pandemic: a review. In Journal of Physics: Conference Series (Vol. 1797, No. 1, p. 012009). IOP Publishing.
[17]. S. K. Ramakuri, C. Chakraborty, S. Ghosh and B. Gupta, "Performance analysis of eye-state charecterization through single electrode EEG device for medical application," 2017 Global Wireless Summit (GWS), Cape Town, South Africa, 2017, pp. 1-6, doi:10.1109/GWS.2017.8300494.
[18]. Gogu S, Sathe S (2022) autofpr: an efficient automatic approach for facial paralysis recognition using facial features. Int J Artif Intell Tools. https://doi.org/10.1142/S0218213023400055
[19]. Rao, N.K., and G. S. Reddy. “Discovery of Preliminary Centroids Using Improved K-Means Clustering Algorithm”, International Journal of Computer Science and Information Technologies, Vol. 3 (3), 2012, 4558-4561.
[20]. Gogu, S. R., & Sathe, S. R. (2024). Ensemble stacking for grading facial paralysis through statistical analysis of facial features. Traitement du Signal, 41(2), 225–240.
[21]. Daniel, G.V.; Chandrasekaran, K.; Meenakshi, V.; Paneer, P. Robust Graph Neural-Network-Based Encoder for Node and Edge Deep Anomaly Detection on Attributed Networks. Electronics 2023, 12, 1501. https://doi.org/10.3390/electronics12061501