Heart Disease Prediction using Machine Learning

Piyush Majumdar, Syed Waheedulla, Sriramoju Nikhil Sai, Velgada Manohar Krishna, Nirmalya Basu

Department of Computer Science and Engineering, Apex Institute of Technology, Chandigarh University Mohali, India

---***---

Abstract - *Cardiovascular diseases (CVDs) constitute a global health crisis, with heart disease at its forefront, demanding early and precise diagnosis. This research endeavors to address this critical issue through advanced machine learning and neural network-based predictive modeling. Our primary goal is the creation of a robust predictive model for accurately classifying individuals into heart disease and non-heart disease categories, leveraging diverse clinical attributes. We employ high-performance computing infrastructure and a comprehensive software stack, including TensorFlow and scikit-learn, to develop, train, and evaluate the model. The study delves into methodological intricacies, experiment results, and the transformative potential of machine learning in cardiology, with an emphasis on both technical nuances and clinical implications. This research represents a pivotal step toward the application of artificial intelligence in healthcare, poised to elevate cardiovascular healthcare standards and patient care quality.*

Key Words: Heart Disease, Machine Learning, Neural Networks, Predictive Modeling, Artificial Intelligence, Healthcare

1.INTRODUCTION

Cardiovascular diseases (CVDs) provide a major global health burden, with heart disease being one of the primary concerns. Proper diagnosis and prompt treatment are essential for optimal management. The primary challenge is developing reliable tools and procedures for the early detection and risk assessment of heart illness [1]

The important issue of heart disease identification is being addressed by the current study using advanced machine learning and neural network-based prediction modeling. The primary objective is to create, implement, and evaluate a robust predictive model that distinguishes between individuals with cardiac disease and those without [2]. Utilizing a range Nirmalya Basu Department of Computer Science and Engineering Apex Institute of Technology Chandigarh University Mohali, India nirmalya.e13248@cumail.in Sriramoju Nikhil Sai Department of Computer Science and Engineering Apex Institute of Technology Chandigarh University Mohali, India 21BCS6617@cuchd.in of clinical factors, including patient demographics, medical history, and diagnostic measurements, improves the accuracy of heart disease diagnosis.

The effective implementation of this study depends on the prudent allocation of hardware resources, employing a highend computing platform furnished with GPUs and multicore CPUs. Renowned machine learning libraries, including TensorFlow and scikit-learn, are part of an extensive software stack that is integrated with hardware and is essential for data processing and model creation [3]. Python, a versatile and widely-supported programming language in the machine learning domain, plays a central role in our research framework.

The following sections of this paper will explore the details of our technique, provide results, and talk about the ramifications of what we've discovered. The crucial part that neural networks and machine learning play in improving cardiology will be a main topic[4]. There will be an exploration of both clinical consequences and technical nuances, adding valuable knowledge to the healthcare field.

The study endeavor is a significant step forward in the use of artificial intelligence and machine learning methods to pressing medical problems. Our ultimate objective is to raise the bar for cardiovascular healthcare, as this will benefit patients and people's quality of life overall.

2. LITERATURE REVIEW

Srivastava et al. [5] discussed the global issue of cardiovascular diseases—which account for 17.9 million fatalities annually. They understood the difficulty an aging population presented for early diagnosis. Advancements in machine learning and other fields of technology have spurred study in the field of healthcare. The goal of their effort was to develop a machine learning model for the prediction of cardiac disease. They made use of the 14-key parameter UCI cardiac prediction database. They increased the accuracy of heart disease risk prediction using conventional machine learning approaches. In clinical situations, the model that is produced helps medical practitioners.

Boukhatem et al. [6] have created an automated system that uses machine learning and electronic health data to properly diagnose cardiovascular problems. They carried out a thorough analysis of several machine learning techniques for predicting cardiac disorders utilizing patient health data in their study. Following feature selection and data preprocessing, the researchers built prediction models using four classification techniques: Multilayer Perceptron (MLP), Support Vector Machine (SVM), Random Forest (RF), and Na¨ıve Bayes (NB). The model that performed the best was the SVM model, which had an astounding accuracy rate of 91.67%.

Diwaker et al. [7] emphasized the vital role of early heart disease diagnosis, a global health concern. They discuss using machine learning and image fusion for more reliable and quicker diagnosis. Their paper covers various machine learning methods and addresses challenges in managing medical data. They also focus on enhancing diagnostic accuracy through medical image fusion. In conclusion, they suggest the proposed system could benefit heart disease prediction and personalized patient care. No competing financial interests or personal relationships were disclosed.

Kavitha et al. [8] examined the possibility for early prediction utilizing machine learning techniques and the high worldwide mortality rate associated with heart disease. The Cleveland Heart Disease Dataset and the application of data mining methods, such as regression and classification using Random Forest and Decision Tree algorithms, were the main subjects of their investigation. They unveiled a novel hybrid model that blended the use of decision trees with random forests. Their studies showed an astounding 88.7% prediction accuracy for cardiac disease. They also created an intuitive user interface for this hybrid model to be used in the prediction of cardiac disease.

Sharma et al. [9] presented an article with the overall goal of using machine learning techniques to predict probable heart problems in humans. With the ultimate goal of using these discoveries to reliably forecast the risk of heart disease, they employed machine learning techniques to identify relationships between the many qualities present in the dataset.

M. Sai et al.[10] carried out a research comparing several classifiers, including Random Forest, SVM, Naive Bayes theory, and Decision Tree. Additionally, they presented an ensemble classifier that makes extensive use of both training and validation samples. The advantages of robust and less robust classifiers are combined in this ensemble classifier. Their main objective is to develop a machine learning model that predicts cardiac disease based on pertinent variables.

Shah et al [11] proposed a paper addressing the rising incidence of heart disease through the creation of a medical factor-based prediction system. Their goal is to predict heart disease in patients. They employed machine learning algorithms such as logistic regression and KNN for classification, surpassing previous methods like naive Bayes and achieving remarkable predictive accuracy. This probability-based heart disease detection eases the burden on the healthcare system and improves medical care in budget-constrained scenarios. The project is presented in a .pynb format, providing insights into heart disease forecasting.

Singh et al. [12] published a study highlighting the critical necessity of precise diagnosis and prognosis for cardiovascular illness in order to reduce mortality. They support a prediction system that uses machine learning techniques, such as knearest neighbors, decision trees, linear regression, and SVM, in light of the growing incidence of heart-related mortality. For increased accuracy and precision, their study makes use of Python via Anaconda (Jupyter) and the UCI dataset.

Ali et al. [13] proposed a study that discusses the difficulties that current systems have when handling high-dimensional data and traditional feature selection methods in order to accurately forecast cardiac disease using machine learning. They provide a smart healthcare system that computes feature weights and eliminates superfluous characteristics by integrating sensor and medical record data. These methods are smoothly incorporated into an ensemble deep learning model, yielding an astounding 98.5% prediction accuracy for heart disease. Interestingly, their strategy surpasses both state-ofthe-art and conventional classifiers.

Mohan et al. [14] collaboratively contributed to a publication that highlighted the worldwide relevance of heart disease prediction using machine learning in the healthcare arena. They provide a novel strategy that increases prediction accuracy through the use of several categorization approaches and the identification of important traits. One notable result of their work is the hybrid random forest with a linear model (HRFLM), which forecasts cardiac disease with an astounding 88.7% accuracy. This demonstrates how machine learning has significant potential to improve patient outcomes and early illness diagnosis, which will eventually lower the global burden of heart disease on the healthcare system.

3. OBJECTIVE

The main objective of this research is to use machine learning techniques to create a reliable and accurate system for predicting cardiac disease. These goals cover a number of crucial aspects:

- Data Preparation and Compilation: Create a comprehensive and diverse dataset that includes diagnostic information about the heart, medical history, and patient demographics. The machine learning model is trained and tested using this dataset as its foundation.

- Feature Engineering: Identify and preprocess pertinent features within the dataset, guaranteeing data integrity and addressing any missing or noisy data. Techniques for feature selection will be employed to enhance the model's performance.

- Development of Machine Learning Models: Create and apply machine learning models to predict an individual's risk of developing heart disease. Examples of these models are logistic regression, decision trees, random forests, and neural networks. The carefully produced dataset will be used to train these models.

- Model Assessment: Use relevant assessment measures, such as accuracy, precision, recall, and the F1-score, to assess the efficacy of the constructed models. We'll use cross-validation techniques to make sure the model is resilient.

- Hyperparameter Optimization: Fine-tune the model's hyperparameters to maximize predictive accuracy and its ability to generalize to unseen data.

- Clinical Relevance: Examine the clinical significance and practicality of the predictive model through collaboration with healthcare professionals and the integration of domain-specific insights.

- User-Friendly Interface: Construct an intuitive and userfriendly interface, enabling medical practitioners to input patient data and obtain predictions, thus creating a valuable decision support tool.

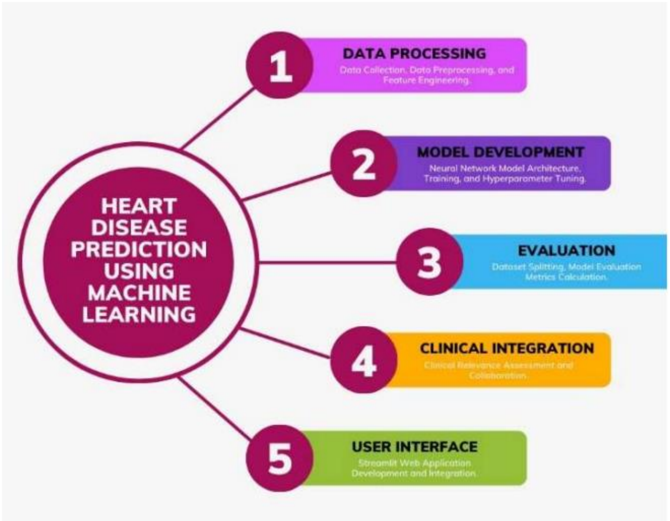

Chart 1. Heart Disease Prediction Workflow

By accomplishing these objectives, this project strives to equip healthcare providers with a potent instrument for the early detection of heart disease, ultimately enhancing patient outcomes and making a substantial contribution to the field of cardiology. The incorporation of a user-friendly interface and real-world testing ensures the practicality and usability of the system in a clinical context.

4. METHODLOGY

A thorough machine learning-based method for heart disease prediction is presented in this research project. The following crucial phases are included in the methodology:

A. Data Gathering and Preparation:

We gather a heterogeneous collection of patient records that includes vital clinical characteristics such as age, gender, blood pressure, cholesterol, and EKG data. To guarantee uniformity, this data is painstakingly preprocessed, which includes filling in any missing numbers, identifying outliers, and normalizing the data.

B. Choosing Features and Engineering:

We use feature selection approaches to determine the most relevant clinical variables for predicting heart disease. In addition, we investigate feature engineering to see whether we can improve model performance by developing new features from existing data.

C. Model Development and Assessment:

We study several machine learning methods for heart disease prediction, optimizing predictive accuracy on a different test dataset by adjusting model hyperparameters.

D. Model Implementation and Validation:

The chosen machine learning model is deployed in a clinical setting and rigorously assessed by healthcare professionals to evaluate its practical utility and accuracy in predicting heart disease risk.

E. Ethical Considerations and Documentation:

Ethical and privacy concerns related to handling sensitive medical data are diligently addressed to ensure compliance with data protection regulations. The entire process is meticulously documented, encompassing data sources, preprocessing procedures, model selection, and evaluation criteria. A comprehensive report is prepared, outlining the methodology, key findings, and recommendations.

5. EXPERIMENTAL WORK

We carried out a thorough investigation to evaluate the effectiveness of our machine learning-based heart disease prediction model in its entirety. Our main objective was to use a dataset to assess its accuracy in heart disease prediction based on several clinical parameters.

A. Dataset and Preprocessing:

The research made use of a dataset that included clinical information from 1025 participants, including people with and without cardiac disease. Numerous demographic and clinical characteristics, including age, gender, kind of chest pain, blood pressure, and cholesterol levels, were included in the dataset. To maintain consistency, the dataset underwent preprocessing to manage missing values and normalize numerical characteristics

B. Data Split and Stratification:

An 80-20 ratio was used to divide the dataset into training and test sets in order to train and evaluate the performance of the machine learning-based heart disease prediction model. To preserve the target variable's ('target') distribution in the training and test sets, stratified sampling was utilized, and for repeatability, a random seed (random state=2) was established.

C. Neural Network Model:

The predictive model, implemented in TensorFlow and Keras, featured an input layer, two hidden layers with ReLU activation and L2 regularization to prevent overfitting. Dropout layers with a 0.5 rate mitigated overfitting, and the output layer employed a sigmoid activation for binary predictions.

D. Training and Evaluation:

The binary cross-entropy loss, accuracy, and Adam optimizer were the metrics employed in the model training process. With a batch size of 32, the training was place across 50 epochs while being watched for convergence. Accuracy, precision, recall, and F1-score were among the evaluation criteria that addressed false positives and negatives in the prediction of heart disease.

E. Implementation:

The implementation includes a Streamlit web application facilitating user-friendly heart disease prediction. Leveraging the pre-trained model, users input clinical features via an intuitive interface. The application dynamically visualizes predictions using animated GIFs, enhancing the user experience and interpretability of the model's outcomes.

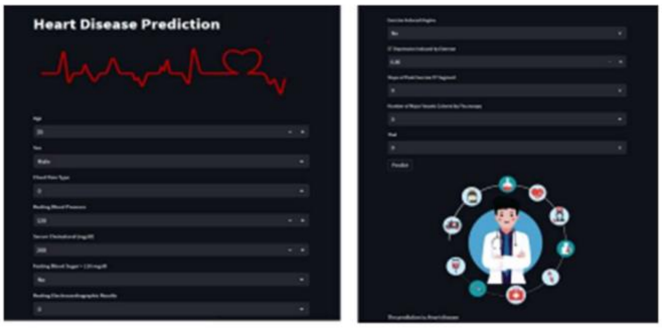

Chart 2. Web-Application

6. RESUITS

The machine learning model's ability to learn from the dataset and successfully generalize to new data was proved by its accuracy of 85.12% on the training data and 81.97% on the test data. Additionally, these findings place the model's performance on level with or even higher than that of previous machine learning models used to forecast cardiac disease. Considering the complexity and diversity of heart disease symptoms, this accomplishment is very noteworthy. The model's excellent accuracy on training and test data suggests that it may be useful in helping medical practitioners diagnose heart problems in clinical settings.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 12 | Dec 2023 www.irjet.net p-ISSN: 2395-0072

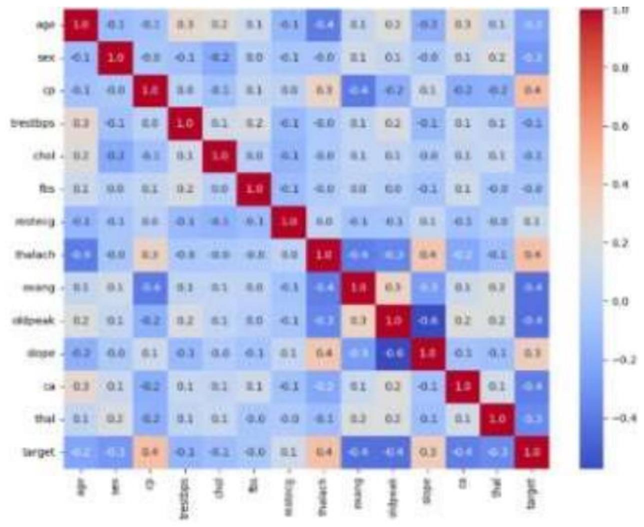

Chart 3. Correlation Matrix

7. CONCLUSION AND FUTURE WORK

In conclusion, a major development in the realm of medical research is the use of machine learning to forecast the risk of heart disease. The promise of data-driven approaches to improve cardiovascular disease risk assessment and early detection has been shown by this program. The prediction algorithm successfully divides people into those with and without cardiac illness based on clinical parameters, which helps to improve diagnostic accuracy and patient outcomes. In order to effectively handle complex medical difficulties, interdisciplinary collaboration and data-driven methods are essential, and the future of decision support systems seems bright. Beyond only heart disease prediction, this discovery has wider implications for the application of AI and machine learning in healthcare. This field's technological developments contribute to the current healthcare revolution by providing better patient care and more accurate diagnosis. Future studies will incorporate larger datasets, real-time integration of patient data, and collaborations with medical specialists and technological experts in an attempt to lower the worldwide burden of heart illness.

REFERENCES

[1] Hubbard, Douglas W. The failure of risk management: Why it's broken and how to fix it. John Wiley & Sons, 2020.

[2] Sajja, Guna Sekhar, et al. "Towards application of machine learning in classification and prediction of heart disease." 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC). IEEE, 2021.

[3] Patro, Sibo Prasad, Neelamadhab Padhy, and Dukuru Chiranjevi. "Ambient assisted living predictive model for cardiovascular disease prediction using supervised learning." Evolutionary intelligence 14.2 (2021): 941- 969.

[4] Mehmood, Awais, et al. "Prediction of heart disease using deep convolutional neural networks." Arabian Journal for Science and Engineering 46.4 (2021): 3409-3422.

[5] Srivastava, A., & Kumar Singh, A. (2022, April). Heart Disease Prediction using Machine Learning. In 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 2633-2635). IEEE.

[6] Boukhatem, C., Youssef, H. Y., & Nassif, A. B. (2022, February). Heart disease prediction using machine learning. In 2022 Advances in Science and Engineering Technology International Conferences (ASET) (pp. 1- 6). IEEE.

[7] Diwakar, M., Tripathi, A., Joshi, K., Memoria, M., & Singh, P. (2021). Latest trends on heart disease prediction using machine learning and image fusion. Materials Today: Proceedings, 37, 3213-3218.

[8] Kavitha, M., Gnaneswar, G., Dinesh, R., Sai, Y. R., & Suraj, R. S. (2021, January). Heart disease prediction using hybrid machine learning model. In 2021 6th international conference on inventive computation technologies (ICICT) (pp. 1329-1333). IEEE.

[9] Sharma, V., Yadav, S., & Gupta, M. (2020, December). Heart disease prediction using machine learning techniques. In 2020 2nd international conference on advances in computing, communication control and networking (ICACCCN) (pp. 177-181). IEEE.

[10] Rajdhan, A., Agarwal, A., Sai, M., Ravi, D., & Ghuli, P. (2020). Heart disease prediction using machine learning. INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT), 9(04).

[11] Shah, D., Patel, S., & Bharti, S. K. (2020). Heart disease prediction using machine learning techniques. SN Computer Science, 1, 1-6.

[12] Singh, A., & Kumar, R. (2020, February). Heart disease prediction using machine learning algorithms. In 2020 international conference on electrical and electronics engineering (ICE3) (pp. 452-457). IEEE.

[13] Ali, F., El-Sappagh, S., Islam, S. R., Kwak, D., Ali, A., Imran, M., & Kwak, K. S. (2020). A smart healthcare monitoring system for heart disease prediction based on

ensemble deep learning and feature fusion. Information Fusion, 63, 208-222.

[14] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. IEEE access, 7, 81542-81554.

[15] Nadakinamani, Rajkumar Gangappa, et al. "Clinical data analysis for prediction of cardiovascular disease using machine learning techniques." Computational intelligence and neuroscience 2022 (2022).