International Journal of Computer Sciences and Engineering
Vol.**11**, Special Issue.**1**, pp.**331-336**, November **2023**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

Research Paper

NTPHD: A Novel Technique to Predict Heart Disease

Ira Nath[1*], Sk Md Toueb Rahaman[2], Arnab Ghosh[3], Suvajit Paul[4], Tathagata Gupta[5], Dharmpal Singh[6]

[1,2,3,4,5]CSE, JIS College of Engineering, Kalyani, India
[6]CSE, JIS University, Kolkata, India

**Corresponding Author: ira.nath@gmail.com*

Abstract: Machine Learning has become a pervasive technology, finding applications in diverse fields across the globe, including the healthcare industry. This transformative technology has the potential to significantly impact medical diagnostics and predictive analysis, aiding in the early detection of various conditions such as locomotor disorders, heart diseases, and many others. By accurately predicting the presence or absence of these ailments in advance, valuable insights can be provided to medical professionals, empowering them to personalize their diagnoses and treatment plans on a patient-by-patient basis, thus revolutionizing the medical field. In this paper, our primary focus lies in predicting possible heart diseases using cutting-edge Machine Learning algorithms. By leveraging the power of these algorithms, we aim to facilitate a comparative analysis of classifiers, including decision tree, K-Nearest Neighbours, Logistic Regression, Support Vector Machine, and Random Forest. Through this analysis, we seek to identify the most suitable classifier that yields the most accurate results for heart disease prediction.

Keywords: Confusion Matrix, K-Nearest Neighbours, Dummies.

1. Introduction

Heart disease, also known as cardiovascular disease, remains a significant global health concern, responsible for a substantial number of deaths worldwide. Cardiovascular diseases account for nearly 17.9 million deaths annually, making it the leading cause of mortality globally. The prevalence of heart disease is expected to rise due to aging populations and lifestyle changes, emphasizing the need for effective early detection and prevention strategies.

In recent years, advances in medical research and technology have paved the way for the development of predictive modelling techniques aimed at identifying individuals at risk of developing heart disease. By analysing a variety of risk factors and patient data, these models can provide valuable insights to clinicians and healthcare providers, enabling them to make informed decisions and design personalized treatment plans.

Heart disease is a major health concern worldwide, contributing to a significant number of deaths annually. Timely and accurate prediction of heart disease can be crucial in enabling healthcare professionals to take preventive measures and provide appropriate treatments. In recent years, machine learning algorithms have shown promising results in predicting heart disease based on patient data.

2. Literature Survey

In 2015, K. Saxena *et. al*. [1] proposed a technique using hill climbing and decision tree algorithms for detecting heart disease. For every actual node decided on by using hill-climbing set of rules a node was chosen by means of a take a look at every stage. Its minimal confidence was zero.25. The accuracy of the device was ready 86.7%

In 2018, Aakash Chauhan, Aditya Jain *et. al*, [2] proposed a technique for prediction heart disease. Data was retrieved directly from electronic records thus reducing manual work. The number of tasks was reduced, and the core number of codes expressed helps in better prognosis of cardiovascular disease. Iterative pattern growing association mining was performed on patient data set to provide stronger associations.

In 2022, Vaibhavi Dhumal, Srushti Gavale *et. al*, [3] developed a method for applying machine learning to predict heart disease. An effective and precise cardiovascular diagnosis method based on machine learning was proposed in this research. For the diagnosis of heart disorders, the raw ECG signal is useful. For analysis, various ECG parameters are used, including heart rate, age, sex, cholesterol level, blood pressure, and ST interval of the ECG signal. Cardiovascular disease was predicted using a variety of machine learning (ML) algorithms.

In 2019, A Lakshmanaro, Y Swathi *et. al*, [4] proposed a Machine Learning Techniques for Heart Disease Prediction. In this paper, machine learning was used to diagnose cardiovascular disease, and modelling techniques was used to deal with imbalanced data. The Framingham_heart_disease data set was publicly available on Kaggle. This data set was used for these analyses. The goal was to determine whether the patient was likely to develop coronary heart disease (CHD) 10 years in the future. The dataset contains 15 items that provided information about patients. Using machine learning.

In 2019, Avinash Golande and Pawan Kumar T *et. al*, [5] developed a Heart Disease Prediction Using Effective Machine Learning Methods, which employed several diagnostic methods to help doctors identify between coronary heart disease and other conditions. K-nearest neighbors, decision trees, and Naive Bayes are frequently used techniques. Packing count, section thickness, minimal streamlining, neural processing, direct kernel self-organization guidance, and SVM (bolster vector machine) were additional comprehensive techniques depending on the primary features employed, respectively.

In 2016, Deepika Kumari *et. al*, [6] Proposed predictive analysis using machine learning techniques for chronic disease prevention and control to ensure accuracy.

In 2019, Senthil Kumar Mohan *et. al*, [7] proposed an Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques with the primary aim of improving accuracy in cardiovascular problems. KNN, LR, SVM, and NN algorithms were applied to achieve a performance improvement of 88.7% accuracy using a cardiovascular prediction model.

In 2022, Md. Mahbubur Rahman *et. al, [8]* proposed a web-based heart disease prediction system using machine learning algorithms. The proposed prediction algorithm uses some health information to predict heart disease. The system used thirteen health data. Eight different algorithms were used for accurate heart failure detection, this study also a website was set up to easily checked their heart status immediately from home.

3. Methodology

Firstly, a dataset collected then the next task is pre-processing the data. After that applying machine learning algorithms to predict the disease. The flowchart is shown in Figure 1.

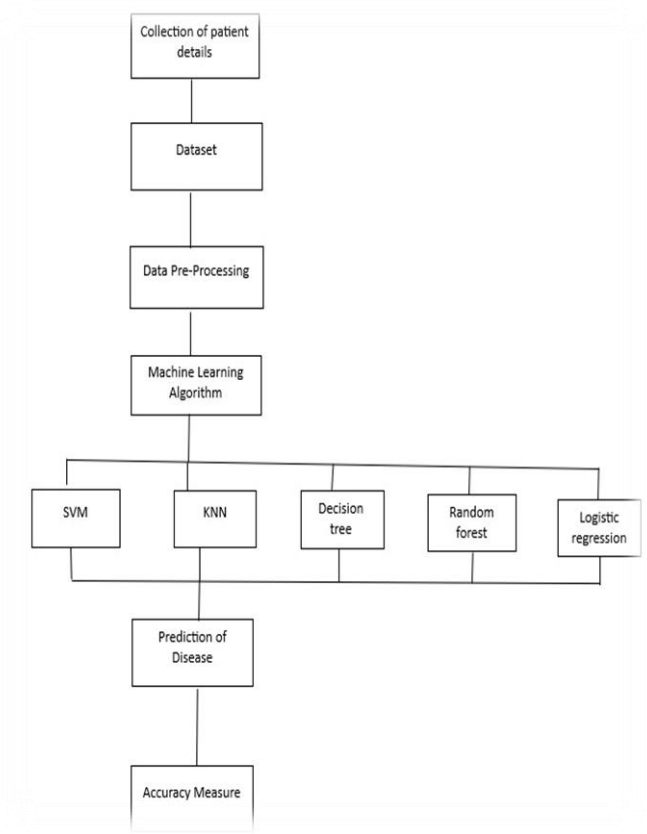

Figure 1. Flowchart of our proposed work

3.1 Dataset Details

To predict heart disease, we must gather some data. After gathering the data, we separated it into training and test sets. The test data are used to assess the prediction model after it has been learned using the training dataset. In this project, 80% of the data are utilized for training and 20% are used for testing. The information was gathered from:
Link-
https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset

All Input Attributes -

- Sex
- Chest Pain
- Fasting Blood Exang – exercise induced angina
- CA – number of major vessels colored by fluoroscopy
- Thal
- Trest Blood Pressure
- Serum Cholesterol (mg/dl)
- Thalach – maximum heart rate achieved
- Age in Year
- Cholesterol
- Restecg

3.2 Correlation Matrix

The correlation coefficients between the variables are displayed in a table called a correlation matrix. The relationship between each possible pair of values in the table is depicted in Figure 2. It is a method for identifying and modelling in the given data, as seen in Figure 2, and for

summarizing a large amount of data. The row-column arrangement of coefficients helps users explore the relationship between two or more variables and how they depend on each other The values of the matrix range from -1 to 1.

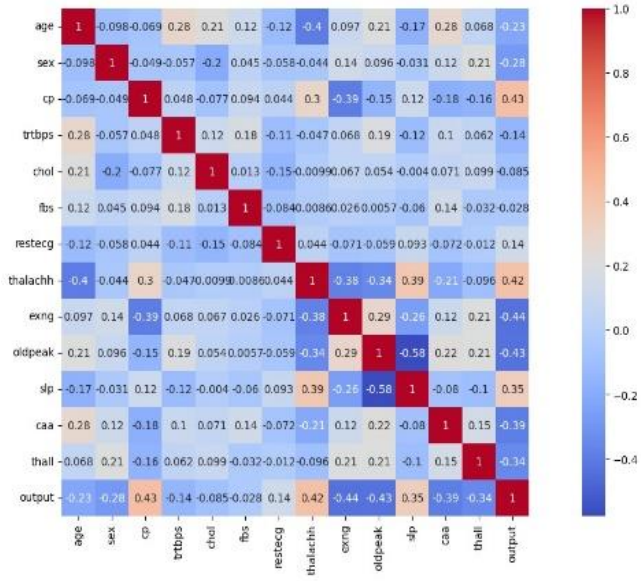

Figure 2. Correlation Matrix

3.3 Distribution of Data

When a problem has to be predicted or classified, data distribution is an important part of it. Here we have shown the distribution of each characteristic such as age, sex, resting blood pressure, adiposity, etc. In Figure 3, we have shown the distribution.

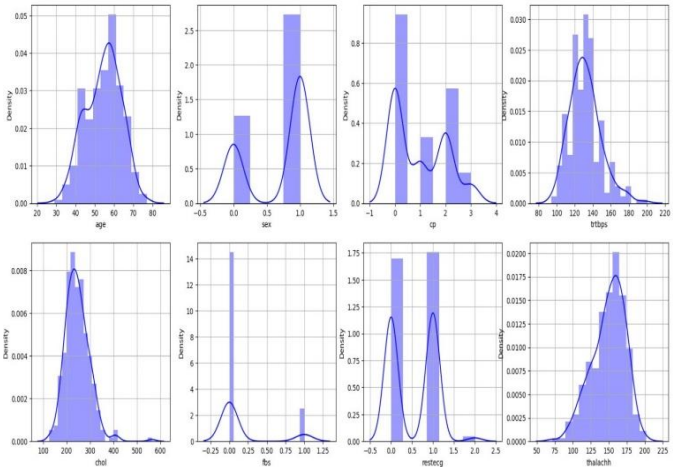

Figure 3. Distribution of Data

3.4 Creating Dummies

A data set may have a variety of values, perhaps even a range of values. Therefore, we generate dummy variables in order to better utilize those categorical data for programming. Dummy variables are binary variables that show if a different categories variable accepts a specific value. Using the get_dummies () method, we may construct dummy variables in Python. Figure 4 shows the code and output of Dummy creating are shown.

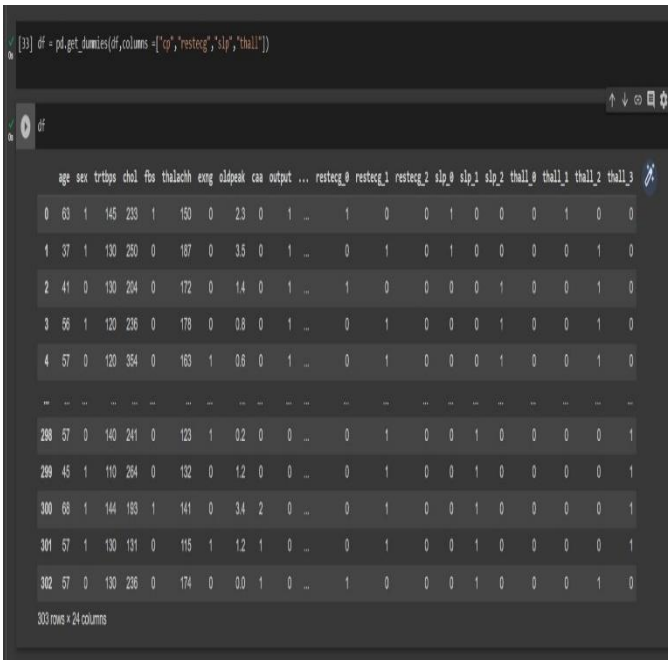

Figure 4. Dummies of Categorical values

3.5 Prediction of Disease

Classification is performed using machine learning algorithms such as SVM, KNN, Decision Tree, Random Forest, and Logistic Regression. A comparative analysis between algorithms is performed and the algorithm that provides the highest accuracy is used to predict heart disease.

4. Performance Analysis

In this study, heart disease is predicted using a variety of machine learning methods, including SVM, KNN, Decision Tree, Random Forest, and Logistic Regression. Only 13 variables from the Heart Disease Prediction dataset are taken into account when predicting heart disease. The accuracy of each method is calculated with the help of its confusion matrix, which is depicted in Figures 5, 6, 7, 8, and 9.

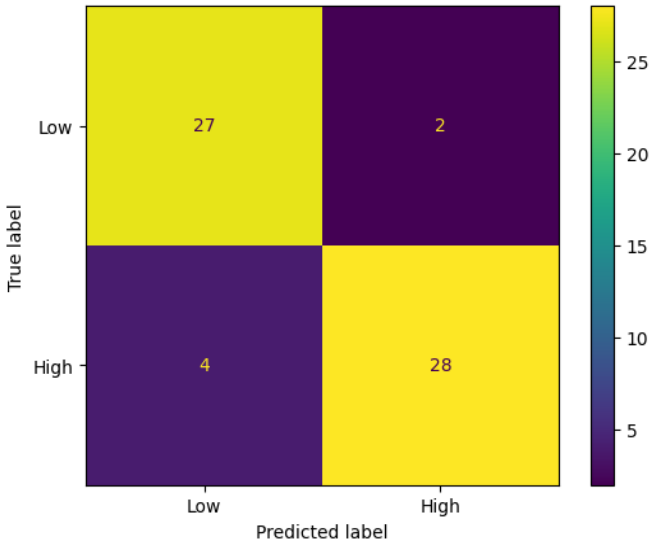

Figure 5: Confusion Matrix using Logistic Regression

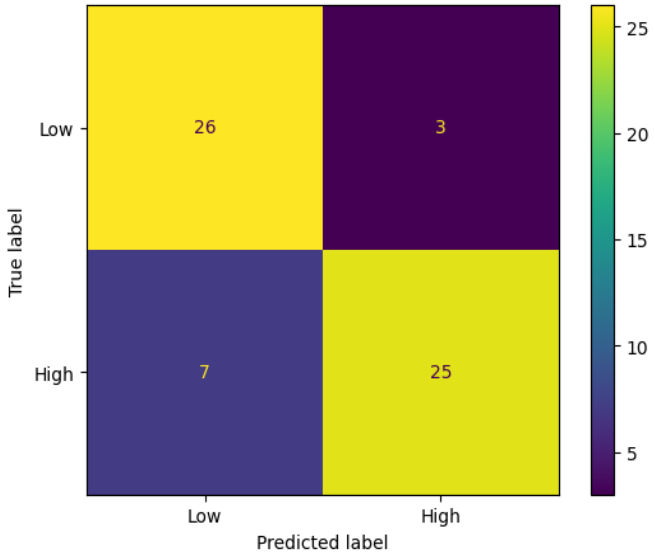

Figure 6: Confusion Matrix using SVM

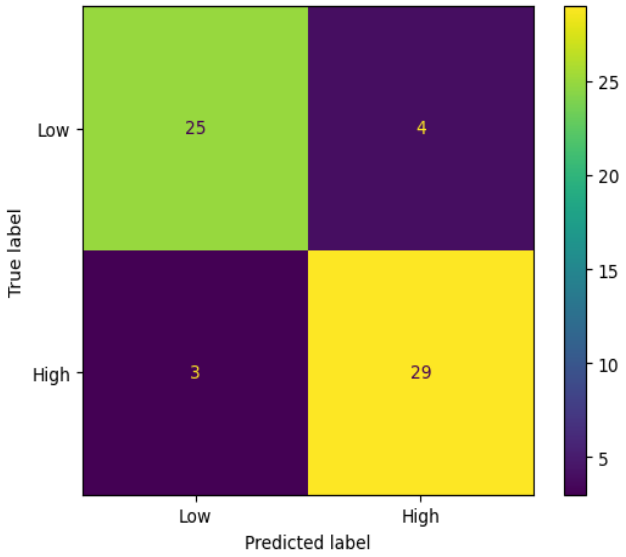

Figure 7. Confusion Matrix using Decision Tree

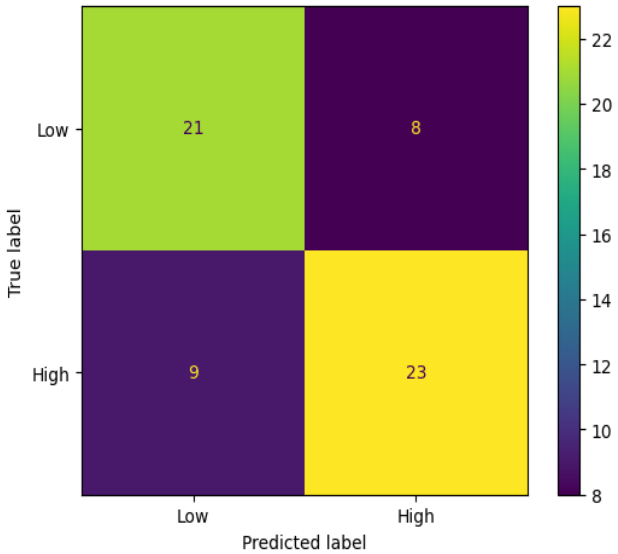

Figure 8: Confusion Matrix using KNN

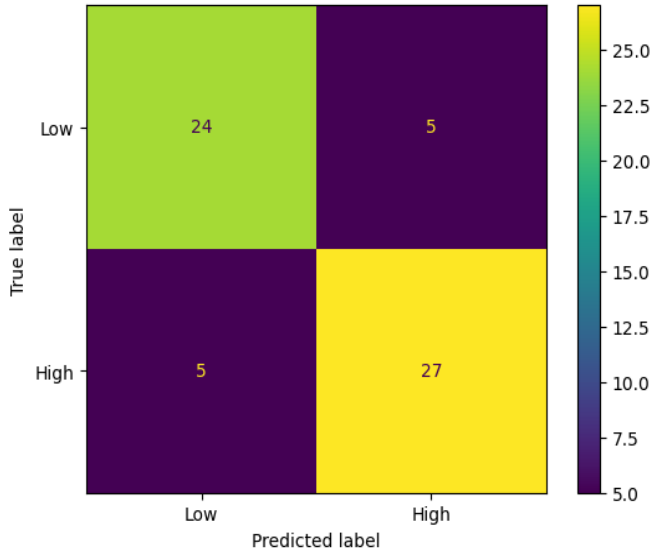

Figure 9: Confusion Matrix using Random Forest

5. Result

After the model is trained, Logistic Regression performs well in comparison to other techniques. The confusion matrix is used to help the calculation of accuracy. The comparison below demonstrates that Logistic Regression has the highest accuracy (90.16%).

Table 1: Accuracy comparison of various algorithms

Algorithm	Accuracy	Precision	Recall	F1-Score
SVM	83.60%	89.65%	78.78%	83.87%
Logistic Regression	90.16%	93.10%	87.09%	90%
Decision Tree	83.60%	82.75%	82.75%	82.75%
KNN	72.13%	72.41%	70%	71.18%
Random Forest	88.52%	86.20%	89.28%	87.71%

5.1. Comparison of Machine Learning Model with Various Literature [11]

A Khan and colleagues developed a method for predicting cardiovascular illness using machine learning algorithms in 2023. Table 2 displays a comparison of accuracy using their model.

Table 2 Accuracy Comparison with various model.

Algorithms	Proposed Model	Other Model
Random Forest	88.52	85.01
Logistic Regression	90.16	83.08
Decision Tree	83.60	83.72
SVM	83.60	83.08

Input & Output:

After putting all attributes in input section that is shown in Figure.10 what output will come shown in Figure. 11.

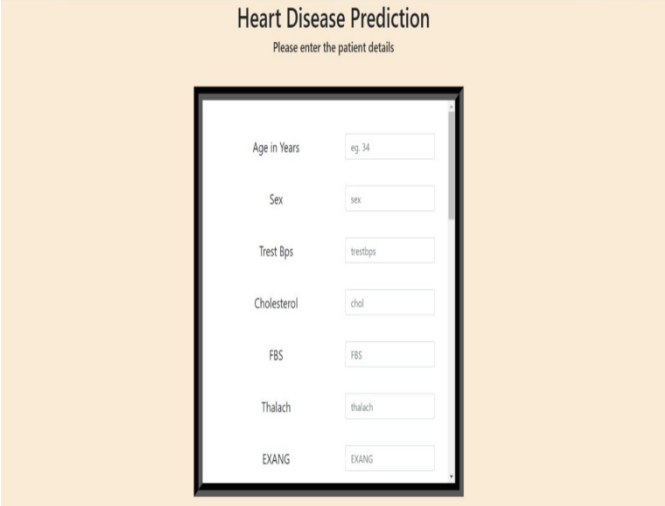

Figure 10: Input

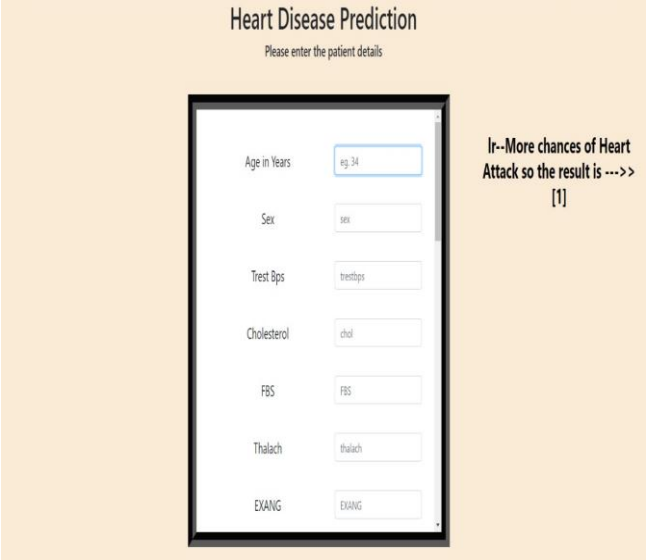

Figure 11: Output

6. Future work & conclusion

Integrating genetic data into Heart disease prediction models could enhance their accuracy. Genetic predisposition plays a significant role in heart diseases, and including this information could provide more personalized predictions.

Using longitudinal data (data collected over a period of time could help in tracking the progression of heart disease risk factors and understanding how they contribute to the development of heart disease over time.

Developing mobile apps that allow users to track their ifestyle factors (such as diet, exercise, and stress) and receive real-time risk assessments could empower individuals to take proactive measures to reduce their risk of heart disease.

Integrating prediction models with electronic health records could facilitate more accurate predictions by considering a wider range of medical history and test results.

Developing methods to explain and interpret the predictions of heart disease models will be crucial for gaining the trust of healthcare professionals and patients.the prediction of heart disease has come a long way thanks to advancements in data science and machine learning.

These models have demonstrated their potential in identifying individuals at risk of heart disease, which can lead to early interventions and improved outcomes. However, it's important to remember that these models are tools to assist healthcare professionals rather than replace their expertise.

References

[1] Saxena, Kanak, and Richa Sharma. "Efficient heart disease prediction system using decision tree." In *International Conference on Computing, Communication & Automation*, pp.72-77, 2015, IEEE.

[2] Chauhan, Aakash, Aditya Jain, Purushottam Sharma, and Vikas Deep. "Heart disease prediction using evolutionary rule learning." In *2018 4th International conference on computational intelligence & communication technology (CICT)*, pp. 1-4, 2018, IEEE.

[3] Prof. Sachin Sambhaji Patil, Vaibhavi Dhumal, Srushti Gavale, Himanshu Kulkarni, Shreyash Wadmalwar, " Heart Disease Prediction using Machine Learning", International Journal of Scientific Research in Science and Technology (IJSRST), Print ISSN : 2395-6011, Online ISSN : 2395-602X, Vol.9, Issue.6, pp.541-546, 2022.

[4] Lakshmanarao, A., Y. Swathi, and P. Sri Sai Sundareswar. "Machine learning techniques for heart disease prediction." *Forest* 95, no.99: 97, 2019.

[5] Avinash Golande, Pavan Kumar T proposed "Heart Disease Prediction Using Effective Machine Learning Techniques". International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Vol.8, Issue.1S4, 2019.

[6] Deepika, Kumari, and S. Seema. "Predictive analytics to prevent and control chronic diseases." In *2016 2nd international conference on applied and theoretical computing and communication technology (iCATccT)*, pp. 381-386, 2016, IEEE.

[7] Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." *IEEE access* 7: pp.81542-81554, 2019.

[8] Rahman, Md Mahbubur, Morshedur Rahman Rana, M. N. A. Alam, Md Saikat Islam Khan, and Khandaker Mohammad Mohi Uddin. "A web-based heart disease prediction system using machine learning algorithms." *Network Biology* 12, no.2: 64-80, 2022.

[9] Khan, Arsalan, Moiz Qureshi, Muhammad Daniyal, and Kassim Tawiah. "A Novel Study on Machine Learning Algorithm-Based Cardiovascular Disease Prediction." *Health & Social Care in the Community* 2023 (2023).

[10] Riyaz, Lubna, Muheet Ahmed Butt, Majid Zaman, and Omeera Ayob. "Heart disease prediction using machine learning techniques: a quantitative review." In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 3*, pp. 81-94. Springer Singapore, 2022.

[11] Sun, Huating, and Jianan Pan. "Heart Disease Prediction Using Machine Learning Algorithms with Self-Measurable Physical Condition Indicators." *Journal of Data Analysis and Information Processing* 11, no.1: pp.1-10, 2023.

[12] Boukhatem, Chaimaa, Heba Yahia Youssef, and Ali Bou Nassif. "Heart disease prediction using machine learning." In *2022 Advances in Science and Engineering Technology International Conferences (ASET)*, pp. 1-6, 2022, IEEE.

[13] Yadav, Anupama, Levish Gediya, and Adnanuddin Kazi. "Heart disease prediction using machine learning." *International Research Journal of Engineering and Technology (IRJET* 8, no. 09, 2021.

[14] Jindal, Harshit, Sarthak Agrawal, Rishabh Khera, Rachna Jain, and Preeti Nagrath. "Heart disease prediction using machine learning algorithms." In *IOP conference series: materials science and engineering*, vol. 1022, no. 1, p. 012072. IOP Publishing, 2021.
[15] Katarya, Rahul, and Sunit Kumar Meena. "Machine learning techniques for heart disease prediction: a comparative study and analysis." *Health and Technology* 11: pp.87-97, 2021.
[16] Ali, Md Mamun, Bikash Kumar Paul, Kawsar Ahmed, Francis M. Bui, Julian MW Quinn, and Mohammad Ali Moni. "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison." *Computers in Biology and Medicine* 136 (2021): 104672.
[17] Patel, Jaydutt, Azhar Ali Khaked, Jitali Patel, and Jigna Patel. "Heart disease prediction using machine learning." In *Proceedings of Second International Conference on Computing, Communications, and Cyber-Security: IC4S 2020*, pp. 653-665. Springer Singapore, 2021.

AUTHORS PROFILE

Dr. Ira Nath is presently working as an Associate Professor in the Department of Computer Science and Engineering of JIS College of Engineering, India. She has about 17 years of experience in teaching and research. She has received her Ph.D in Computer Science & Technology at Indian Institute of Engineering Science and Technology (IIEST), Shibpur, India in 2020. She received the Master of Technology (M.Tech.) degree in Software Engineering from the Maulana Abul Kalam Azad University of Technology, India formerly West Bengal University of Technology, India in 2008. She also received the degree of Bachelor of Technology (B.Tech.) in Computer Science and Engineering from the same university in 2005. She has published 42 papers in referred journal and conferences index by Scopus, DBLP and Google Scholar. Her research interests include Network Security regenerator placement, survivability and routing and wavelength assignment in translucent WDM optical Networks, Wireless Sensor Network and Machine Learning.

Arnab Ghosh is a recent graduate from JIS College of Engineering, having completed his B.Tech. in Computer Science and Engineering. His academic journey reflects a strong foundation in computer science, and he is poised to make meaningful contributions to the field. Arnab is driven by a passion for technology and its applications, as evidenced by his capstone project on heart disease prediction using various machine learning algorithms. This project, undertaken with a group of dedicated peers, showcases his commitment to leveraging cutting-edge techniques for solving real-world problems in healthcare.

Sk Md Toueb Rahaman is a graduate from JIS College of Engineering with a focus on Computer Science and Engineering, stands out as a key author contributing significantly to the project. His expertise in Machine Learning plays a pivotal role in shaping the research's scientific foundation. From the conceptualization of innovative methodologies to their meticulous implementation. Moreover, Rahaman's comprehensive understanding of the existing literature and related work has greatly informed the contextualization of the research within the broader academic landscape.

Tathagata Gupta is a recent graduate from JIS College of Engineering with a degree in Computer Science and Engineering. During his academic journey, he demonstrated a keen interest in the intersection of technology and healthcare. His notable achievement includes a project focused on predicting heart disease using machine learning techniques, showcasing his ability to apply advanced computational methods to real-world challenges. Tathagata's dedication to his studies and the successful completion of this project highlights his commitment to leveraging cutting-edge technologies for the betterment of healthcare.

Dr. Dharmpal Singh received his Bachelor of Computer Science and Engineering and Master of Computer Science and Engineering from West Bengal University of Technology. He has done his Ph.D in year 2015. He has about 17 years of experience in teaching and research. At present, he is with JIS University, West Bengal, India as a Professor and Head of the department. He has published 45 papers in referred journal and conferences index by Scopus, DBLP and Google Scholar and editorial team and senior member of many reputed journal index by SCI, Scopus, DBLP and Google Scholar. He has organized seven national levels Seminar/Workshop, published two patent and has applied for the AICTE Research Project (MRP) in year of 2019.

