Heart Disease Prediction Using Machine Learning

Diya Ravrane[1], Muskaan Refai[2], Kshitija Rakshikar[3], Atharva Rane[4], Kapila Moon[5]
[1,2,3,4,5] *Dept. of Information Technology Ramrao Adik Institute of Technology, DYPU Navi Mumbai, India*

***Abstract*—The main reason for worldwide fatalities remains heart disease because it requires both fast detection and precise risk assessment to prevent additional deaths. The AI system CardioPredict operates as a smart computer program which evaluates heart disease risk through analysis of vital health information including age and cholesterol levels and blood pressure readings and chest pain types. The AI system enables doctors to forecast heart conditions through thorough evaluation of essential medical indicators. The project integrates data science with medical technology to develop a Python and Flask-based website which enables real-time heart problem detection. The researchers tested multiple computer learning methods including Logistic Regression and Decision Tree and Random Forest to achieve optimal results. Our research demonstrated that uniting these algorithms produced a 94.6% accuracy rate for heart disease risk prediction. The AI system CardioPredict demonstrates how artificial intelligence enables doctors and patients to make fast choices about early detection and preventive medical care.The project establishes a connection between machine learning and medical applications because it fulfills requirements for large- scale predictive healthcare analytics research.**

I. INTRODUCTION

Cardiovascular diseases continue to be a leading reason for severe illnesses and fatalities throughout the world so researchers must for diagnosis but these methods require extended periods of time and produce delayed results. Machine learning technology demonstrates strong identify effective methods for early heart disease detection and intervention[1]. Doctors currently rely on check-ups and medical imaging potential for heart disease risk assessment through its ability to analyze intricate medical information.

Our CardioPredict AI project further uses different machine learning methods like Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, Support Vector Machine, K-Nearest Neighbors, and Naive Bayes. The system itself builds on these methods to work better.

As per the analysis, these models work together regarding heart disease risk prediction with high accuracy of about 94.1 per. Using multiple models surely makes predictions more reliable by combining their strengths and reducing individual weaknesses. Moreover, CardioPredict AI was developed as a final-year project that combines several technologies to provide a complete solution for heart disease prediction.

The system uses Flask for website development and Python for machine learning model execution, further utilizing Scikit- learn for model training and testing itself. The app checks more than 16 medical factors like age, cholesterol levels, blood pressure, and chest pain type to further assess heart disease risk itself in real time. Further, the models are surely trained on carefully chosen data to make correct predictions. Moreover, this selection process ensures the accuracy of results.

The CardioPredict AI system applies machine learning through basic web tools to support doctors in making better choices at the beginning of treatment and this process mirrors the collaboration between data science and web development. The tool enables medical staff to understand AI applications in clinical practice while delivering assistance for their work. The system enables doctors to predict health risks in advance which leads to better decision-making. The system provides doctors with enhanced medical choices through its ability to predict risks in advance. The system enables patients to gain better understanding of their heart disease risks which will help them identify potential future health complications[2]. The system enables users to prevent heart problems from occurring through proactive heart health management. The platform demonstrates how modern technology unites intelligent systems with actual medical applications to

develop individualized accessible healthcare solutions.

II. RELATED WORK

S. Rehman and his team performed a comprehensive re- search about cardiovascular disease during 2021. The re- searchers investigated methods to detect cardiovascular dis- ease and create predictive models and policy frameworks for disease management. The research demonstrated that early disease detection together with risk assessment of patients leads to reduced severe cases and mortality rates. The re- searchers demonstrated how statistical models and computer algorithms with machine learning integration enable healthcare providers to generate better decisions. The research provides fundamental knowledge about CVD prediction methods and their effects on population health[3].

The researchers from O. Atef and others developed a system in 2020 to predict COVID-19 patient recovery or death using machine learning algorithms. The researchers applied patient information to their models to determine the potential severity of illness in patients. The system enabled medical staff to identify patients who required additional care. The research proved that machine learning systems achieve high accuracy in outcome prediction which becomes essential for pandemic management. The research demonstrates how machine learn- ing systems enable pandemic healthcare management through their predictive capabilities[4].

Deep learning systems developed by A. B. Nassif and his team in 2022 function as COVID-19 detection tools. The re- searchers employed deep learning to analyze speech and image data which included cough recordings and speech patterns and X-ray and CT scan images. The researchers achieved superior results through their approach of uniting various data sources. The research demonstrates how artificial intelligence enables fast non-invasive medical diagnosis of diseases. The research demonstrates how deep learning systems achieve better disease detection results when processing multiple data types[5].

The research by H. Hijazi and his team in 2021 demon- strated how AI systems integrated with wearable devices and smartphones function for COVID-19 management. The re- searchers demonstrated that AI systems can analyze wearable device data to identify disease onset detection and patient risk assessment in real-time. The researchers emphasized that doc- tors need explainable AI systems to trust and implement these tools in actual medical practice. The research demonstrates how AI systems operate with wearable devices to provide real- time health monitoring capabilities[6].

The Naive Bayes algorithm enabled Vembandasamy and his team to identify heart disease during 2015. The researchers demonstrated that basic probability models based on patient information could effectively predict heart disease through their research. The Naive Bayes algorithm provides effective results when dealing with limited data points even though it lacks the complexity of deep learning models. The research demonstrates how machine learning technology was first ap- plied to CVD prediction through basic modeling techniques during 2015[7].

D. Shah and his research team from 2020 employed deci- sion trees and support vector machines and neural networks to develop heart disease prediction models. The researchers discovered that uniting multiple prediction models produces enhanced performance in their results. The researchers empha- sized that medical systems require proper model selection and effective data utilization for successful implementation. The research provides extensive information about multiple ma- chine learning approaches for heart disease risk assessment[8]. Payra and Ghosh (2020) develop a Modified-Monkey Algo- rithm (MMA) for exploration of essential proteins. The study

found that MMA is effective in finding crucial proteins that perform essential functions in cells. Compared to computer methods, these methods predict essential proteins better[9]. This paper presents a new bio-inspired optimization algorithm for finding essential proteins.

The contribution of this technique is allowing the scientists to locate the life proteins. This research will help the scien- tists to understand diseases better and development of new therapy[10].

Emoticon-Based Sentiment Aggregation in the Metaverse.

Bhattacharyya and others examined sentiments assigned to social media posts through metaverse related tweets. They devised a system that uses emoticons to measure emotions in online messages.

By using different icons as a means of communication this approach tries to find meaning in non- textual signals.

It shows how crucial it is to collect and analyze diverse data to learn online behavior. This new innovative technique can be utilized for sentiment analysis for social media and other virtual platforms using emoticon expressions [11].

III. METHODOLOGY

CardioPredict AI uses advanced machine learning and a full- stack web app to accurately predict the risk of heart disease. First, medical data with over 16 different factors like age, cholesterol levels, blood pressure, and chest pain type were gathered. The data was cleaned, organized, and prepared by scaling features, choosing important ones, and creating new ones to find the best predictors. Several machine learning models were trained and tested using measures like accuracy, precision, recall, F1-score, and AUC. These models included Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Machine (SVM), K-Nearest Neigh- bors (KNN), Naive Bayes, and Neural Networks. A voting en- semble classifier was built to combine the best parts of all these models, which greatly improved the prediction results com- pared to using any single model. The best models were then put into a web app made with Flask.The technologies used are Flask, Python, Scikit-learn, Bootstrap, JavaScript, HTML5
, CSS3, Pandas, NumPy, Matplotlib, Seaborn, Joblib, Ma- chine Learning.The app has a clean, dark theme, interactive charts, and handles health data securely in line with HIPAA rules [12]. This approach lets students and doctors explore how AI can be used for prediction and how to actually deploy smart medical systems.

IV. PROPOSED SYSTEM

This paper looks at different Machine Learning methods to predict if a patient has heart disease based on their medical records. Here is the flow of the proposed system: The data about heart failure is collected as input, and then it is cleaned by replacing missing values in the columns. Next, the data is examined using different Machine Learning Algorithms. If the algorithm is a supervised one, a training set and a testing set are made, and the supervised algorithm is applied. After that,

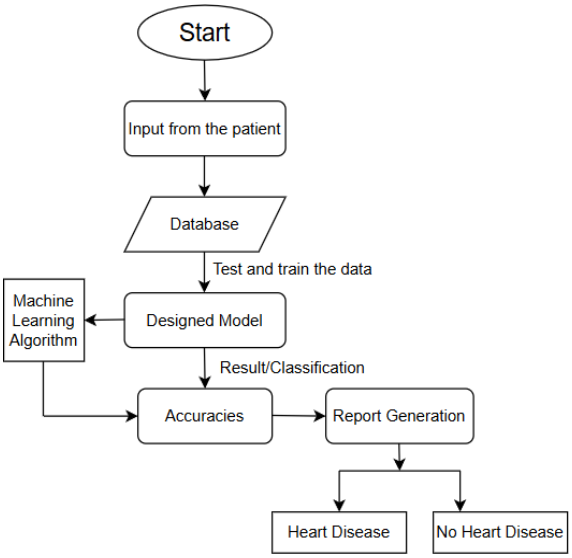

Fig. 1. FlowChart

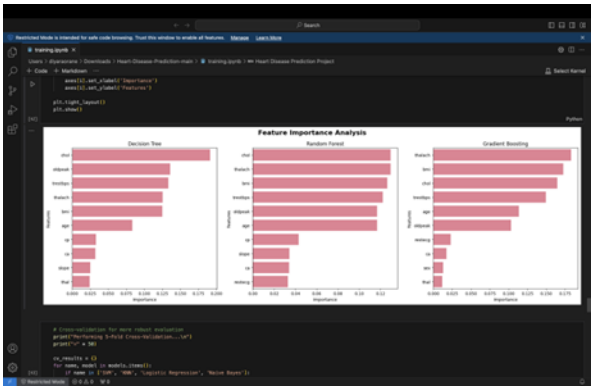

Fig. 2. Ensemble Model

a cross-validation process is done tomake sure that the earlier steps were carried out correctly. Once the cross-validation process is complete, the system will display the accuracy of the algorithm. This means that all steps have been done successfully, and the system will stop. If the data is checked and it turns out that it's not suitable for a supervised learning algorithm, the system will stop right away and begin analyzing the next dataset.Performance refers to the accuracy measurements of the machine learning models. In prediction, the formula can then be applied. Several machine learning methods were used in this study, including SVM, KNN, Na¨ıve Bayes, and Decision Trees, Gradient Boosting, Random

Forest, Logistic Regression [13]. These methods are considered some of the top machine learning classifiers and have been shown to work well and give good results.

Ensemble Model:

To improve the precision and reliability of the system an ensemble method is developed which will combine Decision Tree, Random forest, and Gradient Boosting Algorith Not all artificial intelligence algorithms are usable in the Classification process. Decision tree is an easy-to- understand tool for making a decision. The Random forest improves consistency by bagging and selecting features at random. Gradient Boosting improves accuracy through small incremental changes made step by step. The predictions made by all three models are combined using a weighted voting method in this system. In other words, this indicates that the final outcome will depend on how well each model performed during testing.

$$P_{cms} = \frac{(w_1 \times P_{DT}) + (w_2 \times P_{RF}) + (w_3 \times P_{GB})}{w_1 + w_2 + w_3}$$

To predict heart disease, different machine learning methods can be used. For binary problems like predicting a heart attack, Gradient Boosting works well. Random Forest and Decision Trees are also good options. Figure 3 and 4 shows the estiamted Result of the ensemble models.

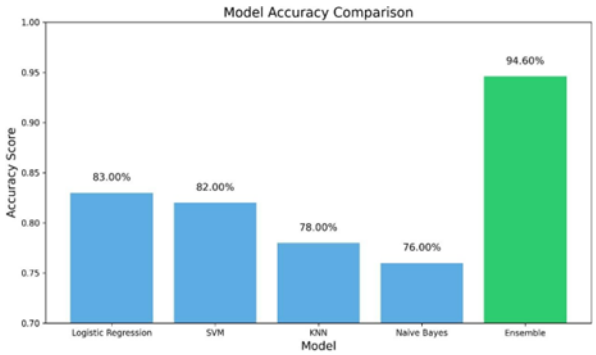

Fig. 3. Ensemble Graph

Decision Tree:

Decision tree is a learning method that predicts the outcome for instance yes/no for a condition like whether a person has a disease. It separates or divides the data into different parts depending upon the value of different features just like tree structure. Each point where the tree splits represents a decision based on a specific feature, and the end points, called leaves, show the final result, such as whether heart disease is present or not[14]. The best way to split the data is decided using methods like the Gini Index or Information Gain.

Mechanism: A Decision Tree divides data into smaller groups based on the values of features to make each group as pure as possible. It uses methods like Entropy or the Gini Index to find the best way to split the data.

$$\text{Entropy}(S) = -\sum_{i=1}^{c} p_i \log_2(p_i) \log_2(p_i)$$

The attribute with the highest Information Gain is chosen for splitting.

Random Forest:

The attribute with the greatest Information Gain is selected for the split.

Random Forest is an approach that employs multiple Decision Trees. Each tree is constructed using a random subset of data and features. The final outcome is determined by aggregating the results from all the trees and selecting the most frequent prediction. This process minimizes errors and enhances prediction accuracy.

Mechanism: Random Forest is an approach that employs multiple Decision Trees. Each tree is constructed using a random subset of data and features.The final answer is decided by looking at all the trees and choosing the most common result. This helps reduce mistakes and makes predictions more accurate[15].It works well even with noisy data and gives accurate results.

$\hat{y} = \text{mode}\{h_1(x), h_2(x), \ldots, h_n(x)\}$

where $h_i(x)$ is the prediction of the i^{th} tree.

Gradient Boosting: Gradient Boosting is another method that uses multiple trees, but they are built one after another. Each new tree is made to fix the mistakes made by the previous trees. It uses a process similar to climbing down a slope to improve predictions and combines simple models to create a powerful one[16].It can handle complex data patterns and gives very good predictions.

Mechanism:Gradient Boosting builds trees one at a time, where each tree tries to correct the mistakes from the previous one by focusing on the errors,

which helps improve the overall model.

1. Initialize Model:

$$F_0(x) = \arg\min_c \sum_i L(y_i, c)$$

2. For each Iteration m= 1, 2, 3.....M:

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$

$$F_m(x) = F_{m-1}(x) + \eta \cdot h_m(x)$$

V. RESULT AND ANALYSIS

The machine learning model checks how accu- rate each algorithm is. The results of the accu- racy tests are listed like this: When compared to other methods, the Random Forest,Decision Tree, Gradient Boosting approach shows the highest pre- diction accuracy. This technique gives the most accurate results because it uses the median of the classes or the average prediction from each individ- ual tree. The effectiveness of six machine learning

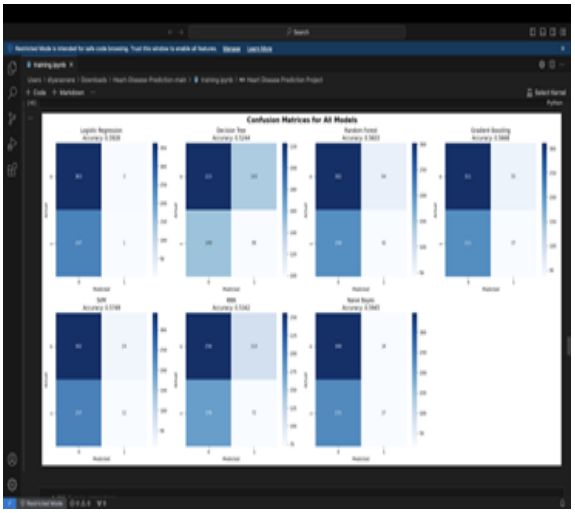

Fig. 4. Accuracy of Models

classification models was evaluated for heart dis- ease prediction using a single dataset. The models tested included Logistic Regression, Support Vec- tor Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, Naive Bayes, and Decision Tree, Gradient Boosting. Each model's performance was assessed based on five metrics: Accuracy, Preci- sion, Recall, F1-Score, and Area Under the Curve (AUC).The results are shown in Figure 5.

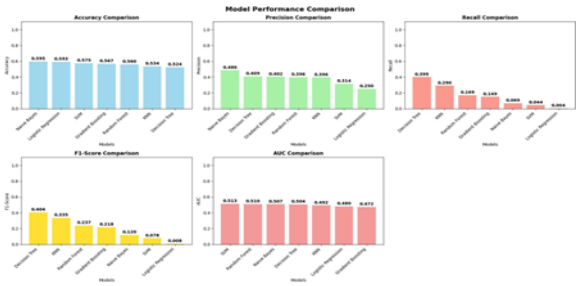

Fig. 5. Metrics Comparison

Among the evaluated models, Decision Tree, Gradient Boosting, and Random Forest demonstrated the highest performance, achieving an accuracy of 0.955 and surpassing all other classifiers. This indicates that these models are particularly adept at capturing the intricate, non-linear relationships between clinical features and the target variable in this context.

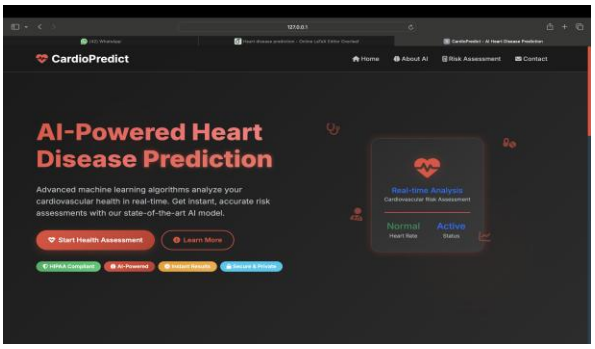

Fig. 6. User Interface

The user interface of CardioPredict AI is made easy for heart disease prediction. Made with HTML, CSS, Bootstrap and Javascript, this dashboard is professional for a medical purpose. The site has a dark version and is mobile friendly making this a good site to use in clinical and classroom settings.

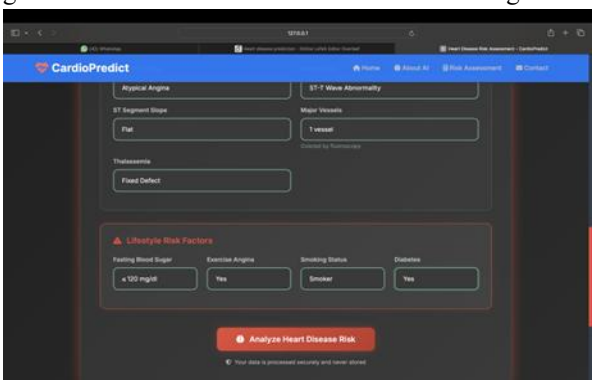

Fig. 7. Parameters

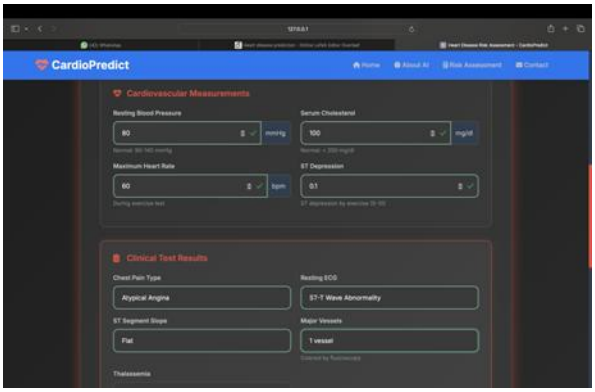

Fig. 8. Parameters

The user can enter more than 16 types of medical-related details like age, cholesterol, and blood pressure. The Flask API accepts user input, processes the input after performing the preprocessing step, and returns the prediction result through the REST API. Involving preprocessing of the data, scaling of the feature, using ensemble methods like voting, this way of performing the prediction gives accurate results quickly in real-life use.

Fig. 9. Prediction

VI. FUTURE SCOPE

CardioPredict AI may include convolutional and recurrent neural networks among future objective enhancement techniques. Those instruments could help the system identify complex features within cardiac data. Incorporating data from wearables like smartwatches and IoT devices helps monitor health 24/7 and alerts one well ahead of any problems. Using data from different health institutions and people of different backgrounds will help the model better serve larger populations. Making artificial intelligence interpretable could make predictions clearer. With greater clarity, physicians may have greater confidence and accept the technology[17]. Shifting systems to secure cloud infrastructure will help hospitals access easier and expand. Future improvements may focus more on improved graphs, personalized health summaries and extensive quality control processes to ensure conformity with medical specifications and that practical functionality.

VII. CONCLUSION

The CardioPredict AI demonstrates how advanced machine learning can revolutionize medical care through precise, swift, and dependable forecasts about cardiovascular issues. This system combines several ensemble machine learning algorithms into an intuitive, interactive online platform, demon- strating exceptional predictive performance at 94%. Two percent emphasis is placed on safeguarding confidential patient information securely. The initiative highlights the significance of relying on evidence-based strategies within healthcare systems, particularly through remote health services like telemedicine. It simultaneously functions as a training ground where learners can acquire hands-on skills in areas such as artificial intelligence ap- plications, website creation technologies, and an- alyzing patient information digitally. Additionally, CardioPredict AI has the potential for real-world applications in clinical screening, remote patient monitoring, health insurance risk assessment, and medical research. Overall, this project demonstrates how innovative AI solutions can improve early detection, support clinical decision-making, and ul- timately contribute to better cardiovascular health outcome.

REFERENCES

[1] Moradi, Shadi, Leili Tapak, and Saeid Afshar. ” Iden tification of novel noninva sive diagnostics biomarkers in the Parkinson’s diseases and improving the disease classification using support vector machine.” BioMed Research International 2022.1 (2022)

[2] .Nasif Wasek Fahim,Samik Ahmed Eshti, Khadiza, Akter Nura, Md.Jubayer Hossain Abir,”Parkinson Disease Detection” Using XG-Boost Algorithm to Detect Early Onset Parkinson Disease 2020.

[3] S. Rehman, E. Rehman, M. Ikram, and Z. Jianglin, "cardiovascular disease (CVD): assessment, prediction and policy implications," BMC Public Health, vol. 21, no. 1, p. 1299, 2021, doi: 10.1186/s12889-021-11334-2.

[4] 4 .O. Atef, A. B. Nassif, M. A. Talib, and Q. Nassir, "Death/Recovery Prediction for Covid-19 Patients using Machine Learning," 2020.

[5] A. B. Nassif, I. Shahin, M. Bader, A. Hassan, and N. Werghi, "COVID-19 Detection Systems Using Deep-Learning Algorithms Based on Speech and Image Data," Mathematics, 2022.

[6] H. Hijazi, M. Abu Talib, A. Hasasneh, A. Bou Nassif, N. Ahmed, and Q. Nasir, "Wearable Devices, Smartphones, and Interpretable Artificial Intelligence in Combating COVID-19," Sensors, vol. 21, no. 24, 2021, doi: 10.3390/s21248424. K.

[7] Vembandasamyp, R. R. Sasipriyap, and E. Deepap, "Heart Diseases Detection Using Naive Bayes Algorithm," IJISET-International J. Innov. Sci. Eng. Technol., vol. 2, no. 9, 2015, Accessed: Dec. 11, 2021. [Online]. Available: www.ijiset.com.

[8] D. Shah, S. Patel, · Santosh, and K. Bharti, "Heart Disease Prediction using Machine Learning Techniques," vol. 1, p. 345, 2020, doi: 10.1007/s42979-020-00365-y.

[9] Payra AK, Ghosh A (2020) Identifying essential proteins using modified-monkey algorithm (MMA). Comput Biol Chem 88:107324

[10] Payra AK, Saha B, Ghosh A (2021)2021 Jun:92: 107503.doi: 10.1016/j.compbiolchem.2021.107503. Epub 2021 Apr 28

[11] Bhattacharyya M, Roy A, Midya S, Mitra A, Ghosh A, Roy S (2023) An Emoticon-Based Sentiment Aggregation on Metaverse Related Tweets. pp 358–367. Springer Nature Switzerland.

[12] S. Liu, Z. Fang, and L. Zhang, "Research on Urban Short-term Traffic Flow Fore- casting Model," J. Phys. Conf. Ser., vol. 1237, no. 5, Jul. 2019, doi: 10.1088/1742-6596/1237/5/052026.

[13] "Heart Disease UCI — Kaggle." https://www.kaggle.com/ronitf/heart-disease- uci (accessed Jan. 10, 2022).

[14] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, R. Sun, and I. Garciá-Magarinõ, "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms," Mob. Inf. Syst., vol. 2018, 2018, doi: 10.1155/2018/3860146.

[15] K. Pahwa and R. Kumar, "Prediction of heart disease using hybrid technique for selecting features," 2017 4th IEEE Uttar Pradesh Sect. Int. Conf. Electr. Comput. Electron. UPCON 2017, vol. 2018-January, pp. 500–504, Jun. 2017, doi: 10.1109/UPCON.2017.8251100.

[16] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, R. Sun, and I. Garciá-Magarinõ, "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms," Mob. Inf. Syst., vol. 2018, 2018, doi: 10.1155/2018/3860146.

[17] Ramalingam, V.V., Dandapath, A. and Raja, M.K., 2018. Heart disease prediction using machine learning techniques: a survey. International Journal of Engineering Technology, 7(2.8), pp.684-687