Journal of Informatics and Web Engineering

Vol. 4 No. 1 (February 2025) eISSN: 2821-370X

Improved Accuracy for Heart Disease Diagnosis Using Machine Learning Techniques

Neeraja Joshi[1], Tejal Dave[2*]

[1,2]Sarvajanik University, Athwalines, Surat, Gujarat 395001, India.

**corresponding author: (tejal.dave@scet.ac.in; ORCiD: 0000-0003-2773-9748)*

Abstract - Accurate diagnosis of cardiovascular diseases (CVDs) is vital as people face many health issues due to CVD. Worldwide, more than 17 million people lose their lives each year due to CVD. This work primarily focuses on diagnosing heart disease before an explicit visit to the expert doctor. Machine learning-based systems have been found helpful in all applications, including medical ones, as they can learn human-like expert knowledge and utilize it subsequently. This work performs the classification of heart disease utilizing the subject's vital parameters. Ordinary people and patients need help understanding pathological laboratory results available after Testing and have to wait till they visit expert doctors for inference. In this paper, traditional methods like linear regression to various machine learning-based systems, including back propagation neural network, support vector machine (SVM), and k-nearest neighbor, are developed for heart disease classification. The proposed system (i) takes 13 vital parameters, including age, sex, chest pain type, fasting blood sugar, resting ECG, etc., as available from the Cleveland database, (ii) processes them with tuned machine learning systems, and (iii) transforms sensor inputs to stroke stage classification. To ascertain the proposed system's efficacy, all methods' performances are compared with similar work performed on the same standard- Cleveland database. Simulation results show 100 percent correct diagnosis and the robustness of SVM-based approaches for test data.

Keywords—Neural Network, Normalization, Classification, Support Vector Machine (SVM), k-Nearest Neighbor (KNN)

Received: 11 September 2024; Accepted: 09 November 2024; Published: 16 February 2025

1. INTRODUCTION

According to the WHO report [1], cardiovascular diseases (CVDs) are one of the major causes of death worldwide, with approximately 17.9 million deaths reported each year. The same report [1] adds that some common symptoms like obesity, hypertension, hyperglycemia, and high cholesterol increase the chances of heart-related diseases. New York Heart Association (NYHA) has derived functional Classification [2] for heart failure according to the severity of their symptoms. Medical history data indicate measurements of clinical parameters for the subject's body and are essential to diagnosing particular details about a given disease. Medical data can extract vital information about diseases from specific stored measurements.

Machine learning approaches have found many uses in domains like surveillance, sensor networks, mobile networks, health care, robotics, etc., for various tasks, including identifying earlier undetected patterns and generating control actions from various systems based on feedback received [3, 4]. In medical imaging, machine learning techniques are used to understand patients' characteristics for model estimation and classification, e.g., Normal and severe heart diseases. Once appropriately trained, machine learning techniques help efficiently classify a given dataset [5].

However, despite the impressive advances in machine learning, several problems still need to be solved. In general, difficulties arise because of uncertainties in the absence of adequate prior information about the surroundings, data range-related limitations and calibrations of cameras or sensors, adverse observation of data being acquired, etc. These all may lead to unpredictability in the values of input variables given to the intelligent systems. Other problems are related to the adaptability of machine learning algorithms to react to unknown environmental conditions that change with timings and classify them correctly.

Various approaches are found in the literature to solve the abovementioned challenges in classifying heart diseases using machine learning. Some approaches focus on logistic regression and decision trees [6]. In [7], comparisons are carried out to evaluate the performance of different machine learning classifiers, such as KNN, SVM, decision trees, random forests (RF), etc. In [8], different ML algorithms were used to identify cardiac abnormalities. Important features were selected using PCA and regression techniques in [9]. Multilayered feed-forward neural network (MFNN) and back propagation neural network (BPNN) algorithms for the prediction of heart disease in four stages [10]. They used a dataset available at the University of California, Irvine [UCI] and achieved an accuracy of 92%. Similar work was done by authors in [11] using WEKA software—several approaches in which classification is done with heart rate variability measured in time and frequency domain. In [12], a cloud-based decision support system was developed using machine learning. A survey was presented in [13] to compare ML methods for diagnosing the heart. Prediction of the vulnerability of a heart disease given primary symptoms was discussed in [14] . A method for detecting chronic heart failure from heart sounds [15] was presented that used four stages to achieve a classification accuracy of up to 92%. Ensemble Machine Learning Techniques were compared on the Cleveland database in the work [16]. The SVM-based method [17] achieved 92.30 % accuracy. At the same time, heart disease prediction through online consultation is discussed in [18], random forest [19], and ensemble learning [20, 21] were also used, but the accuracy for predicting heart conditions needed many improvements. Comparative Machine Learning Models for other applications is presented in [22, 23].

This paper presents various ML-based techniques for the classification of heart diseases. As pathological results are available, there is no need to use deep learning models to keep them light. Inputs are data obtained from patients' medical tests as advised by medical experts. Initially, data is trained with an extensive training dataset. The classification of heart diseases can be done in two steps of the algorithm: (a) training with a large enough dataset and (b) test pairs, input records extracted from a total data set, and exclusively used for testing purposes. The efficacy of the proposed models is shown by comparing the detection accuracy of proposed models to similar work on the Cleveland dataset.

This paper is organized as follows: Section 2 provides the basics of the dataset and our methodology. Section 3 provides the details of the ML methods that we have used. Section 4 demonstrates and discusses simulation results, followed by a conclusion.

2. LITERATURE REVIEW

This section describes various approaches used for heart disease diagnosis in the recent past. We present a multiple linear regression approach that performs Heart Disease Diagnosis using a regression network created using input-output training pairs. A back propagation neural network approach is proposed in which structural parameters of NN, i.e., weights (Ws) plus biases (Bs), are trained using input-output training pairs. An SVM approach is also presented that approximates f(x), functional mapping between input and Output that has at most ε deviation from the desired target data.

2.1 Multi-Linear Regression Model for Heart Disease Diagnosis

Multiple regression problems can be considered as an extension of simple linear regression. MLR helps define a model that relates dependent parameters (output variable) and many independent (predictor) variables. The addition of each input variable creates more relationships among them. In the MLR problem, the input-output relation is given by Equation (1)

Where,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + ... + \beta_p X_p + \varepsilon \tag{1}$$

Y- is called the Output X_1 X_2, X_3 ... X_p - - are inputs, β_0 -intercept, β_1 β_2 β_3 β_p constant slopes, and ε - Prediction error. Estimated multiple linear regression is shown in Equation (2)

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + ----+ b_p X_p \tag{2}$$

Where, $\hat{Y}$ is the predicted value of dependent variables b_o, b_1, b_2,...b_p are the estimates of β_1 β_2, β_3 and β_p.

2.2. Neural Model for Heart Disease Diagnosis

Neural networks have been extensively used in various applications due to their remarkable generalization capabilities once appropriately trained. In this section, we describe our proposed neural-based system for classifying heart disease levels with the help of 13 input attributes derived from the Cleveland database. The proposed algorithm overcomes the shortcomings of earlier methods regarding the learning mechanism used. We consider back propagation neural network (BPNN) architecture, as shown in Figure 1.

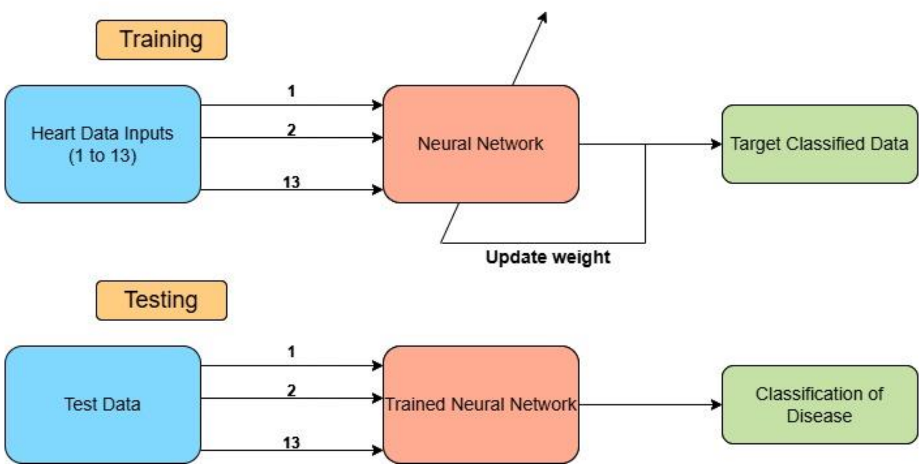

Figure 1 Proposed Neural Network-Based System

The framework contains two phases, namely, training and Testing. The First NN learning phase is carried out utilizing a well-defined and standardized database for the extraction of sufficient numbers of training pairs. This will help to overcome training-related problems that many existing approaches face. In the first training stage, the Input layer is supplied with one of the patterns from the training set. At each layer, computations are done during the forward pass, and actual Output is derived at the end. Henceforth, MSE is computed between desired and actual outputs and backpropagated to the previous layers. Generally, the computed local gradient updates the connection weights. This process is carried out until the input layer is reached. Details of architecture of Proposed Neural Network-Based System for Heart Diagnosis System is shown in Figure 2.

2.2.1 Selection of Parameters for NN Training

2.2.1.1 Identification of The Required Number of Neurons

The computation of the required number of neurons in the middle-hidden layer plays a vital role in successful training. The literature presents a number of heuristics, and one standard way is to select a number equal to the mean of the number of input layer and output layer neurons. If the desired training performance is achieved, stop training; otherwise, repeat training by increasing/decreasing the number of neurons in the hidden layer.

The selection of a proper transfer function is vital in NN, as the required non-linear mapping between desired Output and actual Output can be derived from it. In our work, a 'sigmoid' for the hidden layer and a 'linear' function are used to classify output values.

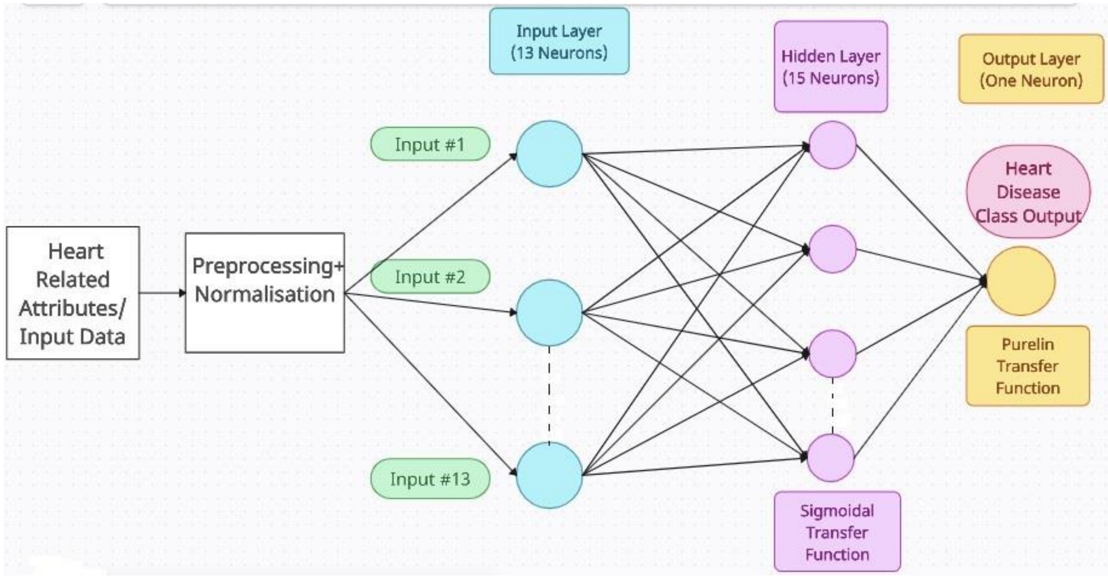

Figure 2 Architecture of Proposed Neural Network-Based System for Heart Diagnosis System

2.2.1.2 Selection of Learning Function and Training Mode

'Trained' is one of the network training functions to update design parameters. It obeys a gradient descent adaptive rule to control MSE. Two training modes are available: (a) incremental training and (b) batch mode. For prediction (classification) problems, we have used batch mode.

2.2.1.3 Initialization of Learning Parameters

Any search algorithm's success depends upon its starting point section. Any algorithm started from a nearby point in the target region will achieve the destination quickly. It should be noted that all zeroes (weights and bias) are also biased points in NN training that can cause catastrophic failure during draining. Hence, we have initialized with a 'random' function for design parameters.

2.2.1.4 Selection of Evaluation Parameter

As most learning algorithms inherently require differentiability, the obvious choice is either MSE or SSE.

2.3 SVM-based Model for Heart Disease Diagnosis

SVM is a supervised training algorithm that uses a model to approximate a function that maps from input real numbers to an output class based on training data. It assigns new examples to either one category or the other. The function of SVMs can be efficiently extended to perform a non-linear classification using proper kernels. Suppose the training set is represented as Equation (3)**Error! Reference source not found.**.

$$D = \{(x_1, y_1), (x_2, y_2), \ldots (x_m, y_m)\} \tag{3}$$

Where xi is the n- vector, yi is the Output for each xi.

The function F(x) relates xi to yi, which is given by Equation **Error! Reference source not found.**

$$F(x) \Rightarrow w^T * x_i + b \tag{4}$$

Where w and b are weights and bias, respectively.

The objective can be defined as the correct estimation of design parameters (w and b) that best fit the data. The condition must be met for every x, as per Equation (5).

$$\begin{aligned} y_i - w^T x_i - b &\le \varepsilon \\ w^T x_i + b - y_i &\le \varepsilon \end{aligned} \tag{5}$$

The training in SVR becomes for soft constrained optimization problems for allowing error and dealing with some noise, as shown in Equations (6) and (7).

$$Minimize : \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{P}(\xi + \xi^*) \tag{6}$$

$$\begin{aligned} &subject\_to \\ &yi - w^T xi - b \le \varepsilon + \xi i; i = 1,....P \\ &w^T xi + b - yi \le \varepsilon + \xi i^*; i = 1,....P \end{aligned} \tag{7}$$

Where ϵ and ξ >=0;i=1, P.

2.4 K Nearest Neighbour (kNN)-based Model for Heart Disease Diagnosis

The K Nearest Neighbour (kNN) is found in literature as an intuitive method to classify unlabelled examples by utilizing their similarity with already known examples in the given training set. Requirements for the KNN algorithm are an integer k-subset, a set of labeled examples (training data), and an evaluation parameter to measure "closeness."

$$\left\|\mathbf{x}^a - \mathbf{x}^b\right\|_2 = \sqrt{\sum_{j=1}^{d}(x_j^{(a)} - x_j^{(b)})} \tag{8}$$

Distance is typically defined per Euclidean norm in Equation (8).

3. RESEARCH METHODOLOGY

In order to compare the performance of various models proposed in an earlier section, we have used the same input-output cardinality for simulating results using MLR, BPNN, KNN, and SVM-based models. The Standard Cleveland dataset, comprising 294 records, is used. Two hundred seventy data points were extracted, normalized, and used as training. There are two layers: the hidden layer comprises 15 neurons, whereas the output layer contains a single neuron. The number of neurons in each layer was decided by trial and error. Two hundred seventy data points were extracted from the Cleveland dataset, normalized, and used as training pairs for feed-forward back propagation type neural networks. The training and adaptation learning functions used are 'trained’ and ‘learned,' respectively. There are 13*15 + 15*1 + 15 + 1= 226 design parameters that comprise weights and biases needed for system training. The evaluation function used is MSE. The network training and summary of various parameters used are shown in Table 1. After training was completed after 10,000 iterations, MSE was reduced to 0.00123022.

For experimenting with the SVM-based method, two design parameters, Cost (C) and precision epsilon, are essential during the training of the record. The penalty's cost should have a value between 0 and infinity. Increasing cost value causes closer fitting to the calibration/training data. The cost value is approximately decided by utilizing the output response characteristics from the training data, which can be approximated by equation (10). In experimentation, a based model for classifying heart diseases was trained as a combination of the two class models. Table 1 shows the class-wise training models used and their respective MSE. While testing SVM as a model for the classification of heart disease gives 100 % performance.

Table 1. Simulation Parameters for Training Back Propagation Neural Network

Total number of training pairs	270
Type	Feed forward back propagation
Training Function	Tracing
Adaption learning function	Learngdm
Performance function	MSE
No. of Layer	2
Number of neurons in the input layer	13
Number of hidden neurons	15
Number of neurons in the output layer	1

4. RESULTS AND DISCUSSIONS

Two feed-forward back propagation network layers are considered for implementing a neural-based model, as shown in Figure 3. The simulation setup's parameters described in Table 2 are considered. The architecture consists of two layers: the hidden layer comprises 15 neurons, and the output layer contains a single neuron. Regarding the number of neurons, experimentation was carried out with different numbers of neurons, like first 10, then 15, and then 20 neurons in the hidden layer. For each experiment, several trials were made. A network type with 15 hidden layer neurons gave min MSE and was further used.

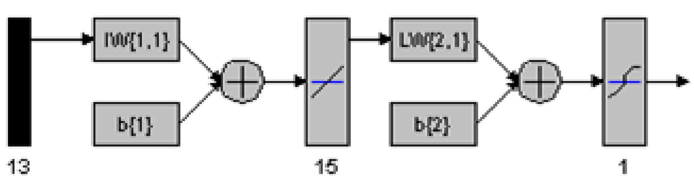

Figure 3 Snapshot of Neural Network Training with Various Parameters

Two design parameters, Cost (C) and precision epsilon, are essential during the record's training to experiment with the SVM-based method. The cost value is approximately decided by utilizing the output response characteristics from the training data, which can be approximated by Equation (9).

$$C = \max(\bar{y} + 3 * \sigma_y, \bar{y} - 3 * \sigma_y) \tag{9}$$

It includes the mean and standard deviation of training data (y). The second parameter, epsilon, depicts the required precision value. Increasing epsilon decreases the number of support vectors, yielding a flat response. As given in [42], the practical prescription for epsilon can be given as Equation (10).

$$\varepsilon = 3\sigma\sqrt{\frac{\ln n}{n}} \tag{10}$$

Where n is the size of the training dataset.

In experimentation, an SVM-based model for classifying heart diseases was trained as a combination of the two class models. Table 2 shows the class-wise training models used and their respective MSE. Testing SVM as a model for classifying heart disease gives 100 % performance.

Table 2. Class-Wise Performance Analysis Using SVM-based Model

Sr. No	Name of Algorithm	Data class	No samples for training	Average MSE	No samples for Testing	Success Rate
01	SVM	0 and 1	200	9.38*10-4	14	100%
02	SVM	0 and 2	203	9.77*10-4	13	100%
03	SVM	0 and 3	201	9.28*10-4	13	100%

After successfully checking individual model results, in order to have fair comparisons, we compared the performances of all methods discussed earlier on the Cleveland dataset. Table 3 details the performance of all methods for extracting 24 test records from the Cleveland database. The first column shows exact record numbers used for comparison, and the second column details the desired class (ground truth values) available. Subsequent columns depict performances of various classification methods under test. Simulation results reveal that MLR methods generally fail for intermediate class 2 and achieve overall poor performance. Similarly, KNN also underperforms in the correct classification in specific classes. BPNN, Naive, and SVM models successfully classify heart disease.

Table 3. Comparison of Various Methods on a Few Test Records From The Cleveland Dataset

Cleveland Dataset	**Ground Truth**	**Output Class Prediction**			
		LR	**k-NN**	**BPNN**	**SVM**
Record #004	0	0	0	0	0
Record #005	0	0	0	0	0
Record #006	0	0	0	0	0
Record #007	0	0	0	0	0
Record #008	0	0	1	0	0
Record #181	1	1	1	1	1
Record #182	1	1	1	1	1
Record #183	1	1	1	1	1
Record #184	1	1	1	1	1
Record #215	2	1	2	2	2
Record #216	2	1	2	2	2
Record #217	2	1	2	2	2
Record #281	3	2	2	3	3
Record #282	3	3	2	3	3
Record #283	3	2	2	3	3
Record #284	3	2	2	3	3

To demonstrate the efficacy of our proposed work, we have compared the success rate of our proposed BPNN, naive-based, and SVM-based methods in heart disease classification with that of similar work in the literature, as reported in Table 4. Again, we have compared methods that were tested on Cleveland datasets. Table 4 shows the efficiency of our work compared to similar work. Figure 4 shows a bar chart comparing our work with other works.

For evaluation, we have used 'Success rate' as a performance evaluation parameter. It is interpreted as ratio of successful detection of an event (heart disease present or absent) to the total number of test samples presented to perform a task or procedure. It can be used to measure the effectiveness of a task or process.

Table 4. Comparison with Similar Work on the Cleveland Dataset

No	Name of Classifier	# Samples for training & testing	Success Rate
01	Logistic Regression [6]	50% training 50%testing	84.80%
02	SVM [17]	10-fold cross-validation	92.30%
03	Random Forest [19]	70% training 30%testing	85.81%
04	BPNN (Proposed)	270 training,24 testing	95.83%
05	SVM (Proposed)	270 training,24 testing	100%

Figure 4. Comparison with Similar Work on the Cleveland Dataset

5. CONCLUSION

In this paper, we analyzed the performance of different machine-learning techniques to classify heart disease in patients. In order to compare the performance of each classifier, we used the same standard Cleveland dataset and kept the number of training and testing pairs the same. We analyzed five classifiers: MLR, Random Forest, BPNN, KNN, and SVM. Compared to other work, we classified heart disease into four classes when only two were discussed. It was asserted that using normalization as a pre-processing step improves the performance of machine learning techniques. Our proposed work gives better accuracy measures than similar work. At the same time, among various classifiers, we analyzed SVM-based techniques that performed better than MLR, KNN, and BPNN-based techniques in our experimentation.

ACKNOWLEDGEMENT

We thank the anonymous reviewers for the careful review of this article.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Neeraja Joshi: Conceptualization, Data Curation, Methodology, Validation, Writing –Original Draft preparation;

Tejal Dave: Project Administration, Supervision, Writing –Review & Editing;

CONFLICT OF INTERESTS

No conflicts of interest were disclosed.

ETHICS STATEMENTS

The paper follows The Committee of Publication Ethics (COPE) guideline.

REFERENCES

[1] "Cardiovascular diseases (CVDs) Retrieved from," [Online]. Available: https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1 . [Accessed on 10 Sept , 2024)].

[2] The New York Heart Association (NYHA), "Classes of heart failure Retrieved from," [Online]. Available: https://www.heart.org/en/health-topics/heart-failure/what-is-heart-failure/classes-of-heart-failure . [Accessed 7 Jun (Last Reviewed: , 2023)].

[3] S. Haykin, Neural Network - a comprehensive foundation, 2nd Edition, Asia,Pearson Education,Prentice Hall International, 1998 .

[4] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. W. Quinn, and M. A. Moni, “Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison,” Computers in Biology and Medicine, vol. 136, p. 104672, Sep. 2021, doi: https://doi.org/10.1016/j.compbiomed.2021.104672.

[5] M. M. Alsaleh et al., “Prediction of disease comorbidity using explainable artificial intelligence and machine learning techniques: A systematic review,” International Journal of Medical Informatics, vol. 175, p. 105088, Jul. 2023, doi: https://doi.org/10.1016/j.ijmedinf.2023.105088.

[6] D. Khanna, R. Sahu, V. Baths, and B. Deshpande, “Comparative Study of Classification Techniques (SVM, Logistic Regression and Neural Networks) to Predict the Prevalence of Heart Disease,” International Journal of Machine Learning and Computing, vol. 5, no. 5, pp. 414–419, Oct. 2015, doi: https://doi.org/10.7763/ijmlc.2015.v5.544.

[7] R. Radhika and S. Thomas George, “HEART DISEASE CLASSIFICATION USING MACHINE LEARNING TECHNIQUES,” Journal of Physics: Conference Series, vol. 1937, no. 1, p. 012047, Jun. 2021, doi: https://doi.org/10.1088/1742-6596/1937/1/012047.

[8] K. M. Mohi Uddin, R. Ripa, N. Yeasmin, N. Biswas, and S. K. Dey, “Machine learning-based approach to the diagnosis of cardiovascular vascular disease using a combined dataset,” Intelligence-Based Medicine, vol. 7, p. 100100, Jan. 2023, doi: https://doi.org/10.1016/j.ibmed.2023.100100.

[9] T. Santhanam and E. P. Ephzibah, “Heart Disease Classification Using PCA and Feed Forward Neural Networks,” Lecture notes in computer science, pp. 90–99, Jan. 2013, doi: https://doi.org/10.1007/978-3-319-03844-5_10.

[10] A. Aaditya., N. Sharad, M. Rahul, S. Atharva, and A. Shubham, “Heart Disease Prediction System using Multilayered Feed Forward Neural Network and Back Propagation Neural Network,” International Journal of Computer Applications, vol. 166, no. 7, pp. 32–36, May 2017, doi: https://doi.org/10.5120/ijca2017914080.

[11] S. K. Sen, “Predicting and Diagnosing of Heart Disease Using Machine Learning Algorithms,” International Journal Of Engineering And Computer Science, Jun. 2017, doi: https://doi.org/10.18535/ijecs/v6i6.14.

[12] E. Maini, Bondu Venkateswarlu, and A. K. Gupta, “Applying Machine Learning Algorithms to Develop a Universal Cardiovascular Disease Prediction System,” Lecture notes on data engineering and communications technologies, pp. 627–632, Aug. 2018, doi: https://doi.org/10.1007/978-3-030-03146-6_69.

[13] V. V. Ramalingam, A. Dandapath, and M. Karthik Raja, “Heart disease prediction using machine learning techniques : a survey,” International Journal of Engineering & Technology, vol. 7, no. 2, pp. 684–687, 2018.

[14] A. Gavhane, G. Kokkula, I. Pandya and K. Devadkar, "Prediction of Heart Disease Using Machine Learning," 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2018, pp. 1275-1278, doi: 10.1109/ICECA.2018.8474922.

[15] M. Gjoreski, M. Simjanoska, A. Gradišek, A. Peterlin, M. Gams and G. Poglajen, "Chronic Heart Failure Detection from Heart Sounds Using a Stack of Machine-Learning Classifiers," 2017 International Conference on Intelligent Environments (IE), Seoul, Korea (South), pp. 14-19, 2017, doi:10.1109/IE.2017.19.

[16] S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia and J. Gutierrez, "A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease," 2017 IEEE Symposium on Computers and Communications (ISCC), Heraklion, Greece, 2017, pp. 204-207, doi:10.1109/ISCC.2017.8024530.

[17] F. S. Alotaibi, “Implementation of Machine Learning Model to Predict Heart Failure Disease,” International Journal of Advanced Computer Science and Applications, vol. 10, no. 6, 2019, doi: https://doi.org/10.14569/ijacsa.2019.0100637.

[18] B. Duraisamy, R. Sunku, K. Selvaraj, V. V. R. Pilla, and M. Sanikala, “Heart disease prediction using support vector machine,” Multidisciplinary Science Journal, vol. 6, pp. 2024ss0104–2024ss0104, 2024, doi: https://doi.org/10.31893/multiscience.2024ss0104.

[19] Y. K. Singh, N. Sinha, and S. K. Singh, “Heart Disease Prediction System Using Random Forest,” Communications in Computer and Information Science, pp. 613–623, 2017, doi: https://doi.org/10.1007/978-981-10-5427-3_63.

[20] A. Rahim, Y. Rasheed, F. Azam, M. W. Anwar, M. A. Rahim, and A. W. Muzaffar, “An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases,” IEEE Access, vol. 9, pp. 106575–106588, 2021, doi: https://doi.org/10.1109/access.2021.3098688.

[21] C. Pan, A. Poddar, R. Mukherjee, and A. K. Ray, “Impact of categorical and numerical features in ensemble machine learning frameworks for heart disease prediction,” Biomedical Signal Processing and Control, vol. 76, p. 103666, Jul. 2022, doi: https://doi.org/10.1016/j.bspc.2022.103666.

[22] T. A. Khan, R. . Sadiq, Z. . Shahid, M. M. Alam, and M. B. . Mohd Su’ud, “Sentiment Analysis using Support Vector Machine and Random Forest”, Journal of Informatics and Web Engineering, vol. 3, no. 1, pp. 67–75, Feb. 2024, https://doi.org/10.33093/jiwe.2024.3.1.5

[23] A. T. Jing Sheng, Z. Che Embi, and N. Hashim, "Comparison of Machine Learning Methods for Calories Burn Prediction", Journal of Informatics and Web Engineering, vol. 3, no. 1, pp. 182–191, Feb. 2024, https://doi.org/10.33093/jiwe.2024.3.1.12

BIOGRAPHIES OF AUTHORS

	**Neeraja Joshi** is currently studying B.Tech (Computer Engineering with Specialization in Data Science) at Sarvajanik College of Engineering and Technology, Sarvajanik University, India. She is planning to further pursue her master's studies in reputed institute. She is an active IEEE student member. Her research focuses on Computer Vision, Data Science and Bio medical signal processing technology. She can be contacted at email: neerajajoshi.co21d1@scet.ac.in.
	**Tejal Dave** is a Senior Assistant Professor in Instrumentation and Control Engineering at Sarvajanik College of Engineering and Technology, Sarvajanik University, India. Her research area of interest are Biomedical Instrumentation and Signal Processing, Industrial Instrumentation and Automation etc. She can be contacted at email: tejal.dave@scet.ac.in.