science, engineering and health studies

Research Article

Science, Engineering and Health Studies
https://li01.tci-thaijo.org/index.php/sehs
ISSN (Online): 2630-0087

Intelligent healthcare for cardiac patients utilizing neural networks, *k*-means clustering, and ad hoc routing

Annwesha Banerjee Majumder[1], Somsubhra Gupta[2], Sourav Majumder[3], and Dharmpal Singh[4]*

[1] Department of Information Technology, JIS College of Engineering, West Bengal 741235, India
[2] Department of Computer Science, Sami Vivekananda University, West Bengal 700121, India
[3] Camgemini Technology Services India Limited, West Bengal 700156, India
[4] Department of Computer Science and Engineering, JIS College of Engineering, West Bengal 741235, India

***Corresponding author**:*
Annwesha Banerjee Majumder
annwesha.banerjee@jiscollege.ac.in

***Received**: 27 August 2023*
***Revised**: 30 November 2023*
***Accepted**: 3 January 2024*
***Published**: 2 October 2024*

***Citation**:*
Majumder, A. B., Gupta, S., Majumder, S., and Singh, D. (2024). Intelligent healthcare for cardiac patients utilizing neural networks, k-means clustering, and ad hoc routing. Science, Engineering and Health Studies, 18, 24050007.

ABSTRACT

This paper introduced a comprehensive healthcare system designed to address the unique challenges faced by cardiac patients in economically and geographically constrained regions. The system operated in three phases, each contributing to the effective care of patients. In the initial phase, an artificial neural network model identified potential cardiac patients with an impressive accuracy of 90.16%, demonstrating its potential for early detection. The second phase employed *k*-means clustering to categorize patients into three groups based on the severity of their condition, facilitating precise prognosis and stratification. Finally, the system utilized an innovative ad hoc routing algorithm to securely transmit patient data to remote servers, enabling expert monitoring and consultation. The study's outcomes demonstrate the system's ability to accurately identify at-risk patients, appropriately categorize their condition, and efficiently route critical data. This holistic approach leverages cutting-edge technologies and methodologies to transform healthcare delivery in underserved areas. The novel system presents a promising avenue for enhancing cardiac care in regions with limited access to advanced healthcare services, ultimately improving patient outcomes and reducing disparities in cardiovascular healthcare.

Keywords: ad hoc routing; artificial neural network; *k*-means clustering; cardiac patient; healthcare

1. INTRODUCTION

Several studies have explored patient clustering for improved healthcare management. Lau et al. (2022) proposed a model for clustering COVID-19 patients based on blood test results. The clustering was conducted using *k*-prototype clustering. Stoitsas et al. (2022) proposed a model to cluster trauma patients, considering both statistical and machine learning metrics. It was observed that clusters obtained through the Bayesian method and deep Gaussian mixture model provided the best results. Majumder et al. (2022) proposed a geo-replicated patient monitoring system using an intelligent, energy-efficient BAN routing algorithm. They proposed a remote patient monitoring system, where patients were clustered into extremely critical, most critical, critical, moderate, and low-risk groups using *k*-means clustering. Additionally, a BAN routing algorithm was proposed where participating nodes were clustered into extremely competent, most competent, competent, moderate, and least competent nodes (Majumder et al., 2021a). Sharma et al. (2020) proposed a model for heart disease prediction using naïve Bayes (NB), decision trees, random forest, and support vector machine (SVM). Random forest achieved the highest accuracy in this model. A model to forecast cardiovascular disease was proposed by Rajdhan et al. (2020) by including NB, decision trees, logistic regression, and random forest. Ramprakash et al. (2020) proposed a deep learning model for predicting heart disease, applying S2 statistics to eliminate overfitting and underfitting. The accuracy levels attained using SVM, *k*-nearest neighbor (KNN), and decision trees were not satisfactory, leading the authors to apply artificial neural networks (ANNs) for accuracy enhancement (Pasha et al., 2020). Using convolutional neural network (NN), Sajja and Kalluri (2020) proposed a heart disease prediction model, comparing it with logistic regression, KNN, NB, SVM, and NNs. A COVID-19 prediction model utilizing logistic regression was proposed by Majumder et al. (2021b). Shao et al. (2022) proposed a mechanism for clustering chronic cough patients using deep unsupervised learning mechanisms on a dataset of 264,146 patients' EMRs. Another model was proposed applying supervised learning methods such as KNN, NB, decision tree, and random forest, with KNN receiving the highest priority (Shah et al., 2020). Muhammad et al. (2020) proposed a combination of different machine learning techniques for the early and accurate diagnosis of heart illness, including KNN, decision tree, logistic regression, random forest, and SVM. Jindal et al. (2021) proposed a cardiac disease prediction model using logistic regression and KNN, claiming better performance than applying logistic regression, decision tree, KNN, NB, random forest, and SVM, Patel et al. (2019) proposed a model for cardiovascular disease prediction. Mohan et al. (2019) proposed a model for heart disease prediction, incorporating random forest, NN, decision tree, and SVM. Rajesh et al. (2018) suggested a model utilizing NB and decision tree, where NB performed better on short datasets and decision tree performed better on large datasets. Haq et al. (2018) proposed a methodology applying different machine learning methods such as ANN, decision tree, SVM, and NB. Zheng et al. (2017) proposed an efficient model for type 2 diabetes detection using different machine learning methodologies. An energy-efficient routing algorithm for BAN was presented by Rakhee and Srinivas (2016) using ANT colony optimization and breadth-first search, employing a cluster-based strategy. Gupta and Banerjee (2015) proposed a model to determine the degree of risk for cardiovascular disease, where lifestyle and heredity were the main determinants. Murthy and Garcia-Luna-Aceves (1996) proposed a wireless routing protocol where the distance and second-last hop count were communicated by nodes, with a comparative analysis presented against Bellman-Ford algorithm (DBF), DUAL (a loop-free distance-vector algorithm), and an ideal link-state algorithm (ILS).

Several studies have predicted diseases or provided support for monitoring. This innovative work predicts and categorizes heart disease severity, enabling tailored monitoring for patients. This integrated method has not yet been widely published in the literature. A key aspect is the seamless transfer of patient data to a remote server for real-time monitoring by medical experts. This integrated approach revolutionizes cardiovascular health management, offering a proactive and personalized solution. The objective of the proposed model is to bridge the gap between advanced technology and the critical healthcare needs associated with cardiac diseases. By doing so, the work strives to improve disease management, enhance the accuracy of diagnosis and classification, and extend healthcare services to underserved populations, ultimately leading to better patient outcomes and reduced healthcare costs. This technique helps individuals who are economically and geographically disadvantaged to receive healthcare, thereby improving heart disease management in medicine. This proposed model of managing cardiac disease encompasses the integration of advanced technologies with a crucial and significant healthcare issue, making the work interesting. The prospect of saving lives, enhancing healthcare accessibility, and mitigating healthcare expenditures renders it an appealing and valuable domain of investigation and advancement. The exploration and advancements in the realm of cardiac illness management possess the capacity to provide significant beneficial outcomes. Enhanced patient outcomes, resulting from timely and precise disease detection, have the potential to augment the quality of life and potentially reduce mortality rates. Furthermore, the implementation of sophisticated technologies has the potential to decrease healthcare expenditures by preventing problems and avoiding expensive treatments. The integration of remote monitoring and support mechanisms has the potential to expand healthcare services to marginalized communities, hence promoting fairness in the availability of medical care. This study also makes a significant contribution to the progress of data-driven decision-making and multidisciplinary collaboration among medical practitioners and data scientists, fostering innovation in healthcare technology and methodology. Investigations conducted in this field have the potential to facilitate the advancement of preventive measures and timely interventions, which results in enhanced well-being among people and general enhancement of healthcare methodologies.

2. MATERIALS AND METHODS

The proposed model consists of three segments outlined as follows:

1. The disease prediction model used an ANN to achieve accurate cardiac disease prediction based on a UCI dataset with selected attributes. Data standardization was performed using StandardScaler.
2. Cardiac patients were clustered into three groups—extremely severe, severe, and moderate—using *k*-means clustering after disease classification.
3. An ad hoc routing algorithm was employed to securely share vital patient information with a remote server, ensuring continuous monitoring and personalized care for cardiac patients, especially in remote or economically challenged areas.

A block diagram of the proposed work is depicted in Figure 1.

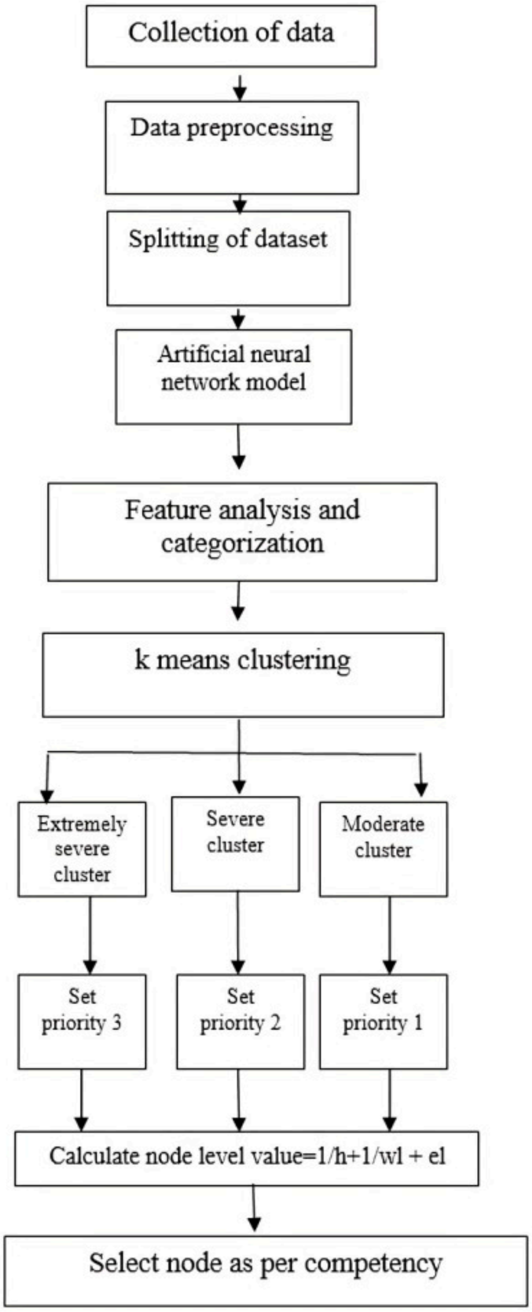


Figure 1. Block diagram of the proposed model

2.1 Data collection

The dataset used in this proposed model was sourced from the UCI data repository, combining four separate datasets from Cleveland Clinic Foundation, V.A. Medical Center in Long Beach, V.A. Medical Center in Zurich, Switzerland, University Hospital in Basel, Switzerland, and Hungarian Institute of Cardiology (Janosi et al., 1988).

2.2 Dataset description

The collected dataset comprises 14 features: age, sex, cp, tresbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, thal, and ca, and the dependent feature target. The distribution of these attributes can be observed in the the histogram shown in Figure 2.

2.3 Feature selection

Feature selection or dimensionality reduction is a crucial phase in any machine learning-based model, as it eliminates irrelevant features and enhances model training. The Extra Tree Classifier, an ensemble method that combines many de-correlated decision trees and is known for its computational efficiency, has been utilized in this model. The eight most relevant features selected through this method are cp, ca, thal, exang, thalach, oldpeak, age, and slope. Feature importance is shown in Figure 3.

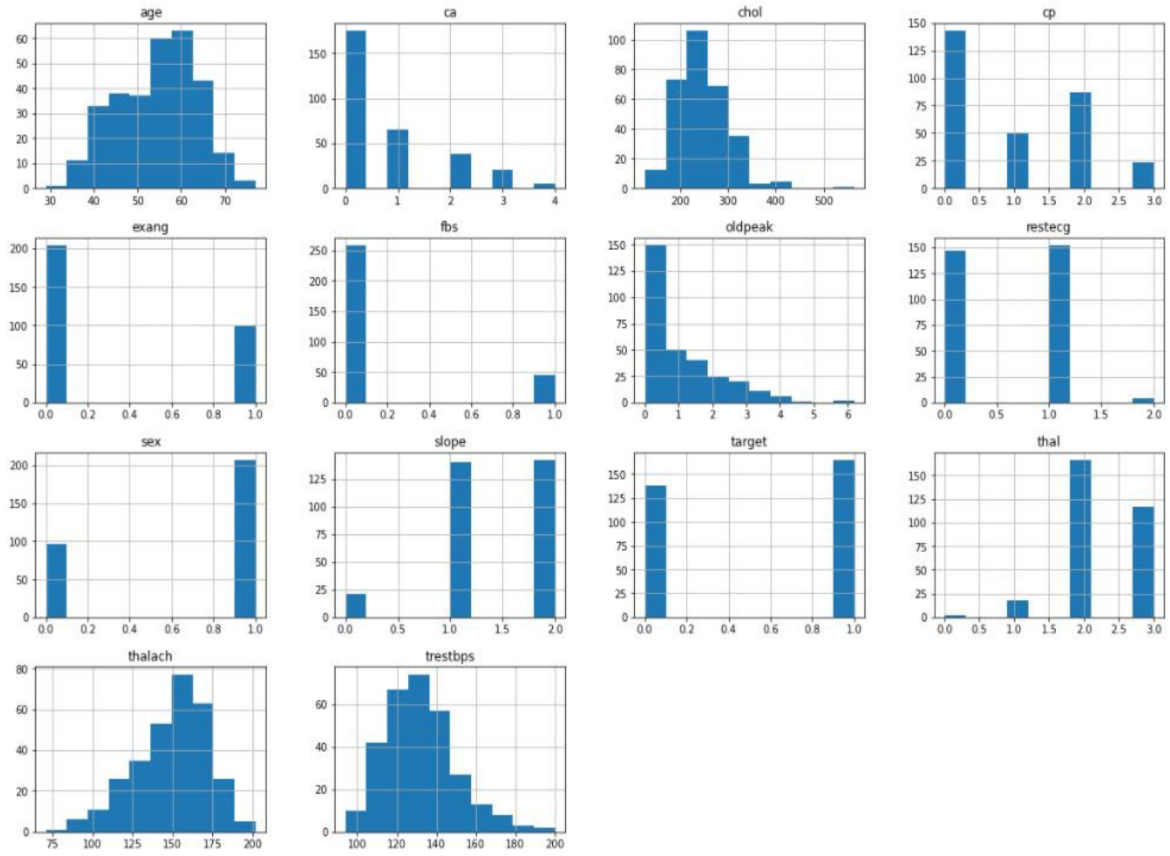


Figure 2. Dataset histogram

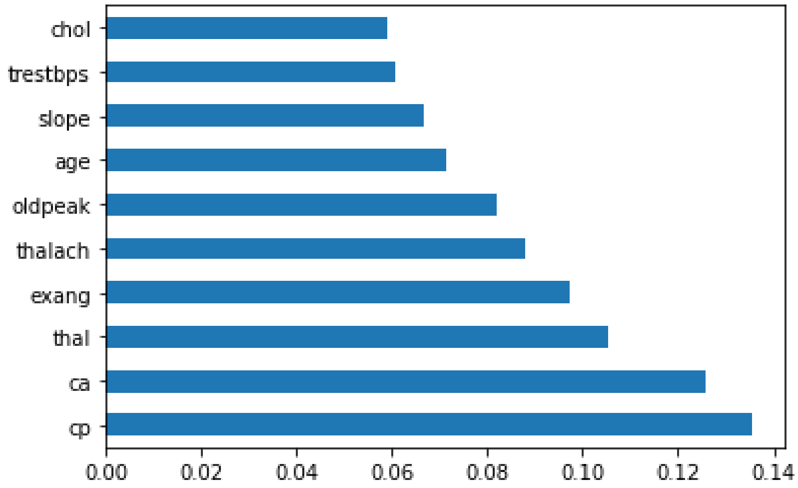


Figure 3. Feature selection
Note: The selected data set is categorized into training and test samples in a 3:1 ratio.

2.4 Data standardization

Data standardization is performed to scale data into a standard format, transforming mean to zero and standard deviation to one. This aids in better model training. The standardization equation (Equation 1) is utilized.

$$z = (x - m) / sd \tag{1}$$

where x is the data point, m is the mean, and sd is the standard deviation. The mean will be zero when with_mean = False, and the standard deviation will be one when with_std = False.

2.5 ANN model development

The next step involves building the model using an ANN inspired by the functioning the human brain. An ANN is a computational network based on biological NNs, where artificial neurons are linked through several layers. In general, the ANN comprises input, hidden, and output layers. It uses inputs to compute a weighted total. Utilizing this weighted total along with an activation function, the output is generated. The top eight features selected by the Extra Tree Classifier form the inputs of the NN model.

For the input, hidden, and output layers, rectified linear unit (ReLU) activation function and sigmoid function were applied. The binary cross-entropy loss function was utilized to quantify discrepancies between the computed and expected results, highlighting the model's performance. Additionally, the He uniform initializer and Adam optimizer were employed. As the NN learns over time, the weights are adjusted based on the outcomes of the loss function, enhancing model performance while minimizing loss. For binary classification problems, binary cross-entropy is appropriate. Activation functions generate the output of an NN, with ReLU efficient activation function is particularly efficient for various types of NNs.

To reduce loss and improve accuracy, optimizer functions are integrated into the NN model. The suggested model utilizes the Adam optimizer, which is a variant of stochastic gradient descent that combines elements from both the RMS and Adagrad algorithms. Additionally, the He uniform initializer is applied, enhancing the model's performance. In an NN, each input parameter is assigned a weight to calculate the weighted sum, and these weights are adjusted gradually based on the learning process's performance. It is crucial to use a good initializer when assigning weights, as initializing weights with small or large values can lead to a decline in model performance.

2.6 Patient clustering

Identifying heart disease and providing appropriate care are crucial aspects of the proposed model. In its second phase, patients are classified into three groups based on major parametric values: extremely severe, severe, and moderate. The parameters used for clustering include thalach, exang, oldpeak, slope, ca, and thal.

2.6.1 Consideration of thalach values

In the dataset, 'thalach' represents the heartbeat during exercise. Significant thalach values are determined based on the following formula:

$$\text{dthalach}=\text{thalach}-(\text{thalach}-\text{age})*\Omega \tag{2}$$

Here, Ω is either 0.64 or 0.76, representing the respective percentages of one's age.

$$\{\text{Low, Medium, High}\}\rightarrow\text{dthalach}\{(0-\lambda),(>\lambda-\text{б}),(>\text{б})\}$$

The deviation in thalach values determines the level of concern, categorized as low, medium, or high based on threshold values λ and б.

2.6.2 Consideration of exang values

'Exang' identifies whether a person experiences angina during exercise. In the dataset, 1 represents 'yes' and 0 represents 'no'. Exang value 1 is considered high, and 0 is considered low.

$$\{\text{Low, High}\}\rightarrow\{0,1\}$$

2.6.3 Consideration of oldpeak values

'Oldpeak' indicates ST depression in the ECG curve during exercise. Values < 2 indicate low risk, values between 2 and 2.55 indicate medium risk, and values > 2.55 indicates high risk (Adeli and Neshat, 2010).

$$\{\text{Low, Medium, High}\}\rightarrow\{<2,\ (2–2.55),>2.55\}$$

2.6.4 Consideration of slope values

'Slope' represents the ST segment during exercise. In the dataset, 0 represents down sloping, 1 represents flat, and 2 represents upsloping. Flat ST segments indicate normal conditions, while deviations indicate heart issues (ECG and ECHO Learning, 2022). In this model, 1 is considered low, 2 is medium, and 0 is high.

$$\{\text{Low, Medium, High}\}\rightarrow\{1,2,0\}$$

2.6.5 Consideration of ca values

'Ca' represents the number of vessels colored to identify blockage. In this model, 3 is considered high, 2 as medium, and 1 as low.

$$\{\text{Low, Medium, High}\}\rightarrow\{1,2,3\}$$

2.6.6 Consideration of thal values

In the dataset, 'thal' represents certain blood disorders. The values are coded as follows: 1 indicates a fixed defect, 2 indicates normal blood flow, and 3 indicates reversible effects. These values are clustered into categories labeled high, medium, and low respectively.

$$\{\text{Low, Medium, High}\}\rightarrow\{2,3,1\}$$

2.6.7 Apply *k*-means clustering

Using *k*-means clustering, classified cardiac patients are categorized into extremely severe, severe, and moderate categories. *k* -means clustering is an unsupervised learning mechanism that identifies clusters in a dataset based on a predetermined number of clusters. In *k* -means clustering, *k* data points are initially selected, followed by the random selection of *k* cluster centers. The Euclidean distance between each point and each cluster center is then calculated. Clusters are formed based on the distances between data points and cluster centers. *k*-means clustering is preferred due to its simplicity, effective performance with large datasets, and guaranteed convergence. The clusters are assigned priority based on the severity level of the patients.

2.7 Routing of patient information

In this phase, patient data for heart disease prediction is distributed for remote monitoring and support. The proposed routing algorithm considers hop counts, waiting list length, and energy levels. Each node is assigned a level according to its competency. Level value of the nodes is calculated using the following formula:

$$\text{NL}=1/\text{h}+1/\text{wl}+\text{el} \tag{3}$$

where NL indicates the node level, h indicates the number of hops, wl indicates the waiting list size (number of jobs in the list), and el indicates the energy level.

The higher the NL value, the more competent the node is for selection in routing. Each patient's clusters are assigned a priority according to the level of severity, as follows:

P=3 for extremely severe cluster
P=2 for severe cluster
P=1 for moderate cluster

When patient data from different clusters need to be routed simultaneously, nodes are selected based on their node level (NL) values. Higher priority clusters are assigned to nodes with higher NL values, which correspond to lower hop counts, smaller waiting list sizes, and higher energy levels. Therefore, high-energy nodes are prioritized for higher priority patients in this routing algorithm.

Proposed algorithm for node selection: Consider PInfo is the set of patients from different clusters, and NInfo is the set of nodes for routing, as follows:

```
PInfo={C1,C2,C3…Cn}
NInfo={N1,N2,N3,…Nn}
If Priority(Ci)=Max(Priority(C)): Then Ci→Max(NL(Ni))
```

Energy is a crucial parameter in real-time data sharing over an ad hoc network. Therefore, energy has been considered an influencing factor in node selection in this proposed model.

3. RESULTS AND DISCUSSION

3.1 Segment 1: ANN model

We utilized an ANN model for disease prediction. The model employs the ReLU activation function, He initializer, Adam optimizer, and L2 Regularizer. The achieved accuracy of the proposed predictive model is 90.16%, indicating its strong performance. The confusion matrix is shown in Figure 4. The proposed NN was developed with 100 epochs. The model's performance across different epochs is shown in Figure 5.

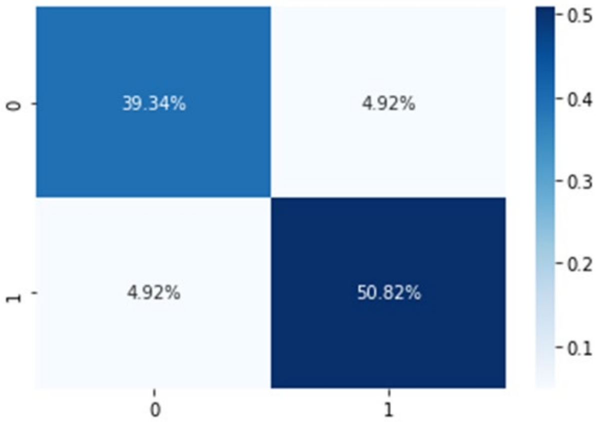


Figure 4. Confusion matrix (Applied ANN with He Initializer, Adam Optimizer, L2 Regularizer, and ReLU Activation Function)

The accuracy of the proposed ANN model, utilizing the He initializer, Adam optimizer, and L2 regularizer, and trained with a new feature set selected through the Extra Tree Classifier, reached 90.16%. In addition to accuracy, other metrics such as precision, sensitivity, specificity, and Cohen's kappa were considered for evaluating the performance of the proposed optimizers. These metrics were used for model validation. Accuracy provides a comprehensive assessment of a system's overall performance. Sensitivity and specificity focus on the system's ability to accurately identify positive and negative examples, respectively. Precision emphasizes the accuracy of positive predictions, while the F1 score aims for a balanced combination of precision and sensitivity. Cohen's Kappa evaluates the agreement level between model predictions and actual observed data, accounting for agreement that could occur by chance. The scores achieved for these metrics are presented in Table 1 and Figure 6, validating the model's performance.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{4}$$

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{6}$$

$$Specificity = \frac{TN}{FP+TN} \tag{7}$$

$$Cohen's\ Kappa = \frac{P_o-P_e}{1-P_e} \tag{8}$$

where TP indicates true positive, TN indicates true negative, FP indicates false positive, FN indicates false negative, Po indicates observed agreement, and Pe indicates expected agreement.

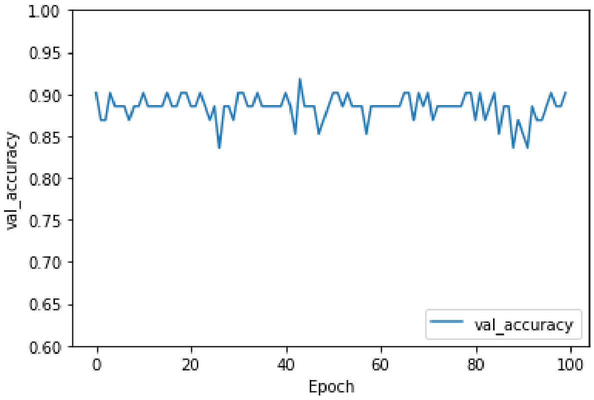


Figure 5. Model performance (ANN performance analysis in 100 epochs)

The classification model demonstrates a high level of performance, as indicated by the provided metrics. With an accuracy rate of 90.16%, it efficiently categorizes the majority of instances. Moreover, the observed high precision rate of 88.85% indicates accurate identification of positive cases while minimizing false-positive results. The sensitivity of 88.85% showcases the model's ability to identify true-positive cases effectively. Additionally, the model exhibits notable specificity (91.18%), indicating minimal false positives in true negative situations. The Cohen's Kappa score, measuring inter-rater agreement, shows a substantial level of agreement (82.44%) between the model's predictions and actual outcomes, surpassing what would be expected by random chance. Overall, these metrics provide strong evidence of the model's effectiveness in classification, making it a reliable tool for effectively differentiating between positive and negative cases. This attribute is particularly crucial in healthcare applications for validating machine learning-based model performance.

Table 1. Model validity measure through performance metrics

Accuracy	Precision	Sensitivity	Specificity	Cohen's Kappa score
90.16%	88.85%	88.85%	91.18%	82.44%

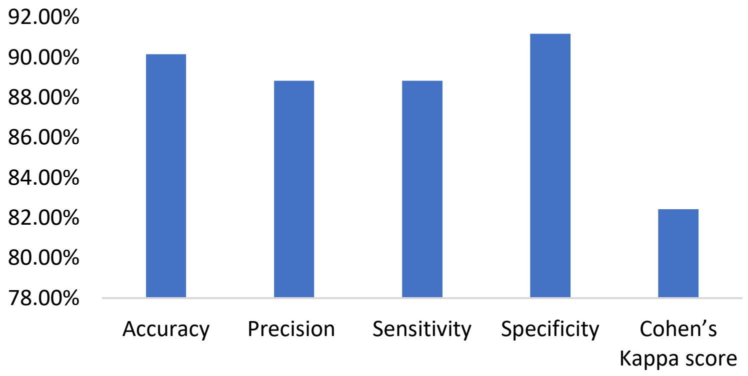


Figure 6. Performance measure of the proposed classifier

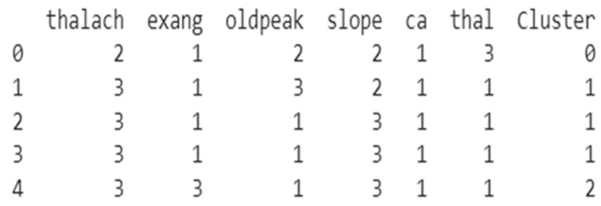
3.2 Segment 2: Formulation of patient clusters

After classifying cardiac patients, clusters have been established categorizing them as extremely severe, severe, and moderate. Since not all patients identified with cardiac disease are at the same level of risk, the required attention and care vary accordingly.

Identification of cardiac disease involves various features. Analyzing these feature values at a micro level helps observe their distribution and identify the risk level of each cardiac patient. The clustering methodology employed is *k*-means clustering. Figures 7 and 8 display the distribution of formulated clusters, where 0 represents moderate, 1 represents severe, and 2 represents extremely severe.

Iteratively examining a range of cluster numbers and evaluating their predictive power allows for an assessment of the clustering outcomes. Quantifying prediction strength helps determine how effectively the clusters segregate data points, serving as a measure of the model's validation.

The ideal number of clusters can be regarded as the one at which the prediction strength reaches a satisfactory level, indicating clear and significant cluster characteristics. This methodology aids in determining the optimal number of clusters that accurately capture the inherent structure of the data, making it a powerful tool for clustering validation and model selection. In this proposed model, the required number of clusters is 3, and the model's performance with 3 clusters is quite satisfactory (>70%, as shown in the graph in Figure 9.



```
   thalach  exang  oldpeak  slope  ca  thal  Cluster
0        2      1        2      2   1     3        0
1        3      1        3      2   1     1        1
2        3      1        1      3   1     1        1
3        3      1        1      3   1     1        1
4        3      3        1      3   1     1        2
```

Figure 7. Formulated patient clusters [0: moderate, 1: severe, 2: extremely severe]

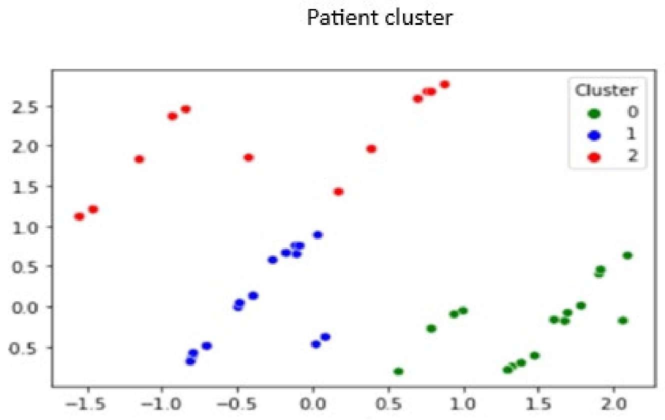


Figure 8. Cluster distribution

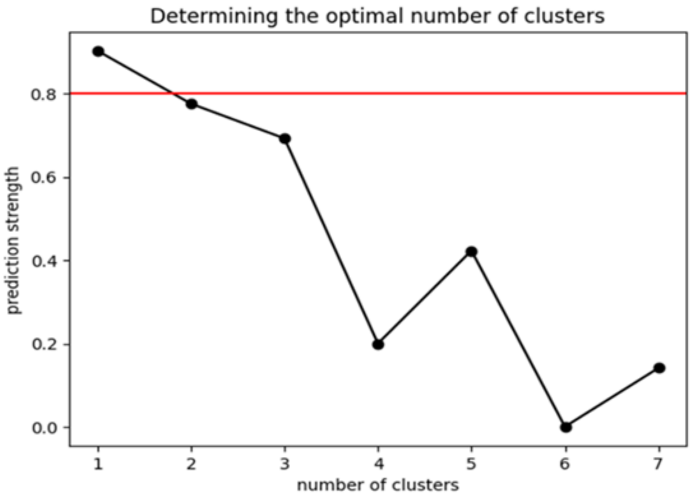


Figure 9. Validation of the number of clusters selected for the model

3.3 Segment 3: Routing of patient data

After identifying and clustering patients, the next step in the proposed model is routing patient data to servers. Sharing cardiac patients' data enables remote healthcare services, providing continuous support to prevent severe health issues. This always-on health service can benefit many individuals. The proposed algorithm is based on three metrics: hop count, waiting list length, and energy level. The NL metric calculates the competency of nodes for routing information. Hop count and waiting list length are inversely proportional to NL, while energy level is directly proportional.

Each patient's data is assigned priority based on the NL value. Higher priority data is routed to nodes with higher NL values. Sample parameter values of the proposed ad-hoc routing mechanism for routing of patient data through the selection of competent nodes are shown in Table 2 and Figure 10.

Validation of the routing mechanism in this proposed model can be achieved by visualizing the calculated NL value and the generated graph displaying a computed path, as shown in Figures 11 and 12, based on a sample execution represented in Figure 13.

This proposed healthcare system not only classifies patients with cardiac disease but also clusters them and routes cardiac disease-related parameters of patients to a remote server for monitoring and consultations. The remote healthcare support facilitated by the ad-hoc routing algorithm establishes a continuous feedback loop between patients and healthcare professionals. This real-time information exchange enables prompt adjustments to treatment plans, medication dosages, and lifestyle recommendations, contributing to proactive disease management and reducing the risk of emergencies. This is particularly crucial in remote or economically challenged areas where access to healthcare might be limited. The holistic nature of this method not only improves the accuracy of disease prediction but also fosters a more personalized and responsive healthcare system, ultimately enhancing patient outcomes and the quality of care for cardiac patients, regardless of geographical constraints.

Table 3 presents a detailed state-of-the-art analysis, addressing the methods, major gaps, and results of existing models compared to our proposed model.

Table 2. Participating selection at different layers for routing of patient data

Layer-1		**Layer-2**		**Layer-3**	
Node level	**Assigned patient data (with priority)**	**Node level**	**Assigned patient data (with priority)**	**Node level**	**Assigned patient data (with priority)**
3.4	1	4.5	1	12.5	3
5	1	8.7	3	11.43	3
6.66	1	14	3	5.77	1
2.79	1	5.6	1	12	3
7	1	7	1	7	1
9.78	2	6	1	16.5	3
2.5	1	11	3	12.78	3
6	1	16	3	8.8	2
7.09	1	6.7	1	9.5	2
10.34	3	5.9	1	8.7	2
12.34	3	4	1	9	2
13.67	3	5	1	6.7	1
9	2	10.89	3	11.08	3
3.45	1	13	3	14	3
5.6	1	15.67	3	3.4	1

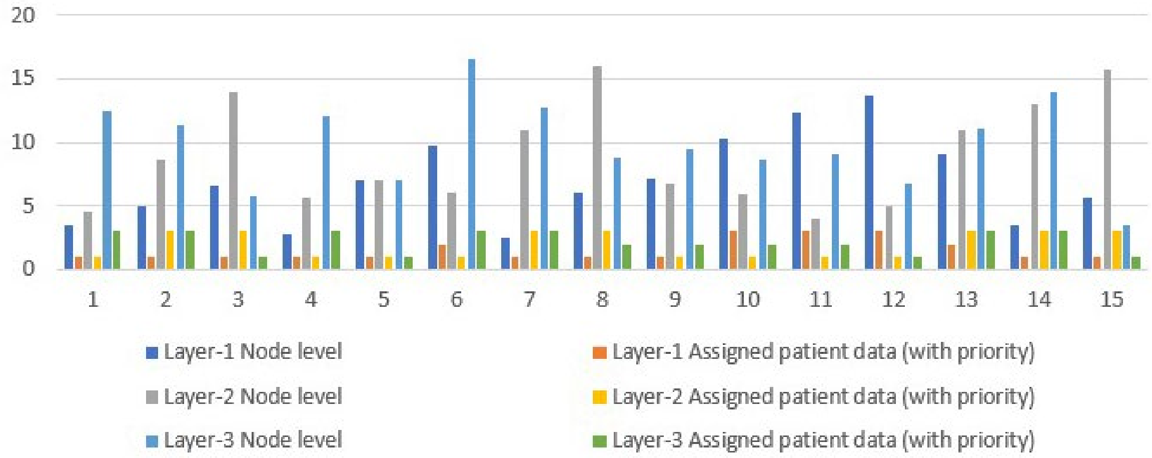


Figure 10. Node competency vs assigned node

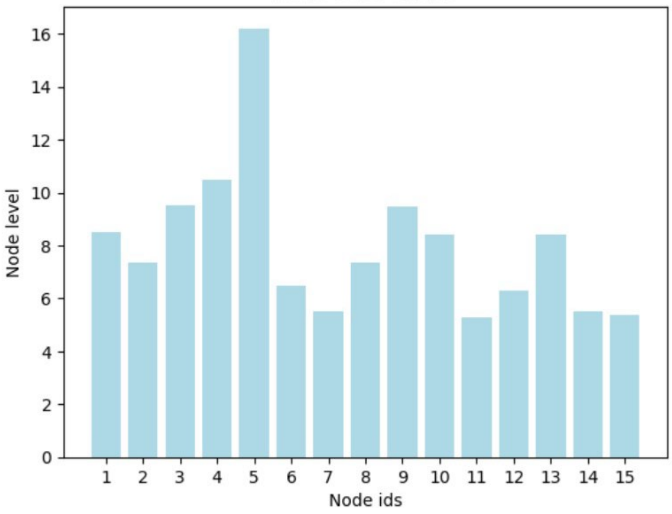


Figure 11. Visualization of the calculated node competency level based on the sample execution presented in Figure13

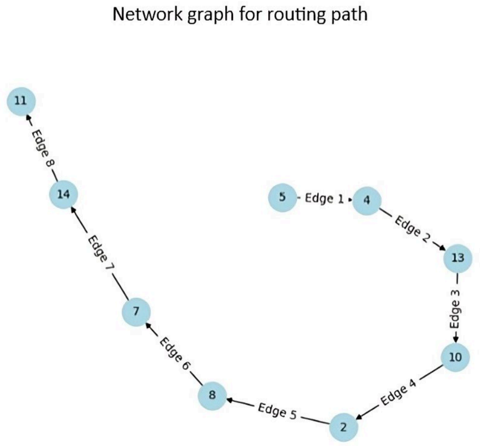
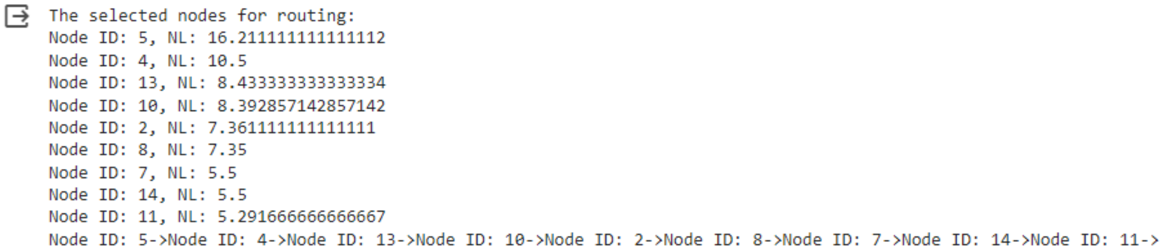


Figure 12. Network graph of the routing path selected based on the node competency level presented in Figure 11



```
The selected nodes for routing:
Node ID: 5, NL: 16.211111111111112
Node ID: 4, NL: 10.5
Node ID: 13, NL: 8.433333333333334
Node ID: 10, NL: 8.392857142857142
Node ID: 2, NL: 7.361111111111111
Node ID: 8, NL: 7.35
Node ID: 7, NL: 5.5
Node ID: 14, NL: 5.5
Node ID: 11, NL: 5.291666666666667
Node ID: 5->Node ID: 4->Node ID: 13->Node ID: 10->Node ID: 2->Node ID: 8->Node ID: 7->Node ID: 14->Node ID: 11->
```

Figure 13. Sample execution of the program

Table 3. Comparative analysis

Proposed model	Performance observations
Zheng et al. (2017)	Applied-Naïve Bayes and decision tree Naïve Bayes has given good performance in small dataset. Decision tree has given good performance in large dataset. This proposed model can only predict disease. Patients clustering according to the severity has not been considered. No further module has been proposed for transferring patients' data to remote server for monitoring.
Rajesh et al. (2018)	Applied following methods with achieved accuracy: Logistic regression: 84% KNN(K=9): 76% Artificial neural network: 74% SVM (kernel = RBF, C = 100, g =0.0001) :86% SVM (kernel = linear): 75% Naive Bayes: 83% Decision tree: 74% Random forest 83% This proposed model can only predict disease. Patients clustering according to the severity has not been considered. No further module has been proposed for transferring patients' data to remote server for monitoring.
Santhana and Geetha (2019)	Applied: Decision Tree: 79.12% accuracy achieved This proposed model can only predict disease. Patients clustering according to the severity has not been considered. No further module has been proposed for transferring patients' data to remote server for monitoring.
Patel et al. (2019)	Naïve Bayes: 75.8% accuracy achieved Generalized linear model: 85.1% accuracy achieved Logistic regression: 82.9% accuracy achieved Decision tree: 85% accuracy achieved Gradient boosted tree: 78.3% accuracy achieved This proposed model can only predict disease. Patients clustering according to the severity has not been considered. No further module has been proposed for transferring patients' data to remote server for monitoring.
Sharma et al. (2020)	Accuracy achieved in different methods: Sharma et al., 2020 Decision tree: 84% Naive Bayes: 83% This proposed model can only predict disease. Patients clustering according to the severity has not considered. No further module has been proposed for transferring patients' data to remote server for monitoring.
Rajdhan et al. (2020)	SVM: 81.75% accuracy achieved Decision tree: 80.43% accuracy achieved Naïve Bayes: 80.43% accuracy achieved Logistic regression: 82.89% accuracy achieved This proposed model can only predict disease. Patients clustering according to the severity has not been considered. No further module has been proposed for transferring patient's data to remote server for monitoring.

Table 3. (continued)

Proposed model	Performance observations
Shao et al. (2022)	The proposed model is deep clustering algorithm with auto-encoder embedding. Patients were being clustered considering electronic medical records. No module for patients' data transferring had proposed for further monitoring.
Majumder et al. (2021a)	This proposed work only focused on a BAN routing by mapping patients' cluster to a node cluster for transferring patient's data. No disease prediction model was proposed.
Stoitsas et al. (2022)	This proposed work used multiple machine learning and statistical models for clustering patients. The classes considered are: Large majority classes and clinically less sensible classes. Only clustering has done. No module has been introduced for transferring patient's data for further monitoring.
Majumder et al. (2022)	A Geo replicated Intelligent BAN routing mechanism has been proposed for transferring patients' data to remote server for monitoring. No model for disease prediction was introduced in this work.
Our proposed model	Our proposed model focused on disease prediction, patients clustering according to severity and transferring patients' data over the network for further monitoring. For disease prediction artificial neural network has been used. Evaluation Matrix used are as follows: Accuracy: 90.16% Precision: 88.89% Sensitivity: 88.89% Specificity: 91.18% Cohen's Kappa: 82.44% For forming the patients cluster *k*-means clustering algorithm has been used. Three clusters have been formed- Extremely severe, server, and moderate. This clustering helps to provide exact care demand for each patient according to severity. In the last module an energy efficient routing algorithm has been proposed for transferring patients' data to remote server so that experts can remotely access and monitor patients' condition and provide suggestions 24X7. These elements combine to offer a comprehensive approach that can greatly enhance healthcare by providing personalized care and efficient data management in the era of remote health monitoring.

4. CONCLUSION

Cardiac disease poses a significant threat to human health, necessitating heightened attention. Machine learning offers a valuable tool for the early prediction of cardiac disease. In this study, a model for cardiac disease prediction, patient clustering, and routing of patient data for monitoring purposes has been proposed. The cardiac disease prediction model was formulated using an ANN and trained on the UCI dataset. Feature selection was conducted using the extra tree classifier for its computational efficiency, resulting in a model with 90.16% accuracy. Subsequently, cardiac patients were clustered into three groups—extremely severe, severe, and moderate—using *k*-means clustering, aiding in identifying disease severity based on various influencing parameters, enabling tailored care for patients according to their risk level. In the next phase, classified patient data related to the disease were routed to a server for remote monitoring. An ad hoc routing mechanism was proposed, considering factors such as hop count, energy level, and waiting list length of each node. This proposed model not only predicts cardiac disease at an early stage but also categorizes patients by disease severity and proposes a routing mechanism for vital patient information to be sent to a remote server, facilitating remote healthcare services anytime and anywhere. This mitigates issues related to limited healthcare access due to economic or geographic constraints. However, the model has limitations, such as being trained on previously collected data from the UCI data repository and lacking real-time data collection from patients using wearable sensors. Looking ahead, future prospects include ongoing refinement of models, advancements in interpretability, personalized medicine integration with wearable devices, fostering interdisciplinary collaboration, and expanding AI-driven healthcare applications to address additional medical challenges.

REFERENCES

Adeli, A., and Neshat, M. (2010). A fuzzy expert system for heart disease diagnosis. In *Proceedings of the International MultiConference of Engineers and Computer Scientists Vol. 1* (Ao. S. I., Castillo, O., Douglas, C., Feng, D. D., and Lee, J.-A., Eds.), pp. 134–139. Hong Kong.

ECG and ECHO Learning. (2022). The ST segment: Physiology, normal appearance, ST depression and ST elevation. *ECG and ECHO Learning*. [Online URL: https://ecgwaves.com/st-segment-normal-abnormal-depression-elevation-causes/] accessed on February 10, 2023.

Gupta, S., and Banerjee, A. (2015). Proposed intelligent system to identify the level of risk of cardiovascular diseases under the framework of bioinformatics. In *Proceedings of the First International Conference on Advancements of Medical Electronics* (Gupta, S., Bag, S., Ganguly, K., Sarkar, I., and Biswas, P., Eds.), pp. 3–12. West Bengal, India.

Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., and Sun, R. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. *Mobile Information Systems*, 2018(1), 3860146.

Janosi, A., Steinbrunn, W., Pfisterer, M., and Detrano, R. (1988). Heart disease. *UCI Machine Learning Repository*. [Online URL: https://archive.ics.uci.edu/dataset/45/heart+disease] accessed on February 10, 2023.

Jindal, H., Agrawal, S., Khera, R., Jain, R., and Nagrath, P. (2021). Heart disease prediction using machine learning algorithms. *IOP Conference Series: Materials Science and Engineering,* 1022, 012072.

Lau, K. Y-Y., Ng, K.-S., Kwok, K.-W., Tsia, K. K.-M., Sin, C.-F., Lam, C.-W., and Vardhanabhuti, V. (2022). An unsupervised machine learning clustering and prediction of differential clinical phenotypes of COVID-19 patients based on blood tests—A Hong Kong population study. *Frontiers in Medicine*, 8, 764934.

Majumder, A. B., Gupta, S., and Dharmpal, S. (2021a). An intelligent BAN routing mechanism for transferring remote patient monitoring data using *k* means clustering. *SSRN Electronic Journal*, Online.

Majumder, A. B., Gupta, S., Singh, D., and Majumder, S. (2021b). An intelligent system for prediction of COVID-19 case using machine learning framework-logistic regression. *Journal of Physics: Conference Series*, 1797, 012011.

Majumder, A. B., Majumder, S., Gupta, S., and Singh, D. (2022). An intelligent, geo-replication, energy-efficient BAN routing algorithm under framework of machine learning and cloud computing. In *Proceedings of the International Conference on Data Management, Analytics and Innovation*, pp. 43–53. Online.

Mohan, S., Thirumalai, C., and Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access,* 7, 81542–81554.

Muhammad, Y., Muhammad, T., Hayat, M., and Chong, K. T. (2020). Early and accurate detection and diagnosis of heart disease using intelligent computational model. *Scientific Reports*, 10, 19747.

Murthy, S., and Garcia-Luna-Aceves, J. J. (1996). An efficient routing protocol for wireless networks. *Mobile Networks and Applications*, 1, 183–197.

Pasha, S. N., Ramesh, D., Mohmmad, S., Harshavardhan, A., and Shabana. (2020, October 9–10). Cardiovascular disease prediction using deep learning techniques [Paper presentation]. *International Conference on Recent Advancements in Engineering and Management*, Warangal, India.

Patel, A. C., Mishra, M., Shameem, A., Chaurasiya, S., and, Saxena, A. (2019). Prediction of heart disease using machine learning. *International Journal of Scientific Development and Research*, 4(4), 354–357.

Rajdhan, A., Agarwal, A., Sai, M., Ravi, D., and Ghuli, P. (2020). Heart disease prediction using machine learning. *International Journal of Engineering Research and Technology*, 9(4), 659–662.

Rajesh, N., Maneesha, T., Hafeez, S., and Krishna, H. (2018). Prediction of heart disease using machine learning algorithms. *International Journal of Engineering and Technology*, 7(2.32), 363–366.

Rakhee, and Srinivas, M. B. (2016). Cluster based energy efficient routing protocol using ANT colony optimization and breadth first search. *Procedia Computer Science*, 89, 124–133.

Ramprakash, P., Sarumathi, R., Mowriya, R., and Nithyavishnupriya, S. (2020). Heart disease prediction using deep neural network. In *Proceedings of the International Conference on Inventive Computation Technologies*, pp. 666–670. Coimbatore, India.

Sajja, T. K., and Kalluri, H. K. (2020). A deep learning method for prediction of cardiovascular disease using convolutional neural network. *Revue d'Intelligence Artificielle*, 34(5), 601–606.

Santhana, K. J., and Geetha, S. (2019). Prediction of heart disease using machine learning algorithms. In *Proceedings of the 1st International Conference on Innovations in Information and Communication Technology*, pp. 53–57. Chennai, India.

Shah, D., Patel, S., and Bharti, S. K. (2020). Heart Disease prediction using machine learning techniques. *SN Computer Science*, 1, 345.

Shao, W., Luo, X., Zhang, Z., Han, Z., Chandrasekaran, V., Turzhitsky, V., Bali, V., Roberts, A. R., Metzger, M., Baker, J., La Rosa, C., Weaver, J., Dexter, P., and Huang, K. (2022). Application of unsupervised deep learning algorithms for identification of specific clusters of chronic cough patients from EMR data. *BMC Bioinformatics*, 23(Suppl 3), 140.

Sharma, V., Yadav, S., and Gupta, M. (2020). Heart disease prediction using machine learning techniques. In *Proceedings of the 2nd International Conference on Advances in Computing, Communication Control and Networking* (Sharma, V., Srivastava, R., and Singh, M., Eds.), pp. 177–181. Greater Noida, India.

Stoitsas, K., Bahulikar, S., de Munter, L., de Jongh, M. A. C., Jansen, M. A. C., Jung, M. M., van Wingerden, M., and van Deun, K. (2022). Clustering of trauma patients based on longitudinal data and the application of machine learning to predict recovery. *Scientific Reports*, 12(1), 16990.

Zheng, T., Xie, W., Xu, L., He, X., Zhang, Y., You, M., Yang, G. and Chen, Y. (2017). A machine learning-based framework to identify type 2 diabetes through electronic health records. *International Journal of Medical Informatics*, 97, 120–127.