

Original Research

Received 27 May 2026

Revised 25 July 2026

Accepted 02 August 2026

Natural Resources for Human Health 2026, Vol. 6 (14s): 476–486

KEYWORDS:

RST, SVM, MREF, Cardiac Arrest, ANN

Prediction of Cardiac Arrest using Multiple Learning Algorithm

¹Niranjan Badu, ²Kakita Murali Gopal, ³Sujogya Mishra, ⁴Stitapragyan Lenka, ⁵Dr.Bibhu Kalyan Mishra

¹Computer Science and Application, GIET University, Gunupur, India
badu.niranjan@gmail.com

²School of Engineering and Technology, GIET University Gunupur, India
kmgopal@giет.edu

³Department of Mathematics, Odisha University of Technology and Research
Bhubaneswar, India

sujogya123@gmail.com

⁴Department of Computer Science and Application, Utkal University Bhubaneswar, India

stitapragyan@utkaluniversity.ac.in

⁵Associate Professor, FET, Sri Sri University, Cuttack
bibhu.m@sisriuniversity.edu.in

ABSTRACT: Cardiac arrest (CA) is a leading cause of death worldwide and often occurs without warning. Early detection and intervention are critical to saving lives. Because there are no clear warning indicators, traditional methods for predicting CA have not been effective. When machine learning (ML) techniques are integrated, survival rates may increase since these techniques provide additional information that can lead to faster detection and treatment.

Objective: This research proposes rapid, early detection methods for Cardiovascular Diseases (CVD). The reason behind CA is an irregular lifestyle, eating habits, smoking and several other factors. This study addresses the handling and prediction of complex cardiac-related diseases using machine learning.

Method and Materials: This study plans to utilise the following machine learning and soft computing tools: Rough Set Theory (RST), Artificial Neural Network (ANN), Random Forest (RF), Support Vector Machine (SVM), and Modified Feature Elimination (MREF).

I. INTRODUCTION

There are several cardiac-related diseases like Pulmonary Tuberculosis, Coronary Artery diseases, Heart Valve diseases, and Cardiac arrest (CA). Cardiac arrest is the most significant cardiac issues, which increase the death rate drastically among all cardiac-related diseases. Usually, the prediction of cardiac arrest (CA) depends to some extent on clinical analysis and normalising the risk scores. Although conventional methods help manage complex cardiac issues, there are situations where they fail to yield fruitful results. Especially when handling and detecting the possibilities of CA in a patient. There are several instances in which a healthy person died of cardiac arrest without any proper symptoms. Other diseases, such as loose motion, malaria, and typhoid, share some general symptoms. However, cardiac-related diseases don't have any common symptoms. This research aimed to identify symptoms that facilitate early detection of cardiac-related diseases. Several researchers have utilised machine learning and soft computing to predict cardiac arrest; in this context, Hota et al. [1] employed machine learning algorithms to predict cardiac arrest. A person with a cardiac disorder without any significant symptoms is, in this context, referred to as Das et al. [2] had designed a model using RST for predicting pulmonary tuberculosis. Several research gaps exist

in predicting cardiac arrest. In addition to cardiac arrest prediction, several researchers have used machine learning tools to predict Alzheimer's disease. In this context, Biswas et al. [3] have predicted Alzheimer's disease using deep learning, machine learning, and statistical methods. There are numerous significant anomalies in contemporary research on predicting cardiac arrest using machine learning tools. The possible challenges in the prediction of cardiac-related diseases using machine learning are as follows:

- a. Several machine learning models used a single data source for training and testing; the models need to validate their claims by using a diversified dataset taken from various hospitals
- b. It's always challenging to gather cardiac-related medical data sets from recognised hospitals, i.e. these standard hospital data are not available universally.
- c. Difficult for the doctors to consider the predicted results until the medical council validates them
- d. Difficulty in understanding the fundamental objective of the contemporary machine learning model used for general-purpose.
- e. Sometimes, predicting a high risk of cardiac arrest for a patient leads to a false negative situation.

The objective of this research is to facilitate early detection of cardiac arrest by identifying the optimal machine-learning model for cardiac arrest, thereby aiding doctors in treating this condition. If a heart attack does occur, giving the medical team advance notice allows them to respond quickly and handle the situation more effectively. The application of an integrated machine learning model reduced mortality rates. The general model for cardiac arrest prediction in contemporary research gives an accuracy level above 92%. However, in reality, these results are generated by using local data, i.e., considering a single data source. Research in data analysis is always challenging, as it requires a large sample size and accurate analysis. Generally, the medical data are vague and imprecise. This study has employed RST to address the data's vagueness, as well as RST-based machine learning tools for accuracy prediction, which are helpful for doctors in the early detection of cardiac arrest.

II. LITERATURE REVIEW

Several researchers have employed various techniques to predict cardiac arrest. In this context, Taasila et al. [4] utilised deep learning to predict cardiac arrest accuracy. Various researchers have employed alternative methods to predict symptoms in this context. Banerjee and Das [5] used machine learning to predict lung cancer. Some researchers have utilised soft computing tools to predict the symptoms of various diseases. In this context, Das et al. [6] employed RST methods to predict pulmonary tuberculosis symptoms. There are several diseases with similar symptoms. To detect these diseases, it is essential to identify the significant symptoms that can help distinguish them. In this context, Nayak et al. [7] employed soft computing methods to predict malaria symptoms, thereby distinguishing it from other forms of fever. In a similar context, Hota et al.[8] had used generalised soft techniques to predict cardiac arrest. Several researchers have employed a combination of soft computing and machine learning techniques, as demonstrated by Nayak et al.[9] had used boosting, i.e., a combination of two or more machine learning techniques, for predicting cardiac arrest. Numerous studies have been conducted to predict cardiac arrest by comparing the predictive accuracy of various machine learning algorithms. Nayak et al.[10] Compare the predictive accuracy of several machine learning algorithms for cardiac arrest. Some researchers have utilised deep learning tools for predicting cardiac arrest. In this context, Kwon et al. [11] employed extended AI techniques to predict cardiac arrest. Among all fatal diseases, cardiac arrest is the most significant one, which can take a human life within a short period of time. The most developed nation, like the United States (US), is unable to contain the death rate due to cardiac arrest. Every year, the US suffers approximately 300,000 deaths due to cardiac arrest. To date, numerous studies have been conducted to achieve fruitful results in predicting cardiac arrest; some researchers have utilised machine learning or deep learning tools for this purpose. Mathematically analysing and developing a model that integrates machine learning tools with rigorous mathematical analysis is always a question of achieving significant accuracy in predicting cardiac arrest. In this context, Nayak et al.[12]had designed a model using RST, a Support Vector Machine (SVM), and a differential equation for predicting cardiac arrest. There are several research gaps in the contemporary studies. The research gaps are given below.

1. Generally, most research conducted on the data set is taken from either a single source or considers data from a single hospital.

2. Few researchers were using a universal dataset. A drawback in this method is that not all symptoms are present in the universal dataset
3. The machine learning tools cannot be applied to vague/imprecise datasets
4. Oversampling and undersampling situations were not discussed in the contemporary research
5. If the sample was taken from a single medical source, it leads to oversampling
6. The dataset was not taken from a random sample
7. Normalisation of a data dataset may lose some of its significant properties
8. Standardising training and testing ratios was unique in the contemporary studies

Some researchers have tried to overcome the above drawbacks. To overcome the above issues in this context, Das et.al.[13,14] developed a model for predicting thoracic diseases with significant predictive value. RST is a technique for handling vague and imprecise data. The RST concept was developed by Z. Pawlak [15] in the early 80s. Fundamental concept of RST derived from conventional set theory, with certain restrictions as discussed by Nayak et al.[16].

A. Background for data analysis

The entire study includes RST and Machine learning tools. The fundamental concepts of RST are Upper Approximation, Lower Approximation, Boundary Region, Reduct, and Core.

a. Skeleton of RST $\langle E, C, D \rangle$, E is the set of classes, C is the set of conditional attributes, and D is the set of decision attributes.

b. Structure of RST

Table I Fundamental structure of RST

E	A ₁	A ₂	A ₃	D
E ₁	1	1	2	N
E ₂	1	1	1	Y

E is the set of class $E = \{E_1, E_2\}$ set of attributes and $\langle A_1, A_2, A_3 \rangle = C$, i.e. set of conditional attributes, and D is the set of decision attributes.

c. Indiscernibility

Indiscernibility is defined using the attributes of interest if two objects cannot be differentiated from each other. Mathematically, indiscernibility denoted as $IND(R) = \{ (a,b) \in K^2 \text{ such that } x(a)=x(b) \} \forall x \in R \}$. Where K is the set of the universe, R is the target set. $x(a)$, $x(b)$ is the objects.

d. Upper Approximation

The object of interest is likely to belong to the target set. The upper approximation mathematically denoted as $K^*X = \bigcup \{ Y \in U/K \mid Y \cap X \neq \emptyset \}$ (1)

e. Lower Approximation

The object of interest definitely belongs to the target set, mathematically denoted as

$$K_*X = \bigcup \{ Y \in U/K \mid Y \subseteq X \} \quad (2)$$

f. Boundary Region

Boundary Region

= Upper Approximation- Lower Approximation (3)

B. Studies and analysis using local data

The dataset was collected from various hospitals and local sources related to cardiac anomalies. Initially, the data was collected from local hospitals, as shown in the following Table.

Table II Initial data set

Records	1	2	3	4	5	6
E ₁	10,000	15,000	20,000	15,000	60,000	1,20,000
E ₂	11,000	10,000	10,000	25,000	56,000	1,12,000
E ₃	10,000	15,000	25,000	25,000	75,000	1,50,000
E ₄	5,000	15,000	15,000	25,000	60,000	1,20,000
E ₅	5,000	15,000	15,000	25,000	60,000	1,20,000

The above Table contains the data set related to cardiac-related diseases E= <E₁, E₂, E₃, E₄, E₅>

E₁ – Sambalpur, E₂ – Cuttack, E₃ – Dhenkanal, E₄ – Anugul, E₅ – Nayagarh, 1- Pulmonary Tuberculosis, 2 – Severe Coronary Artery Diseases (CAD), 3- Cardiac arrest, 4 –Peripheral Artery

Diseases (PAD), 5- Rheumatic Heart Diseases (RHD), 6-Total instances. The data collected on cardiac diseases is approximately 6,22,000, collected from various sources. Logistic regression is applied to data collected from various sources for binary classification. The study consists of two categories, i.e. death rate and recovery rate. Applying logistic multiclass regression to the data in Table II (with each disease 1 to 6 defined above) and more than 600,000 records yields the following results.

The results in Table III were predicted using binary-class logistic regression.

C. Fundamentals of Logistic Regression

Logistic regression is generally used for binary classification. Logistic regression is a supervised learning algorithm. This study has used logistic regression for grouping the data according to dissimilarity present in the raw data; this concept was discussed by Hota et.al.[17] They used various machine learning algorithms and soft computing techniques to predict cardiac arrest. Some researchers have used boosting techniques alongside logistic regression to predict cardiac arrest.

An algorithm to regroup the data according to dissimilarity

1. Initialize using a dissimilar matrix according to the distance measures using the above data.
2. Application of K-means to get group similar data points using a dissimilarity matrix.
3. Train the cluster points using logistic regression, which satisfy the cluster-specific and the patterns label.
4. Calculate the dissimilarity of new data with respect to existing data. Include the new data points in the nearest cluster. Find the predicted class using logistic regression.
5. Find the dissimilarity space of the available data points using logistic regression.

Logistic regression is applied to the multi-class data from Table II to regroup the data for further classification based on the death rate and recovery rate given in Table III. The expected results were given in Table IV. Class is defined as 0 to less than 45 as low, represented as 0, 45 to less than 55 as moderate, represented as high and above 55 as 2.

Table III Expected Rate of recovery

Diseases	Death rate	Recovery rate
1	45%	55%
2	55%	45%
3	65%	35%
4	40%	60%
5	45%	55%
6	55%	65%

Table IV Various Category of Cardiac arrest

P	Low (Below 45%)	Moderate (>45% ,≤55%)	High (>55%)
1	0.2167	0.4513	0.3320
2	0.2219	0.2789	0.4992
3	0.1976	0.1499	0.6525
4	0.1586	0.0726	0.7688
5	0.1192	0.0329	0.8479
6	0.0861	0.144	0.8995

D. Classification of Cardiac related diseases

Further analysis of cardiac-related diseases is being conducted using various parameters, i.e., body weight, age, gender, and smoker/non-smoker status. The entire distribution is given in the following figure. The entire dataset is categorised by gender (male or female), smoking status (smoker or non-smoker), age, and body weight in lb. The data is analysed further to get a refined result for further analysis. This study has used multiclass logistic regression. The refined form of the raw data is given in Figure 1 and Table V

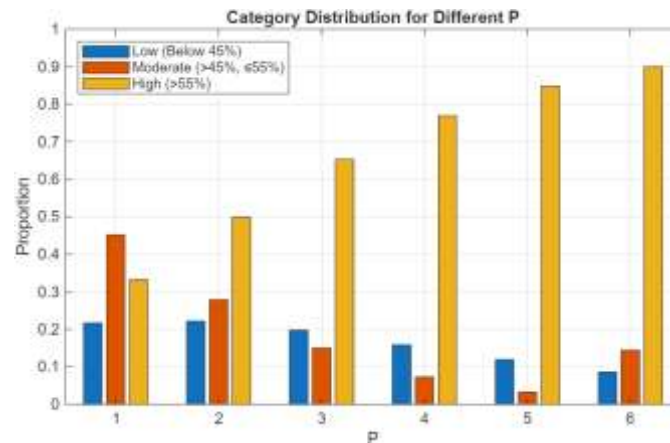


Fig 1 Representing categories of cardiac arrest

The dataset was trained on age, weight, and habit. The bar plot shows the probability P vs. proportion, with age, weight, and habit included. The smoking habits and weight lead to cardiac arrest, as given in the following table.

Table V Cardiac arrest risk

Age	Weights in LB	Smoker	Probability
56	125	0	0.048
57	126	1	0.1305
68	181	0	0.1305
68	181	1	0.8105
77	215	0	0.25
77	215	1	0.87

Clearly observed that person having weight over 181lb, and age of more than 55, with smoking habit are likely to suffer from cardiac arrest. In the subsequent section this research is analysed with respect to (Weight, Age). In the subsequent section., this study has used the RST concept to identify a set of symptoms for cardiac arrest.

E. Data Analysis using RST

The modification of the original data set based on age, weight, habit, gender, and death and recovery rates due to cardiac-related diseases. This study used RST on the refined results derived using data taken from Tables III,IV,V.,

Table VI Initial Information

E	1	2	3	4	5
E ₁	a	A	b	a	a
E ₂	a	B	b	a	a
E ₃	a	A	a	b	b
E ₄	b	B	a	a	b

<1,2,3,4,5> explained in the table II called as conditional attributes, and a,b are called as values of the conditional attributes i.e a- significant ,b- insignificant.

- *Algorithm to find Indiscernibility*

Step 1: The Set of objects is the universal set

Step 2: The Attributes set is A

Step 3 : $\forall(x,y)$ in U Let the set of attributes be T.

Analyse every value of T; if two values of T match, then the two values. If both values are the same, then both are indiscernible

Step 4: Use the equivalence classes to partition the universal set U

Indiscernibility denoted as IN, calculation of Indiscernibility,

$$IN(1)=\{ (E_1,E_2,E_3),E_4\} \quad (1)$$

$$IN(2)=\{ (E_1,E_3), (E_2,E_4)\} \quad (2)$$

$$IN(3)=\{ (E_1,E_2), (E_3,E_4)\} \quad (3)$$

$$IN(4)=\{ (E_1,E_2,E_4),E_3\} \quad (4)$$

$$IN(1,2)=\{ (E_1,E_3), E_2,E_4\} \quad (5)$$

$$IN(1,3)=\{ (E_1,E_2), E_3,E_4\} \quad (6)$$

$$IN(1,4)=\{ (E_1,E_2), E_3,E_4\} \quad (7)$$

$$IN(1,5)=\{ (E_1,E_2), E_3,E_4\} \quad (8)$$

- *Reduct Algorithm*

Step 1: Find all Indiscernibility

Step 2: Compare the Indiscernibility, if it does not form a group, then it produces a reduct

Step 3: Compute the Core = $\cap$ Reduct

Similarly, finding the Indiscernibility with respect to others' attributes and the results {1,2,3}, {2,3,4}, and {2,3,5} does not provide any groups, so {1,2,3}, {2,3,4}, and {2,3,5} are called Reducts.

According to Core = $\cap$ {1,2,3}, {2,3,4}, and {2,3,5}, reduct is now {2,3}, attribute 2 represent CAD and 3 represent cardiac arrest. This research exclusively considers cardiac arrest for further analysis.

III. DATA ANALYSIS USING THE UNIVERSAL DATASET

The universal data set was collected from Kaggle [18]. The dataset is preprocessed as follows.

Study of the data set, which includes analysing the data set, choosing the significant attributes related to cardiac-related diseases, and several missing data points being preprocessed. Using several machine learning algorithms to provide significant prediction accuracy. The exact result is the analysis of evidence with respect to the output.

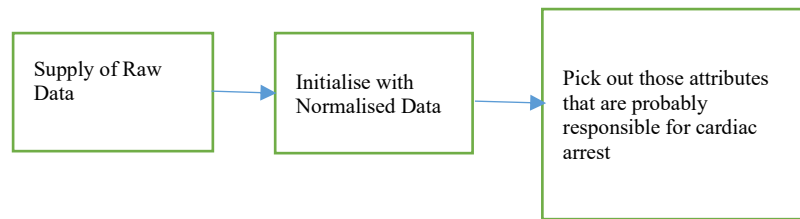


Fig. 2 Representation and analysis of the raw data set

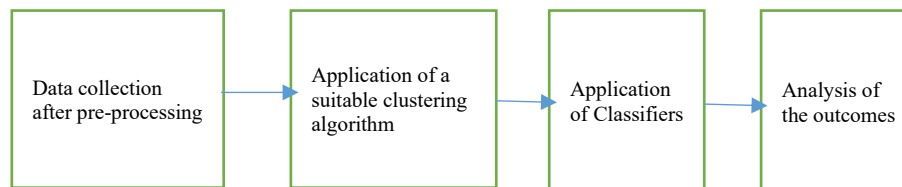


Fig. 3 Result analysis of pre-processing data after the application of clustering and classifier algorithms.

This study considers 14 attributes for the predictions. The attributes are given in the following Table IV. CP- Chest Pain, RBP- Resting Blood Sugar, SCH- Serum Cholesterol, FBS- Fasting Blood Sugar, RER- Resting Electrocardiographic Result, MHRA- Maximum Heart Rate Achieved, EIA- Exercise Induced Angina, STD- ST depression. The above conditional attributes are considered for this research as suggested by Dubey et.al.[19]. The K-means algorithm was used in this study to identify similarities between high- and low-class samples in the current data set. The K-means algorithm clusters based on the arithmetic average, as discussed by Soni et al. [20] and Masethe and Masethe [21].

Table VII Attribute of cardiac data set

Sl no	Attributes
1	Age
2	CP
3	RBP
4	SCH
5	FBS
6	RER
7	MHRA
8	EIA
9	STD
10	Slope of ST
11	Thalassemia
12	Sleeping habit
13	Angiographic
14	Others

A. Application of K-Mean Clustering

Step 1: Initialise the data points according to y_i , where $i = 1, \dots, n$, and divide them into k clusters.

Step 2: Apply Euclidean distance concepts to find the shortest distance of $(n-2)$ clusters from the randomly chosen two clusters

Step 3: Find the arithmetic mean of the newly found data point with the cluster it belongs to, and the arithmetic mean becomes the new cluster. Compare the new cluster with the rest of the data points.

Step 4: Continue steps 2 and 3 until all the data points converge to two randomly chosen clusters.

The k-means algorithm is applied to the dataset to determine the appropriate number of data groups. The patient's attributes are assigned numeric values, and their central point is calculated using the Euclidean distance algorithm. The average value is used for the data sample, which helps predict the patient's physical condition. If the available result for cardiac arrest matches the mean results of the patient's sample, then that patient is more prone to cardiac arrest.

B. Performance Comparison Using Artificial Neural Network

The fundamental concept of an Artificial Neural Network (ANN) is based on three layers: the Input layer, the hidden layer, and the output layer.

Steps for ANN

Step 1: The dataset is provided to the input layer.

Step 2: The dataset is being initialised and sent to the hidden layer.

Step 3: It has been processed and modified in the hidden layer and sent to the output layer.

Step 4: A connection is created from the input layer to the output layer with modifications

Step 5: The output was provided with an activation function for the final result.

The output is calculated by the function $R_j = \sum M_{ij} N_i$, (9)

R_j is the output Neuron

N_i is the input Neuron

M_{ij} is the weight that connects R_j and N_i .

i varies from 1 to k , j varies from 1 to n

Initially, the normalised, formally clustered data obtained using KNN are fed to the input layer, and the cardiac arrest forecast is computed by computing the weights. A specific weight determines the priority of cardiac arrest. The priority is based on the interval from the selected attributes like (male, female), fasting blood sugar present if it is greater than 120 mg/dl else absent, similarly chest pain was decided according to its gravity as type-1 very severe anginal, type-2 severe anginal type 3 pain without anginal, type 4 no symptom, resting electrocardiographic (RECG) having values 1 and 2 (1 for T wave abnormality and 2 for showing left ventricular hypertrophy). This study considers a neural network (NN) with 30% for training and 70% for testing, achieving the following accuracy, as shown in the figure and Table VIII.

Table VIII Performance Metric

Algorithm used	Accuracy	Time taken
Naïve bayes	87	950 seconds
KNN	95	847 seconds
Hybrid	98	752 seconds

C. Statistical validation

This study used a t-test to estimate a confidence interval for the following samples of patients with typical cardiac arrest/sudden cardiac arrest, and found the average age for cardiac arrest is [51 to 52] years. The study has been conducted and found a confidence interval of [47 to 53] years using the formula $t_{cal} = X \pm T\alpha(0.05,14)$, where X is the sample mean, $T\alpha = 2.145$, i.e., the tabular T value at the 5% significance level (95% confidence). The claim of the average age of cardiac arrest falls within that range. This age limit was determined by analysing 14 hospital samples and comparing them to the set of symptoms identified in this research.

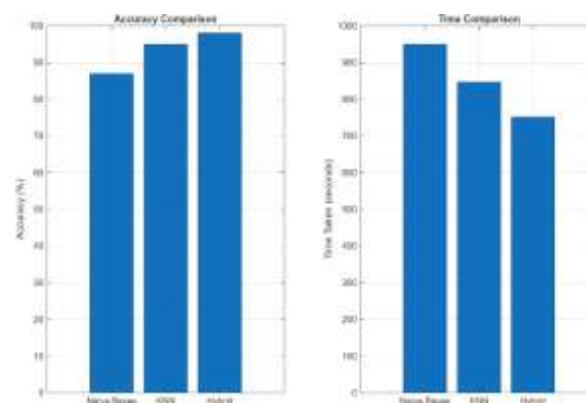


Fig. 3 Accuracy and time comparison

IV. CONCLUSION AND FUTURE WORK

The entire study had included fundamental concepts of soft computing i.e. RST and several ML algorithms for accuracy comparison for the prediction of cardiac arrest. This entire study divided in two parts the first part was using raw and unprocessed data. In the part one Logistic Regression was used to find the dissimilar groups, and RST was applied on these groups to find most fatal diseases it was found to be cardiac arrest in this case. Then Several machine learning algorithms was applied on the normalized universal data set for accuracy comparison. This study can be further extended to the fields of the stock market, entertainment, and the scientific approach to various sports.

REFERENCES

- [1] Hota, Radhanath, Sachikanta Dash, Sujogya Mishra, Sipali Pradhan, P. K. Pattnaik, and Geetanjali Pradhan. "Prediction and diagnosis of thoracic diseases using rough set and machine learning." In *2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 206-213. IEEE, 2023.
- [2] Das, Subhalaxmi, Sateesh Kumar Pradhan, Sujogya Mishra, Sipali Pradhan, and P. K. Pattnaik. "Analysis of heart diseases using soft computing technique." In *2021 19th OITS international conference on information technology (OCIT)*, pp. 178-184. IEEE, 2021.

- [3] Biswas, Soumyajit, Anuvab Sen, and Subhabrata Roy. "A review on early detection of Alzheimer's disease: employing deep learning, machine learning, and statistical methods." *CSI Transactions on ICT* (2025): 1-21.
- [4] Talasila, Vamsidhar, Kotakonda Madhubabu, Meghana Chakravarthy Mahadasyam, Naga Jyothi Atchala, and Lakshmi Sowjanya Kande. "The Prediction of Diseases Using Rough Set Theory with Recurrent Neural Network in Big Data Analytics." *International Journal of Intelligent Engineering & Systems* 13, no. 5 (2020).
- [5] Banerjee, Nikita, and Subhalaxmi Das. "Prediction lung cancer—in machine learning perspective." In *2020 International conference on computer science, engineering and applications (ICCSEA)*, pp. 1-5. IEEE, 2020.
- [6] Das, Subhalaxmi, Sateesh Kumar Pradhan, Sujogya Mishra, Narayan Patra, and Sipali Pradhan. "Classification of pulmonary tuberculosis using mathematical modeling and machine learning." In *2022 International Conference on Machine Learning, Computer Systems and Security (MLCSS)*, pp. 176-182. IEEE, 2022.
- [7] Nayak, Subrata Kumar, Sateesh Kumar Pradhan, Sujogya Mishra, Sipali Pradhan, and P. K. Pattnaik. "Rough set technique to predict symptoms for malaria." In *2021 8th International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 312-317. IEEE, 2021.
- [8] Hota, Radhanath, Sachikant Dash, Sujogya Mishra, Subhalaxmi Das, S. Lenka, and G. Pradhan. "Prediction of Cardiac Arrest using Generalized Soft Computing Techniques." In *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 534-542. IEEE, 2024.
- [9] Nayak, Subrata Kumar, Sateesh Kumar Pradhan, Sujogya Mishra, Sipali Pradhan, and P. K. Pattnaik. "Prediction of Cardiac Arrest by the Hybrid Approach of Soft Computing and Machine Learning." *International Journal of Advanced Computer Science and Applications* 14, no. 7 (2023).
- [10] Nayak, Subrata Kumar, Sujogya Mishra, Subhalaxmi Das, Sipali Pradhan, and Geetanjali Pradhan. "Prediction of Cardiac Arrest Using Artificial Intelligence Methods." In *International Conference on Machine Learning, IoT and Big Data*, pp. 249-261. Singapore: Springer Nature Singapore, 2024.
- [11] Kwon, Joon-myung, Kyung-Hee Kim, Ki-Hyun Jeon, Soo Youn Lee, Jinsik Park, and Byung-Hee Oh. "Artificial intelligence algorithm for predicting cardiac arrest using electrocardiography." *Scandinavian journal of trauma, resuscitation and emergency medicine* 28, no. 1 (2020): 98.
- [12] Nayak, Subrata Kumar, Sateesh Kumar Pradhan, Sujogya Mishra, Sipali Pradhan, and P. K. Pattnaik. "Prediction of cardiac arrest using support vector machine and rough set." In *2022 9th International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 164-172. IEEE, 2022.
- [13] Das, S., S. K. Pradhan, S. Mishra, S. Pradhan, and P. K. Pattnaik. "Analysis of Cardiac Anomalies By Selection And Extraction of Features Using Machine Learning Methods." *Indian Journal of Computer Science and Engineering* 13, no. 6 (2022): 10-21817.
- [14] Das, Subhalaxmi, Sateesh Kumar Pradhan, Sujogya Mishra, Sipali Pradhan, and P. K. Pattnaik. "Diagnosis of cardiac problem using ough set theory and machine learning." *Indian Journal of Computer Science and Engineering* 13, no. 4 (2022): 1112-1131.
- [15] Pawlak, Zdzisław. "Rough sets." *International journal of computer & information sciences* 11, no. 5 (1982): 341-356.
- [16] Nayak, Subrata Kumar, Sateesh Kumar Pradhan, Sujogya Mishra, Narayan Patra, and Sipali Pradhan. "Classification of Symptoms for Malaria using Soft Computing Method." In *2022 International Conference on Machine Learning, Computer Systems and Security (MLCSS)*, pp. 219-223. IEEE, 2022.
- [17] Hota, Radhanath, Sachikanta Dash, Sujogya Mishra, P. K. Pattnaik, and Sali Pradhan. "Early Prediction of Thoracic Diseases Using Rough Set Theory and Machine Learning." In *International Conference on Machine Learning, IoT and Big Data*, pp. 421-433. Singapore: Springer Nature Singapore, 2023.
- [18] <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>
- [19] Dubey, Animesh Kumar, Amit Kumar Sinhal, and Richa Sharma. "Heart disease classification through crow intelligence optimization-based deep learning approach." *International Journal of Information Technology* 16, no. 3 (2024): 1815-1830.