

# Prediction of Cardiac Arrest by the Hybrid Approach of Soft Computing and Machine Learning

Subrata Kumar Nayak<sup>1</sup>, Sateesh Kumar Pradhan<sup>2</sup>, Sujogya Mishra<sup>3</sup>, Sipali Pradhan<sup>4</sup>, P.K. Pattnaik<sup>5</sup>

Department of Computer Science, GVHSS, Sishu Ananta Mahavidyalaya, Balipatna, Odisha, India<sup>1</sup>

Department of Computer Science and Application, Utkal University, Bhubaneswar, Odisha, India<sup>2</sup>

Dept. of Mathematics, Odisha University of Technology and Research Technology, Bhubaneswar, Odisha, 751029, India<sup>3</sup>

Dept. of Computer Science, RBVRR Women's College, Hyderabad, India<sup>4</sup>

Dept. of Mathematics, Odisha University of Technology and Research, Bhubaneswar, Odisha, 751029, India<sup>5</sup>

**Abstract**—Cardiac-related diseases are the major reason for the increased mortality rate. The early predictions of cardiac diseases like ventricular fibrillation (VF) are always challenging for doctors and data analysts. Early prediction of these diseases can save million lives. If the symptoms of these diseases are predicted early, the chance of survival increases significantly. For the prediction of Ventricular fibrillation (VF), several researchers have used Heart Rate Variability Analysis (HRV); various alternatives by combining the features taken from several areas to explore the prediction outcome. Several techniques like spectral Analysis, Rough Set Theory (RST), Support Vector Machine (SVM), and Adaboost techniques have not required any pre-processing. In this work, randomly medical-related data sets are taken from various parts of Odisha, applying regression and Rough Set techniques, reducing the dimension of the data set. Application of Rough Set Theory (RST) on the data set is not only useful in dimension reduction but also gives a set of various alternatives. This work's last section uses a comparative analysis between AdaBoost combined with RST and Empirical mode decomposition (EMD).

**Keywords**—Ventricular fibrillation (VF); heart rate variability (HRV); Rough Set Theory (RST); support vector machine (SVM); regression analysis; Adaboost method

## I. INTRODUCTION

Cardiac-related problems are the most dreadful diseases concerning the death rate. According to World Health Organization (WHO) [1], nearly 18 million people died from cardiac-related problems. The most common cardiac-related diseases are cardiac arrest and sudden cardiac death (SCD). These diseases were responsible for causing nearly 50% of total death caused due to cardiac anomalies. In general, sudden cardiac death occurs due to Cardiovascular problems [2], and this problem doesn't have any prior symptoms of cardiac problems [3,4]. The most general sense of symptoms a person feels just before one hour of the actual occurrences of SCD. A person may not have any symptoms of fatal cardiac conditions but is still prone to SCD [5]. SCD rank second according to mortality rate after cancer [6]. The cause of SCD can be due to various diseases like Ventricular Tachyarrhythmias (VTA), Ventricular Tachycardia (VT), Ventricular Fibrillation (VF), Brady arrhythmia (BA), coronary artery diseases (CAD), valvular diseases (RV), and various genetic factors as discussed by Murukesan et al. [7]. There are several instances where SCD occurs due to

ventricular tachyarrhythmias, including both VF and VT, as Shen et al. discussed [8]; according to their study, ventricular tachyarrhythmias diseases responsible for the improper function of the heart. In most SCD cases, VF action is like a detonator [9-11]. The fatal rate increases by 90% every minute after the VF is detected. So to increase the survival rate, significant emphasis is given to the early detection of VF has notable importance. Forecasting SCD is always significant, as any wrong move can cause this issue to become more severe. There are several instances and examples where SCD is diagnosed wrongly, leading to fatal consequences. Public Access Defibrillation (PAD) method helpful in recovering from sensors can help in saving 2 to 15 lives after the collapse. The success rate of cardiac renewal is the fundamental treatment of SCD which stimulates the heart [12]. It would be ideal to avoid the start of SCD by offering medical support to someone who suffered a cardiac abnormality occurrence, which highlights the question of whether it would be feasible to set up warning devices that could identify cardiac arrest 30 minutes before a crisis as discussed by Fang et al. [13].

## II. BACKGROUND

Several studies have been made with reference to various serious diseases, designing well-organized techniques of forecasting the SCD concluded intrusive and non-intrusive methods as discussed with reference to the article [14–16]. The core objective is to expect the SCD earlier than the ECG signals, as ECG is one of the most important physiological signals to identify cardiac abnormality and electrical conductivity features. Contemporary studies have investigated with properties of both ECG and heart rate variability (HRV), an indication which gets from ECG helps detect the delicate variation that occurs inside the signals earlier than that of the actual occurrence of SCD and to detect the high SCD risk. There are several additional features like time, frequency, and time-frequency. Several researchers have tried to predict SCD by using Machine Learning Algorithm (MLA). Recent studies can detect SCD 25 minutes before its actual occurrence; according to the report [17-18], the researcher used MLA to predict SCD. Several studies were being carried out for perfect diagnosis and early prediction of SCD, but all these data worked in the context of training data, i.e. when the dataset had been properly processed. Several research gaps exist in predicting SCD by various conventional and non-conventional

methods. The following Table I represents the significant research gaps in predicting SCD.

TABLE I. RESEARCH GAPS USING THE CONVENTIONAL METHOD

SL	Conventional research	Drawbacks
1	Soft Computing	Does not work for Random Dataset
2	Linear Regression	Does not work for Outliers
3	Machine Learning tools	Does not work for test data
4	Conventional studies	Machine Error
5	Mathematical Modelling	Only a Numerical solution is possible

A. Literature Survey

Several studies have been carried out to predict SCD or various cardiac diseases. In the context of predicting cardiac arrest, Das et al. [19-20] had various machine learning tools to predict cardiac arrest; they also used hybrid methods like Rough Set Theory (RST) and MLA to boost the accuracy level. RST is a useful tool for prediction, classification, and test and training data. Researchers also used various MLAs like Random Forest Model (RFM) and mathematical modelling to predict meteorological phenomena in this context Das et al. [21-22]. Several research had been conducted to predict various cardiac problems in this context Das et al. [23-25] discussed cardiac diseases using MLA and RST in the same context as Hota et al.[26-27] discussed cardiac problems using various MLA and soft computing techniques. Several researchers have also used mathematical modelling to predict cardiac problems in this context Nayak et al. [28-30] used mathematical modelling and machine learning algorithms to predict diseases like Malaria and Cardiac arrest. Several researchers also used soft computing techniques for social issues in this context Mishra et al. [31] had discussed legal justice for common people in the service sector. The general purpose of research on medical data in computer science is based on the principle of prediction and classifications; in most cases, the researcher used soft computing, MLA, and deep learning techniques for prediction and classifications. To predict cardiac arrest with significant accuracy Manuel et al.[32] had used convolutional neural network (CNN) and ECG images to classify SCD in the same context Kaspal et al.[33] had used a hybrid deep learning method to predict SCD in a similar context Sanchez et al.[34] had used prediction of SCD using wavelet transformation. Scherer et al. [35] had used various methods like pooling operation to significantly detect the objects. Several researchers have used methods like CNN and advanced soft computing methods for prediction and classification other than the medical field in this context Wang et al.[36] had used CNN along with Short-Time Fourier Transform to analyze motor fault analysis. The finding of various researchers considered various parameters like blood pressure and ECG signals for SCD; the ECG signals can predict SCD with significant accuracy, which was observed by monitoring for 24 hours ECG signals from the database of SCDH and considering a sampling frequency of 250 Hz[37]. Several research studies have also been conducted to find the death rate due to cardiac problems in this context Liu et al.[38] had used one-dimensional CNN for health monitoring. This paper used various techniques like RST and machine learning tools to predict cardiac problems. RST was initiated and developed by polish mathematician Pawlak [39]. RST helps

find useful information from vague and imprecise data, as discussed by Das et al. [40]. Several studies have also been conducted using various alternate methods like empirical mode of decomposition (EMD) in this context Huang et al. [41] had discussed the Empirical mode of decomposition and spectrum analysis of Non-linear and Nonstationary time series.

B. Material and Method Used

This work deals with RST, as discussed by Yao and Deng [42], the fundamental concept of RST is based on upper approximations, lower approximations, boundary region, core, reduct, quality of approximation, and indiscernibility. The entire concept of based on indiscernibility, equivalence classes and target set.

C. Fundamentals of RST

The general representation RST is in the form of a table called an information system represented as  $\langle U, C, D \rangle$ , where U is called the universe, C is the set of conditional attributes, and D is the set of decision attributes. The complete skeleton of RST is given in the following Table II.

TABLE II. INITIAL INFORMATION TABLE

E	C	D
E1	1	2
E2	1	1
E3	2	2
E4	2	1
E5	2	1

Description of the Table as follows  $E = \langle E_1, E_2, E_3, E_4, E_5 \rangle$  are the records, C is the set of conditional attributes, D is the set decision attributes/ target attributes, and  $\langle 1,2 \rangle$  represents the values of the decision and conditional attributes.

D. Generalization of RST

The generalization of RST includes a definition of upper approximations, lower approximations, boundary region, core, reduct, quality of the approximation, and indiscernibility and their mathematical representation.

1) *Indiscernibility*: Indiscernibility generally includes a single attribute or multiple attributes. Mathematically indiscernibility defines as

$$IND(K) = \{ (\alpha, \beta) \in R^2 \mid \forall m \in K, m(\alpha) = m(\beta) \} \quad (1)$$

2) *Upper approximation*: Upper Approximation includes the concept possibility, i.e. the set of all objects possibly linked with the target.

$$Upper(AT) = U \{ M \in U/A, : M \cap T \neq \emptyset \} \quad (2)$$

3) *Lower approximation*: Lower Approximation includes the concept of definiteness

$$Lower(AT) = U \{ M \in U/A, : M \subset T \} \quad (3)$$

4) *Reduct*: Reduct forms a set of equivalence classes which is derived from the set of indiscernibility.

E. Various Algorithm used in this Study

1) Algorithm to find Indiscernibility

Step-1  $\forall a \in U$ , where a is the conditional attributes

Step-2  $\forall x, y$  values of the attributes a

Step-3 for  $(i=1; i \leq n; i++)$ , where n is the number of conditional attributes

Step-4 check for indiscernibility i.e. goto step-5

Step-5 if the values of attributes a are either identical or distinct, then indiscernibility

found. i.e. either  $a(x)=a(y)$ , check for the combination attribute 1 to attributes n

else

goto step-3

end if

end for

2) Algorithm to find reduct

Step-1  $\forall IND$  where IND is the indiscernibility

Step-2 for  $(j=1; j \leq n; j++)$ , n is the total number of indiscernibility

Step-3 verify whether the indiscernibility derive reduct

if the values of the indiscernibility are distinct

Reduct R is found else goto step-2

Step-4 end if

Step-5 end for

F. Empirical Mode Decomposition (EMD)

The idea of EMD was initiated by Hung et al., as discussed in the previous section. The EMD approach uses adaptive techniques to analyze stationary and nonstationary one-dimensional signals. EMD has the feature of decomposing a one-dimensional signal in intrinsic mode functions (IMFs). Two significant criteria are essential for the proper function of EMD criteria a. The total number of zero crossing and extrema in the one-dimensional signal either different or equal to numbers only by 1 and b. the lower and upper envelopes must be symmetrical, to estimate the mean value as zero. The following steps are required to calculate each intrinsic mode function (IMFS).

Step-1. Calculate local maxima and local minima of 1 dimensional signal i.e.  $x(t)$

Step-2. Using Step-1 connect the local maxima & minima to calculate the upper & lower envelopes.

Step-3. Find the average of the of the upper & lower envelopes denoted as k

Step-4. New 1dimensional signal to be found out as

$$h_i(t)=x(t)-k(t) \quad (4)$$

Step-5. For  $i= 1$  to n

Check if  $h_k$  satisfies condition 1 and 2

Then Intrinsic mode function  $(IMF_1)=h_k(t)$

Else goto step-5

Step-6 end if

$$\text{Residue signal } r_1(t)=x(t)-IMF_1 \quad (5)$$

Step-7 If the estimated residue signal, function  $r_1(t)$ , has monotonic nature, the method does not work, implying that the estimation of IMFs is completed

Else

Follow the procedure of estimating IMFs and the residue

End if

Step-8 The final

$$x(t)=\sum_{i=1}^n IMF_i - r_n(t), r_n(t) \quad (6)$$

is the last residue found.

On the other hand, if it is not considered a monotonic function, the residual signal is now considered the original 1D signal, which is again evaluated by steps (1 to 3) for estimating the other IMFs.

EMD is the fundamental technique for Ensemble empirical mode decomposition (EEMD) and Complete Ensemble empirical mode decomposition (CEEMD).

G. Ensemble Empirical Mode Decomposition

EEMD, an enhanced form of the EMD technique [23], is categorized by being a noise-assisted technique used to decompose the 1D signals in their basic structures steps for EEMD:

Step-1 Generates new 1D signals by using 1D signal (basic) and basic Gaussian noise  $n_i(t)$

For  $j=1$  to m

$$x_j(t)= x(t) + n_j(t) \quad (7)$$

Step-2 Divide the signal generated by step-2 by the EMD technique

Step-3 Calculate the frequency band or the IMF on the  $k^{\text{th}}$  iteration that will be

$$EEMD_k = \frac{\sum_{j=1}^m c_{jk}(t)}{m} \quad (8)$$

The frequency band or IMF, recognized by k for the trial j is denoted by  $c_{jk}(t)$ .

H. Complete Ensemble Empirical Mode Decomposition

The CEEMD properly described by Torres et al [42] it is the advanced version of EEMD, CEEMD provides better separation than EEMD. Each trial's residue is calculated as  $r_i(t)$  in this context.

$$r_1(t)=x(t)-IMF_1 \quad (9)$$

The flow diagram of the above work is given in the next section for detailing and describing the work. A significant part of the work consists of logistic regression, RST and

EMD, EEMD, CEEMD and comparison of the methods with RST.

1. Fundamentals of Logistic Regression

Logistic regression helps make different classes; this method is also helpful in the case of outliers. This statistical technique is useful for the object with binary outcomes. Logistic regression represented by signum function. The mathematical representation of logistic regression is given as follows as

$$\alpha_a = \left\{ \frac{1}{1+e^{-x}} \right\} \quad (10)$$

III. METHOD AND DATA USED FOR ANALYSIS

This study considered several machine learning algorithms applied to UCI data set for accuracy measure. This work includes three fundamental concepts a. Logistic Regression b. RST to find the significant attributes c. EMD and it two versions for accurate prediction d. Comparison among other methods like Support Vector Machine (SVM), Random Forest Method (RFM) & logistic regression after boosting. The data collected from various districts are given in Table III.

The data is collected from various parts of our states, i.e. Odisha's district, a total of 5,60,000 samples has been taken for the studies. The application of logistic regression reduced the data set into 6 different records. Details of logistic regression are given in the Fig. 1. The logistic regression used for this study has better and more concrete results than linear regression, as logistic regression works better for the outliers. Logistic regression works better for dynamically changing data sets. Training the data set for logistic regression is given as follows. We had considered several cardiac-related diseases from various parts of our state, a total of 5,60,000 data as mentioned in the Table. Using the concept of upper approximation, lower approximation boundary region, i.e., Boundary Region = Upper Approximation - Lower Approximations. Analyzing the Table using the concept of boundary region found that others attribute is insignificant in forming the groups for cardiac problems. This study includes the basis for dividing a larger dataset into groups by training the data set according to the research requirements. This work considered the medical survey according to the research needs; This study divided the entire data sample into categories, i.e. category- 'severe', moderate, average, low, free from cardiac illness, renamed as category-1, category-2,

category-3, category-4, category-5. Using these five categories resulted in 5- distinct groups of 5- records. The records consist of only two values, significant and pointless. Rename the conditional attributes general cardiac problems as 11, Sleeping disorder as 22, Sudden Cardiac death as 33, Pneumonia as 44, Cardiac Asthma as 55, and values significant as a and pointless as b for better understanding and application. When the data samples are divided into categories, this work follows the rules of grouping the data set according to its group; this also follows the concepts of approximations, with a minor difference in diseases between the persons included in one group. The grouping is formed by using the following algorithm. The target values are more than two in our cases with the following assumptions. Target variable can have three or more possible types, which are the five diseases given in the Table III.

In this case, the sigmoid and softmax functions are used for regrouping the sample according to various features. Classes are established using softmax function, and categories or classifications are established using the sigmoid function. P classes softmax function represented by:  $\sigma(x_i) = \frac{e^{x_i}}{\sum_{j=1}^p e^{x_j}}$ , where I varies from 1 to p.

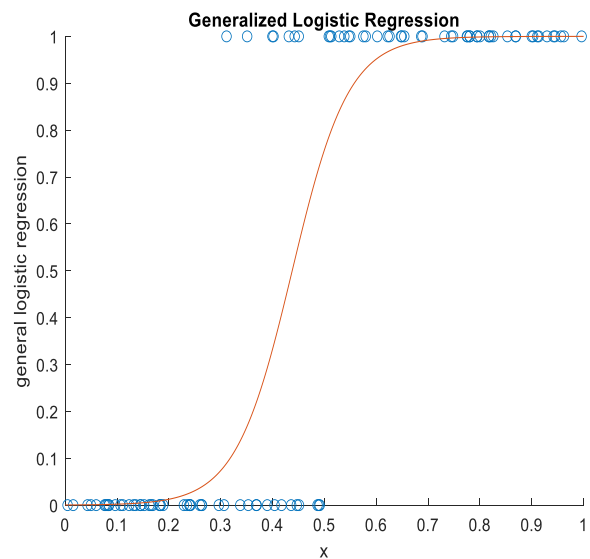


Fig. 1. Application of Logistic regression using the sigmoid function resulted in 6-similar records given in the following Table-IV.

TABLE III. INITIAL DATA TABLE

District	Cardiac problems	Sleeping Disorder	Sudden Cardiac Death	Pneumonia	Cardiac Asthma	Others	Total
Koraput	10,000	50,000	15000	25000	50000	5000	1,55,000
Dhenkanal	20,000	20,000	25000	15000	5000	15000	1,00,000
Jajpur	30,000	10,000	15000	10000	10000	5000	80,000
Cuttack	15,000	25,000	10000	5000	5000	15000	75,000
Puri	25,000	10,000	5000	10000	10000	10000	70,000
Sambalpur	20,000	20,000	15000	5000	15000	5000	80,000

TABLE IV. APPROXIMATION AFTER LOGISTIC REGRESSION

Records	11	22	33	44	55
E1	A	A	a	a	a
E2	A	A	b	b	a
E3	A	B	a	b	b
E4	B	B	b	b	a
E5	B	a	a	b	a
E6	B	a	b	a	b

A. Universal Data Set

Data collected from the University of California Irvine machine learning repository considered UCI repository provided the Heart Disease dataset. This includes 14 different attributes for 304 cases, with target values in deciding whether the patient is suffering from cardiac problems or not, as well as divided the patients according to the symptoms that leads to the finding the number of patients at risk for cardiac problems is extracted from this dataset. There is binary aspect of the dataset i.e. each row corresponds to a single record in this dataset, which has 303 rows and 14 columns, presented in the following Table V and Fig. 2.

The Table represent the detail attribute description and the provided the detail about the target, the figure given below describe the correlation values between -1 to 1. Correlation coefficients range from -1 to +1, indicating a negative or positive correlation. Age, high blood pressure, and high cholesterol levels may strongly correlate with an increased risk of heart disease. Correlation matrices help identify potential risk factors and develop predictive models. This study uses several Machine learning algorithms (MLA). The machine learning algorithms approaches are initially based on weak learner, by combining the multiple algorithm leads to strong learner.

TABLE V. VARIOUS CONDITIONAL ATTRIBUTES RELATED TO CARDIAC

SI	Category	Details	Target
1	Span	Chronic/Not Chronic	Continuous
2	Gender	Patient's Gender	Female/Male
3	CP	Chest Pain	Type(1,2,3,4)
4	Measure	Resting blood pressure	Continuous
5	Ch	Cholesterol range	Continuous
6	Bs	Fasting blood sugar	Less than 120mg/dl
7	Elec	Electrocardiograph	Values(1,2,3,4,5)
8	Tha	Heart beat	Continuous
9	Exa	Induced angina	0/1
10	Op	Amount of rest/ Non rest	Continuous
11	Slope	Peak	Type(Up/Flat/Down)
12	Car	Gives number of major vessels colored by fluoroscopy	Binary(0/1)
13	Tha2	Faulty	Binary(0/1)
14	Num	Cardiac problems	Yes(0,1,2,3,4)

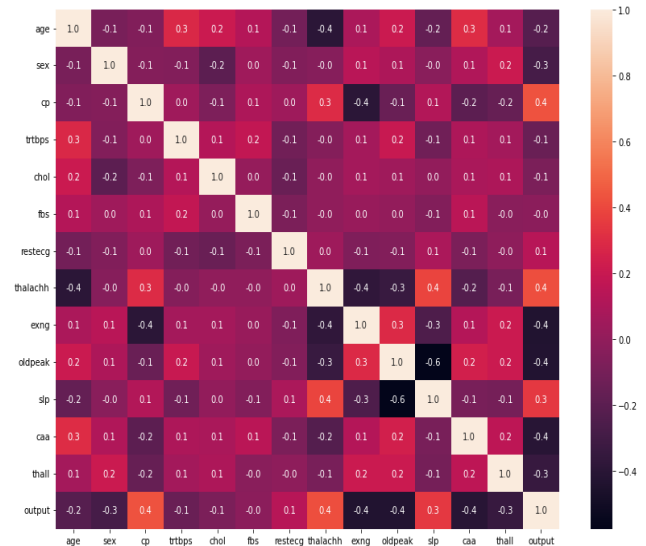


Fig. 2. The correlation matrix.

B. Training by using Naive Bayes Classifiers

Input-: Data sets related to cardiac problems trained with the symptoms given in the rules

Output-: Expected results with binary results Y/N (Y-Yes, N-No)

Step 1. Different category of the dataset trained with binary values (Agree/Disagree). This signifies as an indicator.

Step 2. Expected

$$\text{probability} = \frac{\text{number of objects belongs to the groups (agree)}}{\text{Total samples}}$$

dataset belong to agree category. Similarly, the successive probability calculated by

$$\text{Probability} = \frac{(\text{number of objects belong to group agree after iteration})}{\text{Total samples}}$$

To find the class total number of disagree

Step 3. Find R, where R is the total number of attributes

$$R_y = \sum x_i, \text{ where } x_i \in \text{Agree}$$

$$R_n = \sum y_i, \text{ where } y_i \in \text{Disagree}$$

Step 4. for j 1 to n

1) Find the group level using the formula

Probability for attribute belongs to agree = (Total count for agree belongs to class i)/ (n<sub>i</sub>) Probability for attribute belongs to disagree = (Total count for disagree belongs to class i)/ (n<sub>i</sub>)

Step 5. Divide the new groups of patients belonging to probability R/T, where R represents the groups, and T is its respective features.

Step-6. Compute Probability agree as =  $\prod P(\text{agree})P(\text{Class of attributes belongs to agree class})$

Similarly, compute probability for disagree class by the total product of the probability of disagree × class attributes of disagree class

i=1 to n

g. Assign the new minutes as patients fit to agree class or disagree class, according to higher voting.

### C. Calculating Misclassification before Training

Step-1 Represent P as region belongs to agree class, N as region belongs to disagree class, according to the equation.

$$\sum_{i=0}^{10,000} t_i m_i = 0, k=0.01.$$

Step-2 for i 1 to 10000

Step-3 Verify that the point erroneously classified to the region belongs to agree class

if

{

$$T_i \in N \ \& \ \sum t_i m_i > 0$$

Update  $T_{new} = T_{old} - km_i$

else go to step 5

Step-4 Check for the point misclassified wrongly in the Negative region

$$T \in P \ \& \ \sum t_i m_i < 0$$

Update  $T_{new} = T_{old} + km_i$

else

}

Step-5 end for

### D. Algorithm for Strong Learner using Boosting

Step-1 Adjust the weights as  $1/p$  for  $p$  observation

Step-2 Select the features  $k$  with respect to the smallest Gini index/Concentrated information gain,

Evaluate the over-all error

Step 3 Find the stump of the routine

Step-4 Calculate the new weights with respect to misclassifications

Step-5 All weights to be normalized accordingly

Step-6 For all  $i$  varies from 1 to  $n$

if

{

A configured number of estimators reached the saturation/  
Significant accuracy level

achieved

}

else go to step 2

}

Step-7 end if

Step-8 end for

## IV. RESULT ANALYSIS USING MACHINE LEARNING TOOLS

This study used Max-Min concept to overcome the results with respect to oversampling. This study used UCI sources to provide the diseases related to cardiac. The description of the data set defines in Table-V useful in regulating the cases of a person suffering from cardiac problems. This study divided the domain into two classes: people who suffered from cardiac problems and those who did not suffer from cardiac problems using a set of symptoms. This paper used two concepts of training and testing for people in danger of cardiac problems and recovered from cardiac problems.

### A. Finding Indiscernibility

Finding indiscernibility, reduct and core don't have any generalized algorithm; this study uses the algorithm to find indiscernibility as defined above. This work uses data in Table IV to find indiscernibility, reduct and core. IN denotes indiscernibility.

$$IN(11) = \{(E_1, E_2, E_3), (E_4, E_5, E_6)\} \quad (11)$$

$$IN(22) = \{(E_1, E_2), (E_3, E_4), (E_5, E_6)\} \quad (12)$$

$$IN(33) = \{(E_1, E_3, E_5), (E_2, E_4, E_6)\} \quad (13)$$

$$IN(44) = \{(E_1, E_6), (E_2, E_3, E_4, E_5)\} \quad (14)$$

$$IN(55) = \{(E_1, E_2, E_4, E_5), (E_3, E_6)\} \quad (15)$$

$$IN(11,22) = \{(E_1, E_2), E_3, E_4, (E_5, E_6)\} \quad (16)$$

$$IN(22,33) = \{(E_1, E_5), (E_2, E_6), E_3, E_4\} \quad (17)$$

$$IN(33,44) = \{E_1, (E_2, E_4), (E_3, E_5), E_6\} \quad (18)$$

$$IN(44,55) = \{E_1, E_6, E_3, (E_4, E_5, E_2)\} \quad (19)$$

$$IN(11,33) = \{(E_1, E_3), E_2, (E_4, E_6), E_5\} \quad (20)$$

$$IN(11,44) = \{E_1, (E_2, E_3), (E_4, E_5), E_6\} \quad (21)$$

$$IN(11,55) = \{(E_1, E_2), E_3, (E_4, E_5), E_6\} \quad (22)$$

$$IN(22,33) = \{(E_1, E_5), E_3, (E_2, E_6), E_4\} \quad (23)$$

$$IN(33,44) = \{E_1, (E_2, E_4), (E_3, E_5), E_6\} \quad (24)$$

$$IN(22,44) = \{(E_1, E_6), (E_2, E_5), (E_4, E_3)\} \quad (25)$$

$$IN(22,55) = \{(E_1, E_3, E_5), E_3, E_4, E_5\} \quad (26)$$

$$IN(33,55) = \{(E_1, E_5), (E_2, E_4), E_3, E_6\} \quad (27)$$

$$IN(44,55) = \{E_1, E_3, (E_2, E_4, E_5), E_6\} \quad (28)$$

$$IN(11,22,33) = \{E_1, E_2, E_3, E_4, E_5, E_6\} \quad (29)$$

$$IN(11,33,44) = \{E_1, E_2, E_3, E_4, E_5, E_6\} \quad (30)$$

$$IN(11,44,55) = \{E_1, E_2, E_3, (E_4, E_5), E_6\} \quad (31)$$

$$IN(22,33,44) = \{E_1, E_2, E_3, E_4, E_5, E_6\} \quad (32)$$

$$IN(22,33,55) = \{(E_1, E_5), E_2, E_3, E_4, E_6\} \quad (33)$$

$$IN(22,44,55) = \{E_1, (E_2, E_5), E_3, E_4, E_6\} \quad (34)$$

$$IN(33,44,55) = \{E_1, (E_2, E_4), E_3, E_5, E_6\} \quad (35)$$

$$IN(11,22,33,44) = \{E_1, E_2, E_3, E_4, E_5, E_6\} \quad (36)$$

$$IN(22,33,44,55) = \{E_1, E_2, E_3, E_4, E_5, E_6\} \quad (37)$$

The above equations from (11) to (37) represent the indiscernibility; indiscernibility which don't form groups are called as reduct. The studies lead to the following results. As the following indiscernibility does not produce any groups implies that the following indiscernibility leads to a set of reducts.

$$IN(11,33,44), IN(22,33,44), IN(11,22,33,44), IN(22,33,44,55) \\ \text{Core} = \emptyset \text{ Reduct} \quad (38)$$

In this study, the reduct is found to be

$$\text{Core} = \emptyset IN(11,33,44), IN(22,33,44), IN(11,22,33,44), IN(22,33,44,55) \quad (39)$$

= {33}, as 33 is denoted for SCD as SCD is related to cardiac arrest. This study considered the University of California Irvine machine learning repository considered (UCI). We have used several boosting algorithms to find the accuracy. Further Analysis is being conducted using the strength of RST [27]. The result describes the symptoms of cardiac problems given in the following Table VI.

TABLE VI. FINAL REDUCT TABLE WITH TARGET VALUES

EE	aa <sub>1</sub>	bb <sub>1</sub>	cc <sub>1</sub>	dd <sub>1</sub>	ee <sub>1</sub>	D
EE <sub>1</sub>	1	1	1	0	0	1

In the above Table VI, EE represents the records aa<sub>1</sub>, bb<sub>1</sub>, cc<sub>1</sub>, dd<sub>1</sub>, ee<sub>1</sub> are the conditional attributes. The above attributes are renamed as aa<sub>1</sub> irregular sleeping habit, bb<sub>1</sub> renamed as fluctuation in pulse rate, similarly cc<sub>1</sub>, dd<sub>1</sub>, and ee<sub>1</sub> renamed as breathing difficulties, sweating severely to very severe, and abnormal cholesterol level, respectively. The decision attributes d agree with sudden cardiac death (SCD), <0,1> represents insignificant and significant applicable for both conditional attributes and decision attributes.

### B. Training Phase

The training phase includes the above symptoms taken from Table VII as the input of the study, which is responsible for cardiac arrest or SCD. The training phase used the Naive Bayes classifiers algorithm. The above concept resulted in producing Table VII.

TABLE VII. WEAK LEARNER COMPARISON

Serial	TIDM	Accuracy	Mean AE	Root mean SE	Relative absolute error	Relative SE	F1-Score
Naive B	7	85	0.65	0.85	121	128	0.45
Alternate DT	60	96	0.29	0.49	95	96	0.9
Random Forest	2.23	93	0.32	0.45	84	81	0.87
Reduce Error PT	11.45	97	0.28	0.46	98	98	0.85
C and R tree	57	85	0.29	0.58	99	99	0.87

TABLE VIII. STRONG LEARNER COMPARISON / (ADABOOST COMPARISON)

Serial	TIDM	Accuracy	Mean AE	Root mean SE	Relative absolute error	Relative SE	F1-Score
ABNB	19.5	94	0.5	0.77	132	155	0.87
ABAlternate DT	32	94	0.18	0.28	60	98	0.96
ABRF	11	96	0.11	0.25	37	66	0.98
ABRUPT	65	93	0.21	0.41	46	93	0.85
ABCART	96	95	0.23	0.43	42	92	0.87

The Fig. 3 represents the RMSE (Root Mean Square Error) vs. Mean Absolute Error (MAE) comparison of various classifiers. The classifiers given in the Table define as Naive B-Naive Bayes, Alternate decision tree, reduced error PT-Pruning Tree, C and R- Classification & Regression tree.

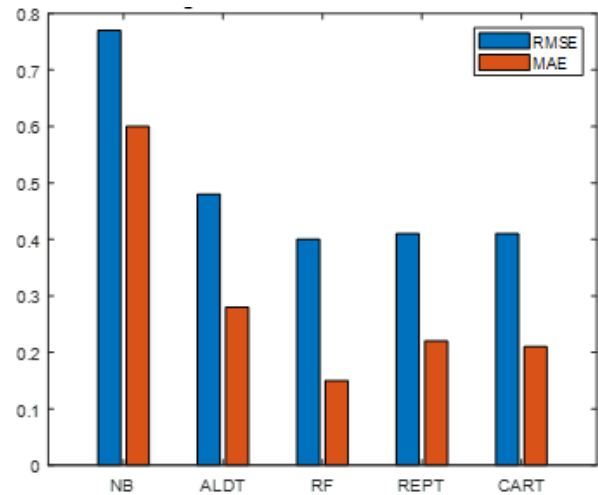


Fig. 3. Error comparison of various classifiers.

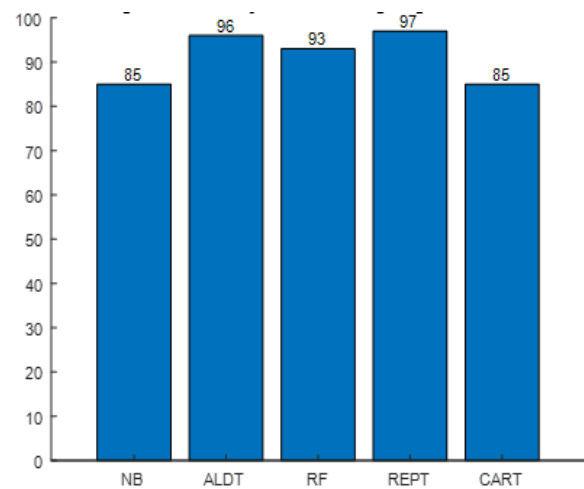


Fig. 4. Comparison of the accuracy of various classifiers.



The Fig. 4 represents the accuracy level of various classifiers with respect to a single classifier application. This study combines various classifiers to get the following results. Applying the combination of two classifiers at a time gives the following output.

ABNB-AdaBoost Naïve Bayes, AB Alternate DT-AdaBoost Alternate decision tree, ABRF- Adaboost Random Forest, ABRUPT- Adaboost Reduced error punning tree, ABCART- Adaboost Correlation and regression tree. There are significant increments of accuracy level after the use of additive strong learner method. The entire result of the output is given in the Fig. 3.

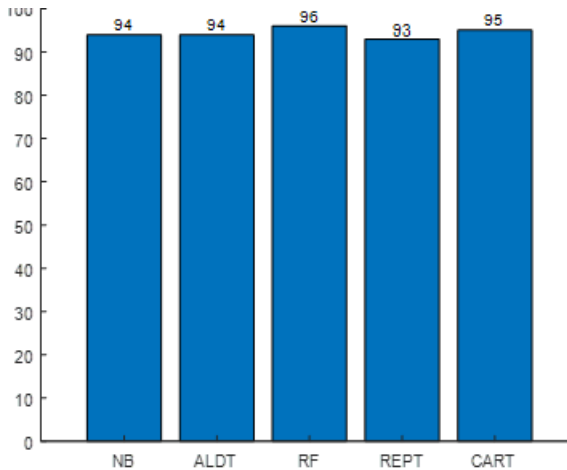


Fig. 5. Accuracy measure of multiple classifiers.

Fig. 5 represents the accuracy level after converting the classifiers from weak learners to strong learner. Subsequently this paper deals with error comparison of single classifiers as well as combinations of classifiers.

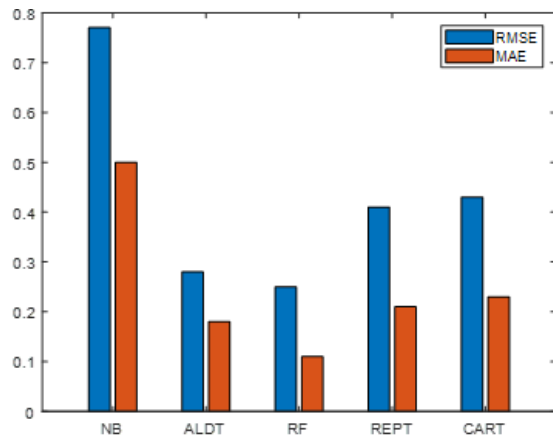


Fig. 6. Error propagation using strong classifier.

The Fig. 6 represents the error rate after using two or more classifiers together. The idea behind this boosting approach is to enhance the accuracy level of prediction and reduced the rate of error margin. This approach also based upon homogeneous classifiers. This concept can be extended for heterogeneous classifiers, with k-fold cross validation and model evaluations. Resulting values using heterogeneous classifiers is given in the following Table IX.

### C. Performance Evaluations

1) *Entropy*: Entropy calculates a structure’s impulsiveness or condition. The idea was given by Rudolf Clausius in 1850 suggested this concept. Mathematical model of this concept was given by

$$\text{Entropy} = - \sum k(x) \log k(x) \quad (40)$$

where  $k(x)$  is the part of the concept of a given class.

2) *Gini index*: Gini Index or Gini Coefficient define as the supply of resources in a given population. In general

$$\text{Gini Impurity} = 1 - \sum_{i=1}^c k_i(x)^2 \quad (41)$$

3) *Information gain*: The decrease of Entropy attained by varying the information (In this study, information derived from the dataset) is called as information gain, and it is normally employed in the training of DT (Decision Tree). This information gain calculation depends upon the Entropy of the prior and posterior transformation of the dataset. The entire information gain is given by the following mathematical formulas.

$$\text{Information Gain } (k_p, f) = I(k_p) - \frac{R_{left}}{R} I(k_{left}) - \frac{R_{right}}{R} I(k_{right}) \quad (42)$$

4) For unbalanced and vague data sets, accuracy measure depended upon the ratio of the number of perfectly classified instances by the total number of sample instances.

$$\text{Accuracy} = \frac{TNs + TP_s}{TNs + TP_s + FN_s + FP_s}, \quad (43)$$

where TN- True negative, TP- True Positive, FN- False Negative, FP- False positive.

Total number positive prediction is called as precision.

$$\text{Precision} = \frac{TP_s}{TP_s + FP_s}, \quad (44)$$

5) *Root mean squared error (RMSE)*, a technique to calculate the forecasting of numeric success

$$\text{Root Mean Squared Error (RMSE)} = \sqrt{\frac{\sum_{i=1}^n (k_i - k_i^t)^2}{n}} \quad (45)$$

6) *Mean Absolute Error (MAE)* is the calculation of the absolute difference between actual and expected values

$$\text{MAE} = \sum_{k=1}^n |l_k - l_k^t| \quad (46)$$

7) *Relative Absolute Error (RAE)*

$$\text{RAE} = \frac{(k_i - k_i^t)^2}{(k_i - k_i^t)^2}, \text{ where } i \text{ varies from 1 to all samples} \quad (47)$$

8) *Total Squared Error*

$$\text{RRSE} = \sqrt{\text{RAE}} \quad (48)$$

Accuracy measure is depicted in table-8

9) *Recall*

$$\text{recall} = \frac{TP_s}{TP_s + FN_s} \quad (49)$$



10) *F1 score*: F1-score is the combination of recall and precision

$$F1\text{-score} = 2 \times \left( \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right) \quad (50)$$

Where  $f$  is the feature split,  $k_p$  is the parent dataset,  $k_{left}$  and  $k_{right}$  are the posterior probability of left and right child node. The equations (39), (40) and (41) are derived from NB classifiers.

**D. Application of the Adaboost Technique**

The Adaboost technique combines multiple weak classifiers to make it a single strong classifier. The most commonly used AdaBoost algorithm is the decision stumps, an alternative name for these decision trees, as discussed by Hussein et al. [43]. This technique generates a model that distributes equal weights at all data points and subsequently provides points to those data points incorrectly classified with higher weights groups. In the next model, significant importance is given to the data points with higher weights. An error margin of  $10^{-3}$  to  $10^{-4}$  is considered. The process continues till the desired accuracy is achieved, as discussed by Reddy et al.[44].

**E. Algorithm for Boosting**

Step-1 Initialize the data points  $(x_i, y_i)$ , for  $i=1$  to  $n$

Step-2 Each  $x_i$  is the instance of space  $X$ ,  $y_i$  is collecting all labels of space  $Y$ .

Step-3 for the training instance  $i$  by rounding  $t$  is given as  $D_{it}$ , Initialize the weights as  $D_{it}(i) = \frac{1}{M}$ ,  $I=1$  to  $M$

Step-4 for the total number of samples, find the weight of misclassified according to the standard algorithm increased each step

$$\alpha_{it} = \frac{1}{2} \ln \left( \frac{L_{+1} - L_{-1}}{L_{-1} + L_{+1}} \right) \quad (51)$$

Step-5 Use basic logistic regression to adjust overfitting

Step-6 Continue with step 1 to 5 till the desired accuracy is achieved.

To overcome oversampling, a technique called Synthetic Minority Oversampling Technique is being used; the following steps are being implemented, and the results are shown in Table VII to Table IX, accuracy and error like RMSE and MAE are shown from Fig. 3 to Fig. 8. The details of SMOTE are described in the subsequent section.

**F. Procedure for SMOTE**

Before the SMOTE algorithm, Data pre-processing is required as the classification algorithm's objective is to collect the raw data and create an outline for every class as conclusive with less error margin to attain accurate prediction. The examples that maintain significant distance from outlines form a group more easily than those that are near the outlines. The examples nearer to the outlines are always a challenge for learning existing classifiers. SMOTE provide a significant result in overcoming these challenges.

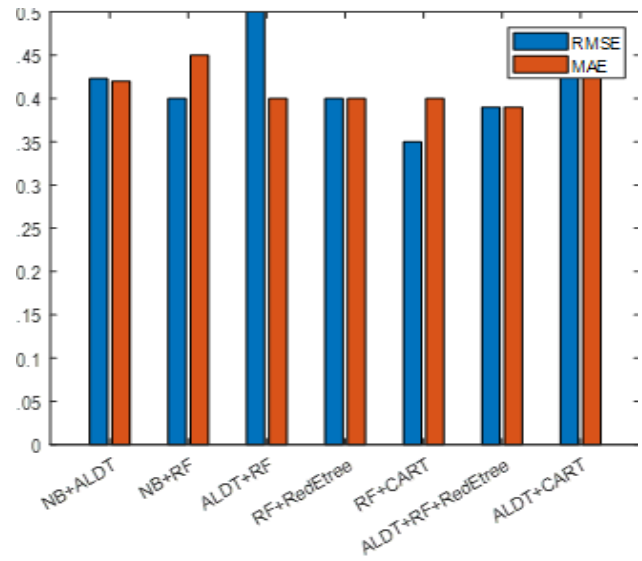


Fig. 7. Error propagation using multiple classifiers.

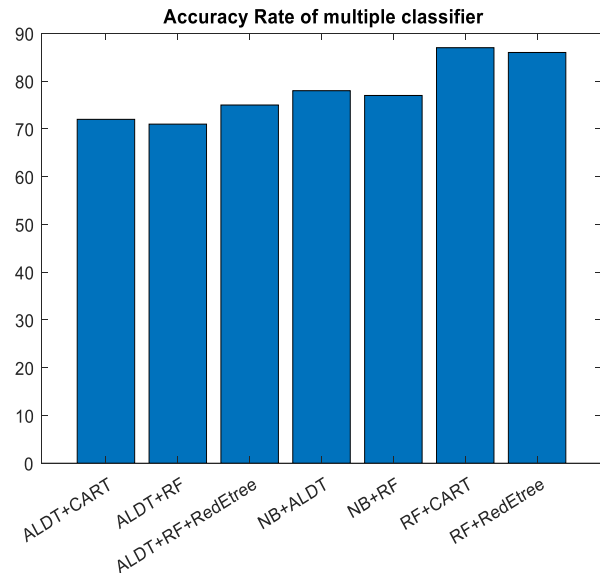


Fig. 8. Accuracy measure using boosting.

TABLE IX. HETEROGENEOUS COMBINATION OF CLASSIFIERS

SL	TIB M	ACCURACY	MAE	RS ME	RA E	RR SE	F1 score
NB+ALDT	29	78	0.42 2	0.42 3	99.23	96.21	0.74
NB+RF	30	77	0.45	0.4	93	98	0.75
AIDT+RF	400	71	0.4	0.5	81	102	0.7
RF+RedEtree	8	86	0.4	0.4	71	92	0.85
RF+CART	7.5	87	0.4	0.35	71	90	90
AIDT+RF+RedEtree	358	75	0.39	0.39	74	94	99.5
AIDT+CART	599	72	0.45	0.45	89	99	0.64

G. SMOTE for Classifications

Step-1 This method's objective is to produce examples, by using the concept

$$T=2 \times (K-M) \tag{52}$$

Where K denotes the majority samples of the group, M denotes the minority samples of the group and T initial examples are newly generated. The basic examples generated by SMOTE will be recognized or excluded depending upon two concepts i.e.

Stage-1{  $\widehat{k}_1, \widehat{k}_2, \widehat{k}_3, \dots, \widehat{k}_n$  } is the group of example and  $\widehat{k}_i(j)$  is the  $j^{th}$  attribute values of  $\widehat{k}_i$  example.

Where j varies from 1 to the total number of samples. Let  $R_a$  and  $R_b$  be the minority and majority examples, respectively. The distance d is calculated between  $\widehat{k}_i$  and  $R_a$  and  $R_b$  for recognition and exclusion. The distance between  $K_{minority}(\widehat{k}_i, R_a)$ ,  $K_{majority}(\widehat{k}_i, R_b)$ . There will be total number of n steps. The distance is calculated as

$$K_{minority}(\widehat{k}_i, R_a) = \sum_{i=1}^p \sqrt{((\widehat{k}_i - R_a))^2} \tag{53}$$

Similarly

$$K_{majority}(\widehat{k}_i, R_b) = \sum_{i=1}^p \sqrt{((\widehat{k}_i - R_b))^2} \tag{54}$$

Using equations 52 and 53, majority and minority distances are calculated, as

$$K_{minority}(\widehat{k}_i, R_a) = (K_1, K_2, \dots, K_n) \tag{55}$$

$$K_{majority}(\widehat{k}_i, R_b) = (L_1, L_2, \dots, L_m) \tag{56}$$

From equation (54), calculate min (  $K_1, K_2, \dots, K_n$  ) to find the minimum of a minority; similarly, calculate min (  $L_1, L_2, \dots, L_m$  ) to find a minimum of majority.

Step-2 If minimum (minority) <= minimum(majority) out coming examples are accepted else rejected.

Step-3 continue with step 1 to 3 for the entire samples to achieve the required outcomes.

TABLE X. A COMPARATIVE ANALYSIS

1	comparing the results of [17]
	Uses local data collected from a particular Hospital/area
	Not a generalized one. This study overcomes these problems of prediction
2	comparing the results of [32]
	Uses image analysis rather than symptoms this method was not general
3	Compared with [35] also uses the method of object recognition is also not a general assumption
4	Overall coAll is this study deals with vague and imprecise data for general purposes, the data not collected from a particular area and not collected from a particular hospital

The detailed comparison of error analysis and accuracy level was given in Fig. 7 and 8. This study has compared its results with several cited papers in Table X.

H. Statistical Validation of the Result

This study uses statistical methods for the validation of the claim. There are several statistical methods available for Analysis and validation. The generally adopted method for validation is the chi-squared statistical test. This study uses two fundamental statistical techniques, i.e., one dimensional  $\chi^2$  (Chi-squared test) and two-dimensional chi-squared test for this study.

1) Statistical validation using one dimensional  $\chi^2$  test

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{57}$$

$H_0$  (Null Hypothesis)-: The above multiple classifiers are not provided accuracy as desired

$H_a$  (Alternate Hypothesis)-: The above multiple classifiers provide desired results as required.

Where  $O_i$  is the observed samples, and  $E_i$  is the expected samples. A survey is conducted over 10,000 population by grouping 10,000 per group total of 10 available groups. The observed values are 35,5,10,5,5,10,3, 7,15,5, unit measure in 10,000. With expected values 10%, 25%, 5%, 5%, 25%, 30%, 45%, 5%, 10%, 15%. So total observed samples are 100,000, and the expected values are 10,25,5,5,25,30,45,5, 10,15.

Calculated  $\chi^2 = 160$ ,  $\chi^2 (9, 0.05) = 16.919$ , much less than the calculated  $\chi^2$  value, so rejected the Null hypothesis. The same data set was also applied on a two-dimensional array and the calculated  $\chi^2 = 21.25$  where the tabular values were calculated as 17.23. The null hypothesis is rejected.

V. CONCLUSION AND FUTURE WORK

This study entirely depended upon a vague and imprecise dataset from various parts of our state's Odisha and medical advisory. The initial idea is to group the data set according to its class. The dataset was divided into six categories using logistic regression. Using two concepts of RST, i.e., indiscernibility and strength, showed that SCD/ Cardiac -arrest was the most significant disease among all cardiac-related problems. The subsequent section of the paper used the boosting technique combining several classifiers to measure the error rate and accuracy level. As the combination increases, the accuracy also increases. There was also a comparative analysis of these studies and EMD and EEMD studies in Table X. These studies can be extended to the fields like sports especially making uniforms on cricket fields worldwide, and entertainments like movies more entertaining for common people.

REFERENCES

[1] WHO, "WHO | Non communicable diseases," Who, 2017. [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs355/en/>. [Accessed: 25-Oct-2017].

[2] Rea T.D. Page R.L. Community approaches to improve resuscitation after out-of-hospital sudden cardiac arrest. Circulation. 2010 Mar 9;121(9):1134-40.

- [3] Deo R, Albert CM. Epidemiology and genetics of sudden cardiac death. *Circulation*. 2012 Jan 31;125(4):620-37.
- [4] Fishman GI, Chugh SS, DiMarco JP, Albert CM, Anderson ME, Bonow RO, Buxton AE, Chen PS, Estes M, Jouven X, Kwong R. Sudden cardiac death prediction and prevention: report from a National Heart, Lung, and Blood Institute and Heart Rhythm Society Workshop. *Circulation*. 2010 Nov 30;122(22):2335-48.
- [5] Myerburg RJ. Cardiac arrest and sudden cardiac death. *Heart disease: a textbook of cardiovascular medicine*. 1997:742-56.
- [6] Passman R. Prevention of sudden cardiac death in dialysis patients: drugs, defibrillators or what else?. *Blood purification*. 2013 Jun 1;35(1-3):49-54.
- [7] Murugesan L, Murugappan M, Iqbal M, Saravanan K. Machine learning approach for sudden cardiac arrest prediction based on optimal heart rate variability features. *Journal of Medical Imaging and Health Informatics*. 2014 Aug 1;4(4):521-32.
- [8] Shen TW, Shen HP, Lin CH, Ou YL. Detection and prediction of sudden cardiac death (SCD) for personal healthcare. In2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2007 Aug 22 (pp. 2575-2578).
- [9] Passman R, Goldberger JJ. Predicting the future: risk stratification for sudden cardiac death in patients with left ventricular dysfunction. *Circulation*. 2012 Jun 19;125(24):3031-7.
- [10] Pagidipati NJ, Gaziano TA. Estimating deaths from cardiovascular disease: a review of global methodologies of mortality measurement. *Circulation*. 2013 Feb 12;127(6):749-56.
- [11] Zheng Y, Wei D, Zhu X, Chen W, Fukuda K, Shimokawa H. Ventricular fibrillation mechanisms and cardiac restitutions: An investigation by simulation study on whole-heart model. *Computers in biology and medicine*. 2015 Aug 1;63:261-8.
- [12] Aziz EF, Javed F, Pratap B, Herzog E. Strategies for the prevention and treatment of sudden cardiac death. *Open access emergency medicine: OAEM*. 2010;2:99.
- [13] Fang Z, Lai D, Ge X, Wu X. Successive ECG telemetry monitoring for preventing sudden cardiac death. In2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2009 Sep 3 (pp. 1738-1741).
- [14] Huikuri HV, Tapanainen JM, Lindgren K, Raatikainen P, Mäkilä TH, Juhani Airaksinen KE, Myerburg RJ. Prediction of sudden cardiac death after myocardial infarction in the beta-blocking era. *Journal of the American College of Cardiology*. 2003 Aug 20;42(4):652-8.
- [15] Hallstrom, A.P.; Stein, P.K.; Schneider, R.; Hodges, M.; Schmidt, G.; Ulm, K.; CAST Investigators. Characteristics of heart beat intervals and prediction of death. *Int. J. Cardiol*. 2005, 100, 37–45.
- [16] La Revere, M. Baroreflex sensitivity and heart-rate variability in prediction of total cardiac mortality after myocardial infarction. *Lancet* 1998, 351, 478-484.
- [17] Vargas-Lopez, O.; Amezcua-Sanchez, J.P.; De-Santiago-Perez, J.J.; Rivera-Guillen, J.R.; Valtierra-Rodriguez, M.; Toledano-Ayala, M.; Perez-Ramirez, C.A. A New Methodology Based on EMD and Nonlinear Measurements for Sudden Cardiac Death Detection. *Sensors* 2020, 20, 9.
- [18] Amezcua-Sanchez, J.P.; Valtierra-Rodriguez, M.; Adeli, H.; Perez-Ramirez, C.A. A novel wavelet transform-homogeneity model for sudden cardiac death prediction using ECG signals. *J. Med. Syst*. 2018, 42, 176.
- [19] Das S, Pradhan SK, Mishra S, Pradhan S, Pattnaik PK. ANALYSIS OF CARDIAC ANOMALIES BY SELECTION AND EXTRACTION OF FEATURES USING MACHINE LEARNING METHODS.
- [20] Das S, Pradhan SK, Mishra S, Pradhan S, Pattnaik PK. Diagnosis of cardiac problem using rough set theory and machine learning. *Indian Journal of Computer Science and Engineering*. 2022;13(4):1112-31.
- [21] Das R, Mishra J, Mishra S, Pattnaik PK. Das S. Mathematical Modeling using Rough Set and Random Forest Model to Predict Wind Speed. In2022 9th International Conference on Computing for Sustainable Global Development (INDIACom) 2022 Mar 23 (pp. 207-213). IEEE.
- [22] Das R, Mishra J, Mishra S, Pattnaik PK. Design of mathematical model for the prediction of rainfall. *Journal of Interdisciplinary Mathematics*. 2022 Apr 3;25(3):587-613.
- [23] Das S, Pradhan SK, Mishra S, Patra N, Pradhan S. Classification of Pulmonary Tuberculosis using Mathematical Modeling and Machine Learning. In2022 International Conference on Machine Learning, Computer Systems and Security (MLCSS) 2022 Aug 5 (pp. 176-182). IEEE.
- [24] Das S, Pradhan SK, Mishra S, Pradhan S, Pattnaik PK. A Machine Learning based Approach for Detection of Pneumonia by Analyzing Chest X-Ray Images. In2022 9th International Conference on Computing for Sustainable Global Development (INDIACom) 2022 Mar 23 (pp. 177-183). IEEE.
- [25] Das S, Pradhan SK, Mishra S, Pradhan S, Pattnaik PK. Analysis of heart diseases using soft computing technique. In2021 19th OITS International Conference on Information Technology (OCIT) 2021 Dec 16 (pp. 178-184). IEEE.
- [26] Hota R, Dash S, Mishra S, Pradhan S, Pattnaik PK. Symptoms Prediction of Tuberculosis using Soft Computing Technique. In2022 OITS International Conference on Information Technology (OCIT) 2022 Dec 14 (pp. 343-347). IEEE.
- [27] Hota R, Dash S, Mishra S, Pradhan S, Pattnaik PK, Pradhan G. Prediction and Diagnosis of Thoracic Diseases using Rough Set and Machine Learning. In2023 10th International Conference on Computing for Sustainable Global Development (INDIACom) 2023 Mar 15 (pp. 206-213). IEEE.
- [28] Navak SK, Pradhan SK, Mishra S, Patra N, Pradhan S. Classification of Symptoms for Malaria using Soft Computing Method. In2022 International Conference on Machine Learning, Computer Systems and Security (MLCSS) 2022 Aug 5 (pp. 219-223). IEEE.
- [29] Navak SK, Pradhan SK, Mishra S, Pradhan S, Pattnaik PK. Rough set technique to predict symptoms for malaria. In2021 8th International Conference on Computing for Sustainable Global Development (INDIACom) 2021 Mar 17 (pp. 312-317). IEEE.
- [30] Navak SK, Pradhan SK, Mishra S, Pradhan S, Pattnaik PK. Prediction of Cardiac Arrest Using Support Vector Machine and Rough Set. In2022 9th International Conference on Computing for Sustainable Global Development (INDIACom) 2022 Mar 23 (pp. 164-172). IEEE.
- [31] Mishra S, Mohanty SP, Pradhan SK. Reasons for employees need justice from legal bodies: A rough set approach. In2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom) 2016 Mar 16 (pp. 2018-2022). IEEE.
- [32] Manuel A. Centeno-Bautista, Angel H. Rangel-Rodriguez, Andrea V. Perez-Sanchez, Juan P. Amezcua-Sanchez, David Granados-Lieberman and Martin Valtierra-Rodriguez, Electrocardiogram Analysis by Means of Empirical Mode Decomposition-Based Methods and Convolutional Neural Networks for Sudden Cardiac Death Detection, *Applied science, Appl. Sci.* 2023, 13(6), 3569; <https://doi.org/10.3390/app13063569>
- [33] Kaspal, R., Alsaadon, A., Prasad, P.W.C., Al-Saiyd, N.A.; Nguyen, T.Q.V.; Pham, D.T.H. A novel approach for early prediction of sudden cardiac death (SCD) using hybrid deep learning. *Multimed. Tools Appl*. 2021, 80, 8063-8090.
- [34] Perez-Sanchez, A.V.; Valtierra-Rodriguez, M.; Perez-Ramirez, C.A.; De-Santiago-Perez, J.J.; Amezcua-Sanchez, J.P. Epileptic seizure prediction using Wavelet Transform, Fractal Dimension, Support Vector Machine, and EEG signals. *Fractals* 2022,30,2250154, <https://doi.org/10.1142/S0218348X22501547>
- [35] Scherer, D.; Müller, A.; Behnke, S. Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition. In Proceedings of the Artificial Neural Networks-ICANN 2010, Thessaloniki, Greece, 15-18 September 2010; pp. 92-101
- [36] Wang, L.H.; Zhao, X.P.; Wu, J.X.; Xie, Y.Y.; Zhang, Y.H. Motor Fault Diagnosis Based on Short-Time Fourier Transform and Convolutional Neural Network. *Chin. J. Mech. Eng. Engl. Ed*. 2017, 30, 1357-1368.
- [37] Goldberger, A., L. Amaral, L. Glass, J. Hausdorff, P. C. Ivanov, R. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley. "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new

- research resource for complex physiologic signals. *Circulation* [Online]. 101 (23), pp. e215–e220." (2000).
- [38] Liu, T.; Xu, H.; Ragulskis, M.; Cao, M.; Ostachowicz, W. A Data-Driven Damage Identification Framework Based on Transmissibility Function Datasets and One-Dimensional Convolutional Neural Networks: Verification on a Structural Health Monitoring Benchmark Structure. *Sensors* 2020, 20, 1059
- [39] .Z.Pawlak *International Journal of Computer & Information Sciences* volume 11, pages341–356(1982).
- [40] Das S, Pradhan SK, Mishra S, Pradhan S, Pattnaik PK. Prediction of Heart Diseases Using Soft Computing Technique. In *Intelligent Systems: Proceedings of ICMB 2021* 2022 May 4 (pp. 155-167). Singapore: Springer Nature Singapore.
- [41] Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proceedings of Royal Society A Mathematical, Physical and Engineering sciences*, 1998.
- [42] Yiyu Yao, Xiaofei Deng, Quantitative rough sets based on subset hood measures, *Information sciences*, Volume 267, 20 May 2014, Pages 306-322
- [43] Hussein AS, Li T, Yohannese CW, Bashir K. A-SMOTE: A new preprocessing approach for highly imbalanced datasets by improving SMOTE. *International Journal of Computational Intelligence Systems*. 2019 Jan 1;12(2):1412-22.
- [44] Reddy PC, Chandra RM, Vadiraj P, Reddy MA, Mahesh TR, Madhuri GS. Detection of plant leaf-based diseases using machine learning approach. In *2021 IEEE International conference on computation system and information technology for sustainable solutions (CSITSS) 2021* Dec 16 (pp. 1-4).