A Hybrid Machine Learning Approach For Heart Disease Classification Using KNN And SVM Method

Shweta Bharati, Rajeev Kumar Gupta, Vinay Singh, Nikhlesh Pathik

Abstract: In the current scenario, heart disease is one of the main cause of death in all over the world. If hard diseases predict in advance, it will very helpful for the doctor to diagnose at the early state and prescribe the medicine accordingly. Machine learning approach can be used to predict heart disease in advance. Main aim of this study is developed a heart disease prediction model using Cleveland Heart Disease dataset. In order to evaluate the performance of the heart disease prediction model three measurements named accuracy, specificity, and sensitivity are used. The identifications of symptoms in patients are called sensitivity. Usually the incorrect identification of patient's stage is the specificity. The proposed model is compare with the tradition SVM and KNN and found that proposed model diagnose heart disease more accurately.

Index Terms: Data Mining, Coronary Heart Disease, KNN-classifier, Support Vector Machine, Machine Learning, Confusion matrix, Precision, Recall.

——————————————— ◆ ———————————————

1. INTRODUCTION

The main part of physique to pump and purify blood in heart especially for humanoid form the heart performance is an important role towards the appropriate blood provision by means of essential part of the body, without proper work of heart humanoid life is unable to survive. Heart is a body organ playing a beautiful role from all others organs to blood vessels in circulatory system. When the system of blood circulation is not proper, some organs especially like head can be improper to work also function of heart can be irregular or stop and death can be happen. The heart disease refers to the function of heart as well as cardiovascular system. Heart is an essential body organ for life of livings especially heart plays a great role in human body. Heart is the most useful part of humanoid body means life is properly based on the appropriate function of heart. When heart function is improper it may be great harm to other parts of physique also. We are discussing the heart is a significant structure for all living wage to pumping the blood into all other body parts of physique via blood vessels of the circulatory system.

Increases harm factors of heart syndrome:
- Domestic antiquity
- Stage
- Smoking
- Reduced food
- In height plasma compression
- In height plasma cholesterol
- Fatness
- Mental dormancy
- Headache

————————————————————

- Shweta Bharati is currently pursuing masters degree program in Computer Science and Engineering in Sagar Institute of Science and Technology, India, shwetabharati88@gmail.com.
- Dr. Rajeev Kumar Gupta is currently working as an associate professor in the department of Computer Science and Engineering at Sagar Institute of Science and Technology, India, rajeevmanit12276@gmail.com.
- Prof. Vinay Singh is currently working as an assistant professor in the department of Computer Science and Engineering at Sagar Institute of Science and Technology, India, vinaysingh@sistec.ac.in.
- Prof. Nikhlesh Pathik is currently working as an associate professor in the department of Computer Science and Engineering at Sagar Institute of Science and Technology, India, nikhleshpathik@sistec.ac.in.

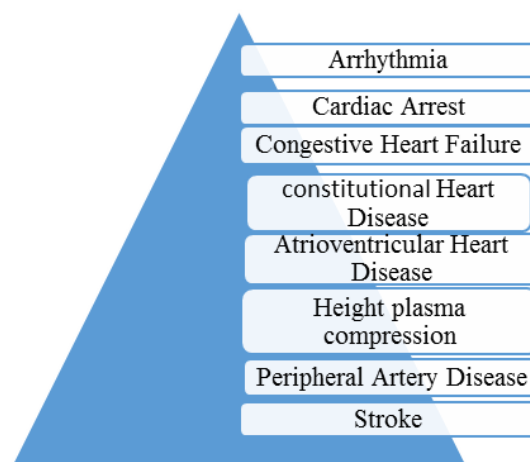### 1.1 Classification of Heart Disease: -

Figure 1: Types of heart disease

1.2 Coronary Heart Disease

This kind of heart disease is the great cause to death. In this heart disease patients are becoming in a large number every year as per summarization from World Health Organization (WHO) founded that 175 million populations is caused towards death in 2015, presenting 31% of totally death besides an average of 7,4 million supposed by coronary heart infection. The actual indication of CHD (Coronary Heart Disease) discover chest discomfort or else angina, this looks imprecise and angina contain be happening in conditions that may not be accompanied by coronary heart disease.

1.3 Symptoms of CHD (Coronary Heart Disease)

Coronary Heart Disease describes lots of symptoms such as given below:-

- Angina (Chest Pain): Angina is a physical task or challenging situation. It is a type of chest pain effect by decreases oxygen in the blood flow to the heart.
- Heart Attack (Arteries blocked): Heart attack is a section of death of the heart muscle starting from by a blood supply loss. Heart attack is a section of death of the body muscle starting from by a blood supply loss. It is a serious

2790

medical exigency in which the blood contribution to the heart is instantly choked, using by a blood clot and a cardiac arrest includes the heart stopping and the pumping of blood around the body. A Heart attack can conduct to cardiac heart arrest.

Symptoms of a heart attack can include like:
- Feeling lightheaded or feeble
- Shortness of breath
- Chest pain
- Drudgery

3) Heart Failure (Weak Blood Pump): Heart failure is a very serious condition, and there is usually no care, though with the accurate treatment, the patients can recreation controlling an agreeable, strategic and creative life. It can be continuing (chronic), otherwise your situation may well perhaps begin immediately (acute). At the present heart failure, is core pumping chambers of your heart may be transformed inflexible not and properly fill stuck between thumps. In some incident of heart failure, your heart influence may come to be injured and weekend that the heart cannot pump lifeblood hyperactive efficient throughout your physique. Heart Failure can contain the left side (left ventricle), right side (right ventricle) or left/ right both on the sideways heart displeasure of your sentiment. Basically heart failure activates with the left ventricle, especially in left side presented important pumping chamber.

Heart failure signs and symptoms include:
- Breathing difficulty at rest
- Feeling exhausted
- Swelling in your legs and feet
- Irregular or quick heart tired
- Moderate ability of training

Table 1: Classification of Heart failure

Types of heart failure	Explanation
Left-Ventricle heart failure	Left heart is responsible for in the body pumping of blood is to be rest condition and fluid may back up in your legs, causing fluid buildup and shortness of breath.
Right-Ventricle heart failure	Right heart pumps plasma to the lungs where it collects oxygen and fluid may back up into your abdomen, legs and feet, caused by way of swelling in limbs.
Diastolic heart failure	Heart muscle is unbending but then normal and left side cannot contract vigorously, satisfying issues.
Systolic heart failure	Heart inability to pump efficiently after filling with blood and left side cannot contract vigorously, pointing forcing problems.

2. LITERATURE SURVEY

Sarath Babu et al. [1] in this paper heart disease diagnosis using data mining techniques likes- Genetic algorithm, K-mean algorithm, MAFIA algorithm and decision tree methods. This classification detecting heart disease and put on predicting the treatment of heart, which heart disease diagnosis handlings identification of treatment specified reasonable price. P.K. Anooj et al. [2] in this paper presented by the weighted fuzzy rule based clinical support system for computer aided diagnosis of the heart disease. This system will inevitably recover knowledge from the patient's data. The clinical system for the prediction of heart disease consist of two stage such as:- (1) Generated weighted fuzzy rules (2) Apply data mining procedure rules (3) Expanding weighted fuzzy rule based on the decision support system. These fuzzy rules construct clinical decision support system using Mamdani fuzzy assumption system. These decision support system upgrade important network based system in terms of accurate, efficient and sensitive. Kensaku Kawamoto et al. [3] presented a decision support system and develop the patient care momentous. Implementation of medical decision system that included features similar flexible and appropriate. Nidhi Bhatla et al. [4] presented a various studies in data mining techniques, which are used in heart disease prediction. Observation discloses to the neural networks and decision tree and work on 15 attributes. These attribute show the 100% highest accuracy in neural network and 99.62% accuracy show in decision tree. Decision tree offer the 99.20% efficiency. Boshra Bahrami et al. [5] presented by diagnosis heart disease classification methods such as: Decision Tree, KNN, and Naïve Bayes. After the classification and performance evaluation the Decision tree is considered as the best classifier for heart disease diagnosis from the dataset.

Deepika N et al. [6] in this paper presented by Pruning Classification Association Rule (PCAR). PCAR rule comes prefer Apriori algorithm from Data Mining technique. These method deletes lowest frequent item with minimum frequent item-sets and deletes infrequent items from item-sets then generated frequent item-set. Subbalakshmi et al. [7] in this paper proposed by needing specific patient data and prefer long time medical experience. The health maintenance data which are appreciatively improve unseen information for effectual collecting vast amount of decision making process.

3. DATASET

In this research total employee's 303 data of Cleveland heart disease (CHD) dataset from UCI machine learning repository which have seven data of misplaced values.

Table 2: Summary of Cleveland Heart Disease (CHD) Dataset

Attribute	Attribute Description	Description Type	Typical Range
Age	Age in years	Numeric	25-65
Sex	Sex in number	Nominal	0:Female; 1:Male;
Cp	Chest pain	Nominal	1:typical angina; 2:atypical angina; 3:non-anginal pain; 4:asymptomatic
Trestbps	Resting Blood Pressure	Numeric	95-200
Chol	Cholestrol in 120mg/dl	Numeric	125-565
Fbs	Level of Fasting Blood Sugar	Nominal	0: False; 1: True;
Restecg	Resting electro cardiographic results	Nominal	0: Normal; 1: having ST-T wave abnormality;
Thalac	Max heart Rate Achieved	Numeric	82-185
Exang	Exercise induced	Nominal	Yes = 1;

	angina		No = 0;
Oldpeak	Segment ST depression induced by exercise relative to test	Numeric	70-200
peak Slope	The slope of the peak exercise ST segment	Numeric	(1-3) 1: usloping; 2: flat; 3:downsloping;
Ca	Number of major vessels (0-3) colored by fluoroscopy	Numeric	(0-3) 0,1,2 & 3
Thal	The defect type of the heart	Nominal	(3,6 & 7) 3: Normal; 6: fixed defect; 7: reversible defect;
Disease	Identification of the heart attack	Nominal	Yes=2; No=1 ;

4. METHODOLOGY

Describe the mythological approaches such as follows: Pre-processing, normalization, neural network, SVM and KNN are involve the process of data mining techniques
.

4.1 Preprocessing

Preprocessing is a data mining methods that included transforming rare statistics into a comprehensible arrangement. Rare amount of data-set is extremely predisposed to noisy, missing value and irregularity.

4.1.1 Data Preprocessing

a) **Data Cleaning:** - It is the progression of deleting and removing the inaccurate records from the huge datasets.

   Missing value – usage the best probable assessment to fill in the missing value.

   Noisy data – Regression, Clustering.

b) **Data Integration/ Normalization:** - It is the evolution of combining to participate (compile) multiple data. It is used to syndicate various data from different source into target data and meaningful info.

c) **Data Transformation: -** It is the progression of representing the map and transforming the data as one interested format to another format or structure. Which data are providing the appropriate forms for mining by conduct compact or concentricity operation?
   - Smoothing - Cleaning of data
   - Accumulation - Reduction of data
   - Generalization – Reduction of data
   - Standardization – min-max, z- score, decimal scaling

d) **Reduction/Feature extraction:-**
   - Decrease of data cube
   - Collection of quality subsection
   - Dimensionally reduction
   - Selection of data
   - Hierarchical concept

e) **Discretization/Selection: -** It stands for the progression of determining the relevant data source, further collect

acceptable tools. The situation exteriors the genuine practice of data collection.

4.2 Normalization

Min-Max standardization, Z-Score standardization and Decimal Scaling methods are the techniques of normalization with respect to privacy and accuracy. Practically applied to the unique k-mean clustering algorithm and personalized data to authenticate the effectiveness and the precision of our proposed approach. Algorithm of Normalization
1) Appoint any size of the data range.
2) Variety of dataset to read the source code.
3) Using scale range techniques.
4) Used generated scale data interested in further handling as per our requirement.
5) Then, newly generated scale up (if needed).
6) Finish

4.2.1 Min-Max standardization

It executes a direct modification arranged to the unusual information. The standards are regulated inside the specified variety. Designed intended for plotting a significance, of an element X since variety [minX, maxX] to an original range [new_minX , new_maxX], the calculation is specified through, Transmute the information from measured components to an original interval from new_minX to new_maxX for feature F:

$$\left(X = \frac{X - \min value\ of\ X}{\max value\ of\ X - \min value\ of\ X} \right)$$
$$* (S - R) + R$$

4.2.2 Z-Score Standardization

Z score normalization, correspondingly known as Zero mean standardization. At this time the data is regular constructed arranged the mean as well as ordinary deviancy. Then the formula is,

$$Z = X - min / max - min$$

Here, Z is the standardized value of member of standard observed values x, min means minimum and max is known as maximum. X given its range. To transform it into particular range, then;
The general formula is;
   v´= (v-min)/ (max-min) * (new_max-new_min) + new_min
Where, v = old variable, v´ = transformed variable, new_min = minimum of the stabilized dataset, new_max = maximum of the stabilized dataset.
   v = [min, max], and v´ = [new_min, new_max]

4.3 Neural Network

It is determined the set of algorithm that has transformed machine learning. They are superlative by genetic neural network and which are co-ordinate the established of artificial characterization of the humanoid head that make an effort on the way to simulate the aforementioned statistics processing.
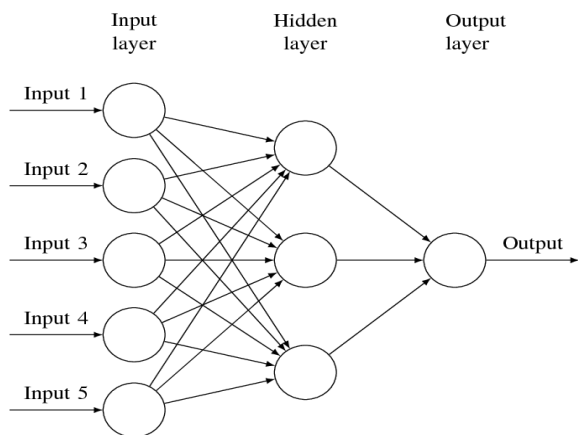
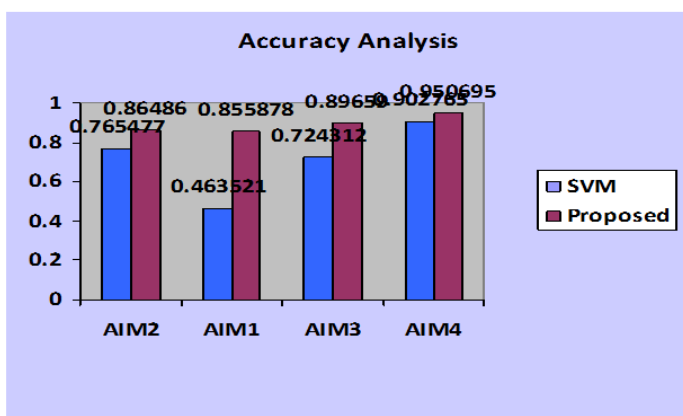**Figure 2:** *Feed Forward Neural Network*

Figure 3: *A neurons in the hidden or output layer*

This process is defined as follows:

$$y = F(w_i x_i - t)$$

Here,
variables (X1, X2, Xi, . . . ,Xn) = inputs variables (W1, W2, . ,Wi, . ,Wn) = weights related to the inputs
t = threshold
F = activation function of the neurons
y = output

4.4 Support Vector Machine

A SVM is a discriminatory category authorized distinct by an extrication hyper-plane. It is a type of deep learning algorithm that performed the classification or regression of supervised learning methods. However, it is mostly used in classification problems. SVM is used for determined volume of high dimensional dataset. But, calculation of SVM is identical expensive in expressions of consumption in time taken and memory space. SVM provided straight classical method used for classification and reversion challenges. This challenge is solving linear and nonlinear problems and generated work performance is very well for any other practical problems. SVM is a standard machine learning technique, which remains magnificently applied on the real lifetime association from many types of fields. It has been very prevalent amongst the machine learning and data mining investigators. The classification of SVM method consists of two stages like: preparation (training) and challenging (testing). The classification of SVM techniques performance was verified using cross-validation of 10-fold. SVM method selected the direct kernel function.

4.5 KNN Classifier

It is unique of the closest work in progress. It is used in data mining and machine learning techniques. It is based on the two things such as: regression and classification. K-NN regression, in this regression produces the significance property for the object. This significance property is the regular value of K-NN (k-nearest neighbor). K-NN classification, in this classification production is an association class for the object. This object is confidential by a plurality division of its neighbors. K is known as positive integer. If k = 1, now entity is allocated towards the particular nearest neighbor class.

Follows KNN's stages are: -
1. Select an uncategorized dataset.
2. Previously classified dataset extent the space.
3. Small distance measurement taken K constraint.
4. Checked gradient classes and consumed the straight space.
5. Calculate the each and every class's total amount.
6. Takings by means of perfect class that give the idea of maximum times.
7. New dataset classified with the class that acquired in stage 6.

5. RESULT EVALUATION

5.1 Performance Evaluation

To performance evaluation of the suggested strategy and used three measurements like- accuracy analyze, specificity analyze, sensitivity analyze. Contemporary investigation of heart infection identification, imprecision is the design of the perceptive conclusion among the patients and strong issue curriculums. In the incident that the consequences of classification do not offer faithful perception surrounded by alternate conditions of fitness, then and there the inaccuracy is not noteworthy although truthful discernment delivers in height accuracy. Probability of the patient classification is known as the sensitivity.

$$Sensitivity = \frac{TP}{TP + FP}$$

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

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

Where, TP is known as the number of true position, FP is known as the number of false position, TN is known as the number of true negative and FP is known as the number of false positive. However, obviously show that three solution metrics such as: accuracy, sensitivity and specificity, are relatively adequate towards exhibit the class of the organization outcomes.

5.2 Result Analysis

Solution metrics is a called as grid matrix that is used to calculate the quality of classifier outcomes. It is concert the measurement techniques intended for ordering of machine learning method. In metrics each and every row represented

2793

to the predicted class and each and every column represented to the actual class. It is also known as error metrics.
Scheming process of error metrics:
1. Select a validated or tested dataset with predictable results.
2. Create an estimated data for tested dataset in each and every row.
3. As of expected results and predictable amount.
Solution Metrics for SVM and Solution Metrics for K-NN is showing below in table 3 and table 4:

Table 3: Solution Metrics for SVM

Solution Metrics of 's2'	
95	2
33	15
0	0
Solution Metrics of 's2'	
70	15
75	12
0	0
Solution Metrics of 's3'	
93	3
48	14
0	0
Solution Metrics of 's4'	
125	5
5	6
0	0

Table 4: Solution Metrics for Proposed Techniques

Solution Metrics of 's2'	
246	21
14	22
0	0
Solution Metrics of 's2'	
224	15
20	31
0	0
Solution Metrics of 's3'	
259	17
14	19
0	0
Solution Metrics of 's4'	
289	8
5	6
0	0

5.3 Accuracy Analysis

It is defining the measurement tool and calculates the discrimination outcomes among the classes of healthy subject and patients. If concerns of the classification don't give the true perception constrained by healthy of substitute states. If accuracy is not important while true discrimination outcomes gives highest inaccuracy. Now, in which every aim of K-NN technique is compared with the every aim of SVM technique. SVM method specified four AIMS such as: AIM1, AIM2, AIM3, and AIM4. Accuracy of AIM1 is 0.463521, accuracy of AIM2 is 0.765477, accuracy of AIM3 is 0.724312, and accuracy of AIM4 is 0.902765. K-NN methods recommended four AIMS such as: AIM1, AIM2, AIM3, and AIM4. Accuracy of AIM1 is 0.86486, accuracy of AIM2 is 0.855878, accuracy of AIM3 is 0.89659, and accuracy of AIM4 is 0.950695.

$$\text{Accuracy Analyze} = \frac{\text{Total accuracy of AIMS}}{\text{No. of AIMS}} \times 100$$

In this consideration represented by the evaluation between

SVM method and K-NN scheme is implemented. In which it is established that the 70% accuracy rate of the SVM then our recommended method is approximately 91% which means our method generated better performance.

Table 5: Accuracy of the SVM to the K-NN method.

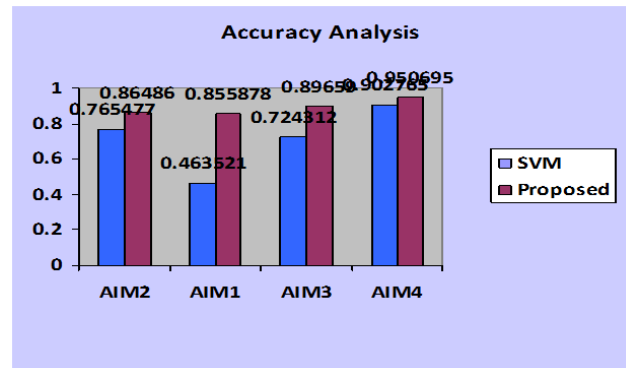|	SVM	Proposed
AIM2	0.765477	0.86486
AIM1	0.463521	0.855878
AIM3	0.724312	0.89659
AIM4	0.902765	0.950695

Figure 4: Accuracy Analysis the graph between SVM and suggested to the K-NN Method

5.4 Specificity Analysis

It is also called as true negative rate and measurement the section of actual negative that is appropriately recognized. It is responsible for the patients of data misclassified by way of standard substance as per.
Now, in which each and every aim of K-NN technique is compared with every aim of SVM technique. SVM method specified four AIMS such as: AIM1, AIM2, AIM3, and AIM4. Specificity of AIM1 is 0.444444, specificity of AIM2 is 0.7777779, specificity of AIM3 is 0.754626, and specificity of AIM4 is 0.33333. K-NN methods recommended four AIMS such as: AIM1, AIM2, AIM3, and AIM4. specificity of AIM1 is 0.6889, specificity of AIM2 is 0.512548, specificity of AIM3 is 0.552541, and specificity of AIM4 is 0.4654.

$$\text{Specificity Analyze} = \frac{\text{Total specificity of AIM}}{\text{No. of AIMS}} \times 100$$

In this consideration represented by the evaluation between SVM method and K-NN method is implemented. In which it is established that the 60% accuracy rate of the SVM then our recommended method is approximately 52% which means our method generated method is given better performance of specificity analysis.

Table 6: Specificity result of the SVM to k-NN

Specificity Analyze	
SVM	Proposed

AIM2	0.7777779	0.512548
AIM1	0.444444	0.6889
AIM3	0.754626	0.552541
AIM4	0.33333	0.4654

Figure 5: Specificity graph between SVM to k-NN
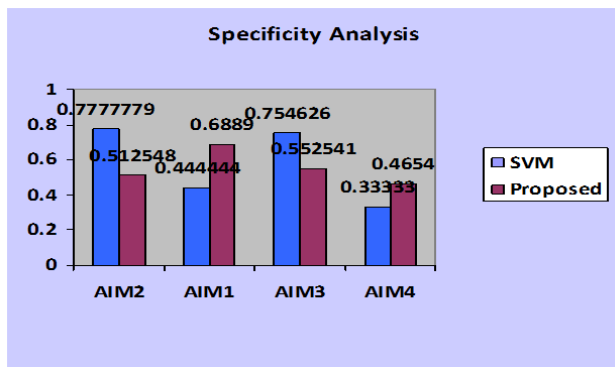

Figure 6: Senstivity graph between SVM to k-NN

5.5 Sensitivity Analysis

It is also known as accurate progressive ratio and measurement the section of actual progressive that is appropriately recognized. Classification of patient probability is known as sensitivity. Now, in which each and every aim of K-NN technique is compared with the every aim of SVM technique. SVM method specified four AIMS such as: AIM1, AIM2, AIM3, and AIM4. Sensitivity of AIM1 is 0.472581, sensitivity of AIM2 is 0.745732, sensitivity of AIM3 is 0.714965, and sensitivity of AIM4 is 0.928761. K-NN methods recommended four AIMS such as: AIM1, AIM2, AIM3, and AIM4. sensitivity of AIM1 is 0.902545, sensitivity of AIM2 is 0.935424, sensitivity of AIM3 is 0.92477, and sensitivity of AIM4 is 0.954214.

$$\text{Sensitivity Analyze} = \frac{\text{Total sensitivity of all AIMS}}{\text{No. of AIMS}} \times 100$$

In this consideration represented by the evaluation between SVM method and K-NN method is implemented. In which it is established that the 70% accuracy rate of the SVM then our recommended method is approximately 94% which means our method generated method is given better performance of sensitivity analyze.

Table 7: Sensitivity result analysis of the SVM Suggested k-NN method

	Sensitivity	
	SVM	Proposed
AIM2	0.745732	0.935424
AIM1	0.472581	0.902545
AIM3	0.714965	0.92477
AIM4	0.928761	0.954214
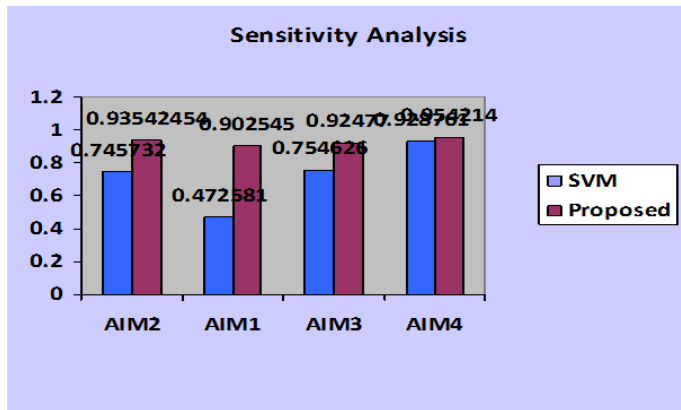
6. CONCLUSION

Our experimental outcomes analyze accuracy, specificity and sensitivity using KNN (K — Nearest Neighbors) classifier and SVM (Support Vector Machine). Now, in which each and every aim of K-NN technique is compared with the each and every aim of SVM technique. In this consideration represented by the evaluation between SVM method and K-NN method is implemented. In which it is established that the 70% accuracy rate of the SVM then our recommended method is approximately 91% which means our method generated method is given better performance of accuracy analysis then the prevailing SVM techniques. In which it is established that the 60% accuracy rate of the SVM then our recommended method is approximately 52% which means our method generated method is given better performance of specificity analyze then the prevailing SVM techniques. In which it is established that the 70% accuracy rate of the SVM then our recommended method is approximately 94% which means our method generated method is given better performance of accuracy analysis then the prevailing SVM techniques.

7. REFERENCES

[1] Sarath Babu, Vivek EM, Famina KP, Fida K, Aswathi P, Shanid M, Hena M: "Heart Disease Diagnosis Using Data Mining Technique", International Conference on Electronics, Communication and Aerospace Technology(ICECA), page no 750-753 2017.

[2] P.K. Anooj: "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules", Journal of King Saud University – Computer and Information Sciences, page no 27–40, 2012.

[3] Kensaku Kawamoto, Caitlin A Houlihan, E Andrew Balas, David F Lobach: "Improving clinical practice using clinical support systems: a systematic review of trials to identify features critical to success", BMJ, doi:10.1136/bmj.38398.500764.8F, page no 1-8, 2005.

[4] Nidhi Bhatla, Kiran Jyoti: "International Journal of Engineering Research & Technology (IJERT), vol. 1 Issue 8, ISSN: 2278-0181, page no 1-5, 2012.

[5] Boshra Bahrami, Mirsaeid Hosseini Shirvani: "Prediction and Diagnosis of Heart Disease by Data

Mining Techniques", Journal of Multidisciplinary Engineering Science and Technology (JMEST), ISSN: 3159-0040, vol. 2 Issue 2, 2015.

[6] UCI, "Heart Disease Dataset", 2017. [Online], Available: https://archive.ics.uci.edu/ml/machinelearning-databases/heart disease/.[Accessed:1-Feb-2019].

[7] Alberto Palacios Pawlovsky: "An Ensemble Based on Distances for a kNN Method for Heart Disease Diagnosis", Faculty of Biomedical Engineering, Dept. of Clinical Engineering, Toin University of Yokohama, Kanagawa, Japan.

[8] František Babič, Jaroslav Olejár: "Predictive and Descriptive Analysis for Heart Disease Diagnosis", Proceedings of the Federated Conference on Computer Science and Information Systems, DOI: 10.15439/2017F219, ISSN 2300-5963 ACSIS, Vol. 11, page no 155-163, 2017.

[9] Balasaheb Tarle, Sudarson Jena: "An Artificial Neural Network Based Pattern Classification Algorithm for Diagnosis of Heart Disease", Research Scholar, Computer Science and Engineering GITAM University Vishakhapatnam, Hyderabad, 978-1-5386-4008-1/17/$31.00 ©2017 IEEE.

[10] Meenal Saini, Niyati Baliyan*, Vineeta Bassi, "Prediction of Heart Disease Severity with Hybrid Data Mining", 2nd International Conference on Telecommunication and Networks (TEL-NET 2017).

[11] Burak Kolukisa*1, Hilal Hacilar1, Gokhan Goy1, Mustafa Kus2, Burcu Bakir-Gungor: "Evaluation of Classification Algorithms, Linear Discriminant Analysis and a New Hybrid Feature Selection Methodology for the Diagnosis of Coronary Artery Disease", IEEE International Conference on Big Data (Big Data), 1Department of Computer Engineering, Abdullah Gül University, Kayseri, Turkey, 978-1-5386-5035-6/18/$31.00 ©2018 IEEE, page no 2232-38, 2018.

[12] K.Polaraju, Durga Prasad: "Prediction of Heart Disease using Multiple Linear Regression Model", IJEDR, Vol. 5, Issue 4, ISSN: 2321-9939, 2017.

[13] K. Rajalakshmi and K. Nirmala, "Heart Disease Prediction with MapReduce by using Weighted Association Classifier and K-Means", Indian Journal of Science and Technology, Vol. 9(19), DOI: 10.17485, ISSN (Print): 0974-6846, ISSN (Online): 0974-5645, May 2016.

[14] Alberto Palacios Pawlovsky: "An Ensemble Based on Distances for a kNN Method for Heart Disease Diagnosis", 2018.