International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8295
***Article Received:** 01 July 2023 **Revised:** 30 August 2023 **Accepted:** 16 September 2023*

A Comprehensive Analysis on Risk Prediction of Heart Disease using Machine Learning Models

Dr. Pokkuluri Kiran Sree[1] , Dr.M. Prasad[2], Dr. Raja Rao PBV[3], Mr Chintha Venkata Ramana[4], Mr. P T Satyanarayana Murty[5], Mr. A. Satya Mallesh[6], Mr. P J R Shalem Raju[7]
1-7, Department of Computer Science And Engineering , Shri Vishnu Engineering College for Women
Bhimavaram,India, 534202
e-mail: drkiransree@gmail.com

Abstract— Most of the deaths worldwide are caused by heart disease and the disease has become a major cause of morbidity for many people. In order to prevent such deaths, the mortality rate can be greatly reduced through regular monitoring and early detection of heart disease. Heart disease diagnosis has grown to be a challenging task in the field of clinically provided data analysis. Predicting heart disease is a highly demanding and challenging task with pure accuracy, but it is easy to figure out using advanced Machine Learning (ML) techniques. A Machine Learning approach has been shown to predict heart disease in this approach. By doing this, the disease can be predicted early and the mortality rate and severity can be reduced. The application of machine learning techniques is advancing significantly in the medical field. Interpreting these analyzes in this methodology, which has been shown to specifically aim to discover important features of heart disease by providing ML algorithms for predicting heart disease, has resulted in improved predictive accuracy. The model is trained using classification algorithms such as Decision Tree (DT), K-Nearest Neighbors (K-NN), Random Forest (RF), Support Vector Machine (SVM). The performance of these four algorithms is quantified in different aspects such as accuracy, precision, recall and specificity. SVM has been shown to provide the best performance in this approach for different algorithms although the accuracy varies in different cases.

Keywords- Heart Disease Prediction, Machine Learning (ML), Random Forest (RF), Decision Tree, Support Vector Machine (SVM) K-Nearest Neighbors (K-NN).

I. INTRODUCTION

One of the body's most significant organs, after the brain, is the heart. Blood circulation throughout the body is the heart's primary function. Heart disease is any condition that has the potential to disrupt the heart's function. There are numerous types of heart diseasesworldwide.

The World Health Organization (WHO) reports that heart attacks and strokes are the leading causes of death worldwide. According to WHO, heart diseases cause many deaths worldwide each year. Cardiovascular disorders were the cause of more than 50% of deaths in the US and other countries. In many regions, it is one of the leading causes of death. It is regarded as the main cause of death in adults.

Heart disease is one of the leading causes of death worldwide in developed nations. Heart failure risk can be increased by a number of circumstances. Medical experts have classified risk factors into two categories: risk factors that cannot be adjusted and risk factors that can be changed. Unchanged risk factors include family history, gender, and age. Modifiable risk factors include obesity, smoking, physical diseases, high blood pressure, and excessive happinessCardiovascular problems early identification can lower the mortality rates, due to lack of information, many are not aware of the earlier factors that lead tocardiovascular diseases.

Healthcare organisations are attempting to diagnose the disease in its early stages. The disease is typically only discovered in the last stages or after death. This occurrence has led healthcare organizations to aim to identify the diseases at an early stages.

Heart diseases can be categorised as either coronary heart disease or cardiovascular disease. The term "cardiovascular disease" refers to a number of conditions that have an impact on the heart, blood vessels, and the body's circulatory and pumping systems. Various diseases, disabilities, and deaths are carried on by cardiovascular diseases. Disease diagnosis is a critical and difficult task in medicine.

A major problem faced by the healthcare industry in today's lifestyle is that it has become very difficult to predict the disease [18] even after a person is ill. Even in the medical industry, records or data are so large that even now the data in the world may be incomplete and inaccurate. In the past it was very difficult to effectively diagnose the disease in the early stages of every patient and thus it may not be possible to treat patients under these conditions. Many researchers have tried to develop a model that can predict heart disease in its early

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8295
***Article Received:** 01 July 2023 **Revised:** 30 August 2023 **Accepted:** 16 September 2023*

stages but have not been able to develop a well-developed model. Every suggested system has disadvantages in its own way. Diagnosis of heat illness is based on examination of patient's health data by invasive based methods and analysis of relevantsymptoms by health professionals.

Since cardiac disease has a complex nature, it requires careful treatment. The impact on the heart is not recognized so gradually that it increases or results in sudden death. Given that the number of individuals dying from heart disease continues to increase at an alarming rate worldwide, the prediction of heart disease is one of the most significant issues in clinical data analysis The healthcare sector has produced a lot of information about heart disease. Can make predictions with an acceptable level of accuracy given the vast amount of data that ML has given the medical services industry. A non-invasive method has been developed in response to conventional methods for early heart disease detection. These conventional methods are followed by medical reports and these methods take a lot of time to complete and also these methods are initiated by human beings and therefore these methods are not accurate for the diagnosis of heart disease and therefore give false results. An automated system is essential to remove these problems and achieve better and faster results. Using different machine learning techniques [17], researchers have recently found that medical data sets may be extensively analysed. Machine learning algorithms will be provided these data sets directly, and they will perform in accordance with everything they were designed

To detect disease at an early stage and assist traditional diagnostic methods to cure the disease they are detected using Machine Learning algorithms. Several Machine Learning methods were used to create the model and address the issue of identifying the key morbidity factor when forecasting cardiovascular diseases. Additionally, these algorithms tackle a variety of issues, such as the detection of absent values, outliers, and significantquantities.

Remaining analysis is arranged as follows:

The methodology for predicting heart disease is described in Section III, while Section II provides an explanation of the earlier analysis, in Section IV the analysis of results is described and in Section V the conclusion is described.

II. LITERATURE SURVEY

Singh, A et al. [1] The study examines various machine learning methods for calculating heart disease. The Decision tree, KNN algorithm, SVM, and linear regression procedure are the classification and regression models that are utilized for prediction in this study. The KNN algorithm was the most accurate, as shown by the experiment's results. However, in real-time environments or applications this model is implemented.

Hashi, E.K. M.S.U and Zaman et al. [2] Cognitive methods were used to predict cardiac disease. In this analysis, the accuracy of five machine learning algorithms used for predictions are examined. Heart disease is predicted using the bagging models, and the logistic model tree is implemented to enhance predictive performance. It is common knowledge that studies using random forest estimations have produced highly accurate results.

A. Javeed, A. Noor, L. Yongjian,

S. Zhou, I. Qasim, and R. Nour et al. [3] By overcoming the issue of overfitting, the development of a model was improved in terms of its ability to predict cardiac disease. When a model predicts cardiac disease with perfect accuracy on training data but performs better than anticipated on test data, this is referred to as overfitting. A model that has been built with good effectiveness on both training and testing data has been created to address these problems. Two algorithms, a random forest algorithm and a random search algorithm, both predict that model. They received better results from both the testing and training data using thedescribed method.

S. Mohan, C. Thirumalai, and G. Srivastava et al. [4] In order to predict coronary disease, they created an artificial intelligence model that combines the advantages of the Linear Method (LM) and Random Forest (RF). In this way, the prediction model achieves the highest level of accuracy. Keeping unnecessary symptoms and signs to a minimum will help patients undergo fewer tests.

L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan et al. [6] An IoT-based system on ensemble that works for diagnostic prediction and patient monitoring produced the best output using random forest at 93%. Additionally, bagging, Random Forest, Naive base, and K-Nearest Neighbour approaches were utilized.

K. Deepika and S. Seema et. al., [8] Predictive analytics such as Naive basis, Support Vector Machine, Random forest, Decision tree, and Artificial neural network were developed to prevent and control chronic cardiac illnesses to measure the accuracy of UCI machine learning repository datasets. They were utilized with the help of Machine Learning techniques. Gives the best accuracy Support vector machine among them.

Soni, Jyoti et al. [11] The authors address the use of data mining tools to forecast cardiac disease. Some methods like Bayesian classification algorithms i.e. KNN algorithm, Decision Tree algorithm, Neural Network classifiers have been studied and evaluated. The use of Genetic Algorithm in feature selection has also been studied for some important risk factors for heart disease. Evaluated the study and highly accurately analyzed the decision tree model.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8295
***Article Received:** 01 July 2023 **Revised:** 30 August 2023 **Accepted:** 16 September 2023*

Vanisree K and JyothiSingaraju et al.[12] To solve many types of problems Machine learning is useful. The necessary variable from the autonomous factor estimations, Estimates can also be made using this procedure. The health care industry is an information mining application area because it contains substantial amounts of physically challenging data resources. Even in developed nations, the major cause of death is considered to be heart disease. Among the most causes of death from heart disease is that the dangers either go unnoticed or are discovered later. While Neural Networks followed in accuracy, selection trees underperformed. SVM was determined to be the most accurate predictor.

LathaParthiban and Subramanian R. et. al.,[15] Produce adequate outcomes for cardiac disease prediction, Neural systems, fuzzy logic, and genetic calculations are all connected in the CoActive Neuro-Fuzzy Inference System (CANFIS) model. As a result, a heritability estimate was utilized to fine- tune the CANFIS parameters and to choose the best set of capabilities. Healthcare providers are urged to use the model as a helpful tool for predicting cardiac disease.

III. RISK PREDICTION OF HEART DISEASE

Figure 1 represents the suggested heart disease risk prediction block diagram. This section presents and discusses in detail Machine Learning techniques for risk analysis of heart disease.

In the field of medical diagnosis, machine learning is used extensively in areas where computer analysis reduces manual error and increases accuracy. The most reliable way to know the diagnosis is to use Machine Learning algorithms. This technology uses Machine Learning to automate medical diagnosis in an effort to forecast heart problems. To learn about heart disease in a different way, ML systems are used to process the raw data and provide different approaches. Three important characteristics are present in inputs classified as objective, examination, and subjective: 1. Information from the patient objective characteristics are collected for factual; 2. 3. Subjective characteristics are obtained from the patient in order to gather information. Test characteristics are obtained following the medical examination. The dataset features used in the model were age, glucose, height, presence of cardiovascular disease, weight, physical activity, sex, alcohol use, systolic blood pressure, smoking, diastolic blood pressure, and cholesterol.

A binary classifier and a multiclass parameter are suggested for the features of a given dataset. The multiclass parameter is used to determine if cardiac disease exists or not. The patient has heart disease if the value is 1. The patient has no heart illness, according to state 0 of the patient's condition. During the pre-processing phase, the detection values are converted from the medical records. The selection of features has received particular consideration. This improves the suggested scheme's integration into current systems and makes the extraction process easier. After being finalised, the dataset was split into two sets: training and testing. The remaining 70% of data was used to train the ML models. After the model has been trained, the final model is tested with the remaining 30% of data. Then, ML implementation was done using these split datasets.

Instead of using regression, supervised Machine Learning techniques utilize classification algorithms [19] since experiment outcomes are true or false values rather than continuous numerical values. Five Machine Learning (ML) classifier techniques were used to train the presented model. These are the algorithms:

(1) Decision Tree (2) K-NN (3) RF (4) SVM). Among the 13 attributes in the data set, two attributes, age and gender, are utilized to recognize personal information about patients. The remaining 11 symptoms are considered important because they have significant clinical records.

Most significantly, medical records are used to identify and diagnose cardiac disease. Several ML methods like DT, K- NN, RF and SVM are used in this approach as mentioned earlier. For the experiment, 13 features in all were revised using ML methods. This classification model is given with a decision tree as a prediction model. Different types of choice trees exist. It maintains attributes (tree branches) for inferences with respect to target values (tree leaves).

Target parameter classification trees and tree models can take finite values. These tree frameworks are represented by the leaves in their class labels and the branches describe combinations of attributes directing those class labels. Regression trees can handle continuous values because fixed trees inthe target parameter (usually realnumbers).

A popular and effective classifier is a growth ratio selection tree. This is one of the most important and least difficult ways to keep things in order when the customer doesn't know much or doesn't understand how the information is spread out. When a test to understand this ordering technique is discovered or when certain pure parametric controls of probability densities are unknown, such figures are prepared to do discriminant estimation. A specified region of K-nearest neighbours is chosen using a preparation dataset. Euclidean partitioning was used to evaluate the data set to determine how close each person was to the production objective.

K-NN's approach is to split the collection to the line being estimated. For unusual columns in the objective set the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8295
***Article Received:** 01 July 2023 **Revised:** 30 August 2023 **Accepted:** 16 September 2023*

present method is re-hashed. A classification algorithm is called k-nearest neighbor algorithm. Based on the class k of a given data point's k closest neighbours, where k is a tiny positive integer, one can determine the class of a given data point. Analyses are presented for Random Forest (RF) categorization. A regulation that uses a decision tree. Many decision trees are detailed that were created by building them throughout training. Is utilized to classify objects, leading to classes mode. To use mechanism Random Forest is shown to be very near and easy. To make a single decision, Random Forest regression combines multiple decisions. Training is carried out using random sampling for features and arbitrary sub-features for sample nodes. Into test set and train the dataset is divided. Contains an attribute subset that is made to ensure that each subset.

The same feature may be considered more than once because the tree building samples are produced using bootstrapping. There may be a numerical limit to the total number of node splitting features. This fitting issue is simplified by this algorithm. A supervised learning simulation is the Support Vector Machine (SVM)[16]. It implicitly transfers the inputs of those spaces into high dimensional feature spaces. This is a method for classifying both linear and nonlinear data that is reported. An SVM utilizes a nonlinear mapping to transform the initial training data. It searches out the hyper-plane that best linearly divides tuples corresponding to various classes. The SVM finds this hyper-plane using support vectors and margins. Several standard performance indicators, including accuracy, precision, and classification error, have been taken into consideration when calculating the performance efficacy of this model.

IV. RESULT ANALYSIS

This section shows the outcome analysis of the offered risk prediction of heart disease using the ML technique. The performance of the suggested model is evaluated using the following definitions for False Positive (FP), True Positive (TP), True Negative (TN), and False Negative (FN):

True Positive (TP): The total probability, or TP, is the sum of all correctly categorised instances of actually positive predictive performance.

True Negative (TN): TN stands for the total number of accurately identified and actually negative predictive instances.

False Positive (FP): The overall number of cases with positive predictions that were incorrectly classified as errors but still had a positive FP is maintained.

False Negative (FN): The total number of cases in which negative predictions were classified as error rather than actual is maintained as a negative FN.

Accuracy: The explanation of it is given as the ratio of correctly detected instances to all occurrences, and it is represented as

$$\text{Accuracy} = \frac{TP+FN}{(TP+FP+TN+FN)} * 100$$

Precision: The ratio of data points reflects the accuracy with which a model claims to be relevant and actually is relevant. Can only provide and express relevant contexts with precision this means that classification models

$$\text{Precision} = \frac{TP}{(TP+FP)} * 100$$

Recall: In the dataset recall reveal ability to find all relevant contexts. In recall and mean classification patterns identify all relevant contexts.

$$Recall = \frac{TP}{TP+FN} * 100$$

The performance analysis of the presented risk prediction of heart disease made with ML models is detailed in table 1.

Table 1: PERFORMANCE ANALYSIS

Performance Metrics	Decision Tree	K-NN	RF	SVM
Accuracy (%)	79	88	81	91
Precision (%)	84	86.4	87.1	90.5
Recall (%)	70.6	76.2	79.7	94.8
Specificity (%)	50.0	56.5	60.0	82.6

The SVM, which is used to detect heart disease, provides high precision, recall, accuracy, and specificity, as represented in the above table. SVM is hence more accurate than Decision Tree, K-NN, and RF. SVM has a higher precision in this comparison, as demonstrated by the graph above. Therefore SVM approach has better Recall than the other approaches. Therefore using Machine Learning to identify the risk prediction of heart disease has potential with high specificity.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8295
***Article Received:** 01 July 2023 **Revised:** 30 August 2023 **Accepted:** 16 September 2023*

V. CONCLUSION

A comprehensive analysis using Machine Learning models was performed in this approach to identify the risk of heart diseases. Reducing human interaction and using clinical laboratory testing to forecast an appropriate Machine Learning model for early disease prediction has improved treatment with a satisfactory level of accuracy. Is to develop proper machine learning models for the prediction of heart disease in the current era due to the unimaginable increase in mortality due to heart disease, it is said that one of the imperatives. To improve accuracy in heart disease prediction this analysis suggests an approach that aims to detect critical features by using Machine Learning approaches. SVM was determined as the best algorithm to predict heart disease based on the analysis results. Based on the calculated performance metrics of the four algorithms in different aspects such as accuracy, precision, recall and specificity, SVM outperformed decision tree, K-NN and RF in this research.

REFERENCES

[1] Singh, A., (2020, February). Heart Disease Prediction Using Machine Learning Algorithms. In 2020 International Conference on Electrical and Electronics Engineering (ICE3) (pp. 452-457). IEEE

[2] Hashi, E.K. and Zaman, M.S.U., 2020. Developing a Hyperparameter Tuning Based Machine Learning Approach of Heart Disease Prediction. Journal of Applied Science & Process Engineering,7(2), pp.631-647.

[3] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R. Nour, "An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection," IEEE Access, vol. 7, pp. 180235– 180243, 2019, doi: 10.1109/ACCESS.2019.2952107.

[4] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," IEEE Access, vol. 7, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

[5] L. Ali "An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure," IEEE Access, vol. 7, pp. 54007–54014, 2019, doi:10.1109/ACCESS.2019.2909969

[6] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan, "An Automated Diagnostic System for Heart Disease Prediction Based on $\chi 2$ Statistical Model and Optimally Configured Deep Neural Network," IEEE Access, vol. 7, pp. 34938–34945, 2019, doi: 10.1109/ACCESS.2019.2904800.

[7] R. Ani, S. Krishna, N. Anju, M. Aslam and O. Deepa, "Iot based patient monitoring and diagnostic prediction tool using ensemble classifier", in 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, 2017, pp. 1588-1593.

[8] K. Deepika and S. Seema, "Predictive analytics to prevent and control chronic diseases," Proc. 2016 2nd Int. Conf. Appl. Theor. Comput. Commun. Technol. iCATccT 2016, no. January 2016, pp. 381–386, 2017, doi: 10.1109/ICATCCT.2016.7912028

[9] Hazra, S. Mandal, A. Gupta, and A. Mukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review," Advances in Computational Sciences and Technology, 2017, 10, 2137-2159.

[10] Ashish Chhabbi, Lakhan Ahuja, Sahil Ahir and Y. K. Sharma, "Heart Disease Prediction Using Data Mining Techniques", IJRAT Special Issue National Conference "NCPC-2016", pp. 104-106, 19 March 2016.

[11] Soni, Jyoti, "Predictive data mining for medical diagnosis: An overview of heart disease prediction." International Journal of Computer Applications 17.8 (2011): 43-48.

[12] Vanisree K, JyothiSingaraju. Decision support system for congenital heart diseasediagnosis based on signs and symptoms using neural networks. Int J Comput ApplApril 2011;19(6). (0975 8887).

[13] Fida Benish, Nazir Muhammad, Naveed Nawazish, Akram Sheeraz. Heart disease classification ensemble optimization using genetic algorithm. IEEE; 2011. p. 19–25.

[14] Y. Roche, Risques médicaux au cabinet dentaire en pratique quotidienne: Identification patients évaluation risques prise encharge: prévention précautions, 2010.

[15] LathaParthiban, Subramanian R. Intelligent heart disease prediction system usingCANFIS and genetic algorithm. Int.

[16] Satish Kumar Satti, Suganya Devi K., Prasad Maddula, N.V.Vishnumurthy Ravipati, Unified approach for detecting traffic signs and potholes on Indian roads, Journal of King Saud University - Computer and Information Sciences, Volume 34, Issue 10, Part B, 2022, Pages9745-9756, ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2021.12.006.

[17] Pokkuluri, K. S., & Nedunuri, S. U. D. (2020). A novel cellular automata classifier for covid-19 prediction. Journal of Health Sciences, 10(1), 34-38.

[18] Sree, P. K., Babu, I. R., & Devi, N. U. (2009). Investigating an Artificial Immune System to strengthen protein structure prediction and protein coding region identification using the Cellular Automata classifier. International journal of bioinformatics research and applications, 5(6), 647-662.

[19] Sree, K., & Babu, R. (2010). Identification of Promoter Region in Genomic DNA Using Cellular Automata Based Text Clustering. International Arab Journal of Information Technology (IAJIT), 7(1).

[20] Sree, P. K. (2008). Exploring a novel approach for providing software security using soft computing systems. International Journal of Security and Its Applications, 2(2), 51-58.

[21] Mangalampalli, S., Sree, P. K., Swain, S. K., & Karri, G. R. (2023). Cloud Computing and Virtualization. Convergence of Cloud with AI for Big Data Analytics: Foundations and Innovation, 13-40.

[22] Pokkuluri, K. S., Nedunuri, S. U. D., & Devi, U. (2022). Crop

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8295
***Article Received:** 01 July 2023 **Revised:** 30 August 2023 **Accepted:** 16 September 2023*

Disease Prediction with Convolution Neural Network (CNN) Augmented With Cellular Automata. INTERNATIONAL ARAB JOURNAL OF INFORMATION TECHNOLOGY, 19(5), 765-773.

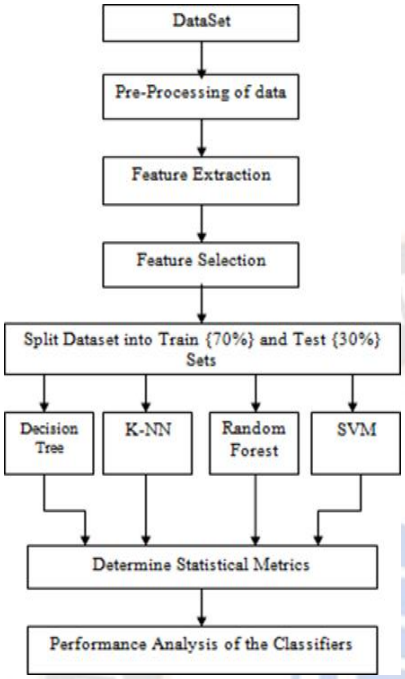

Fig. 1: Presented risk prediction of heart disease

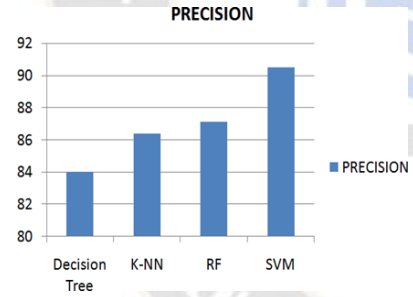

Fig. 2: Accuracy performanceComparison between methods

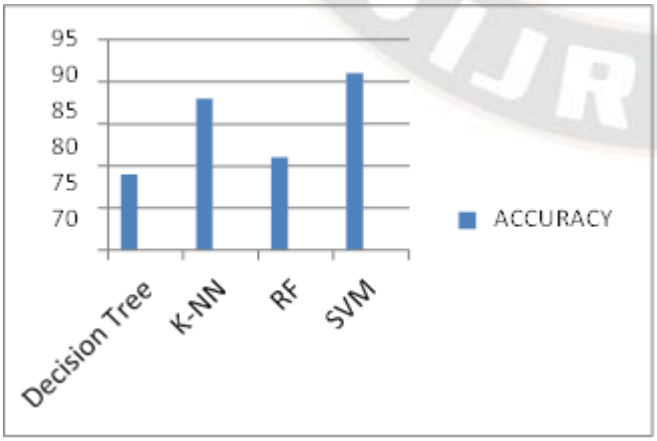

Fig. 3: Precision performance comparisonbetween methods

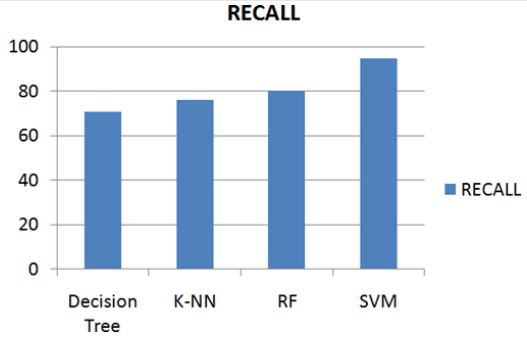

Fig. 4: Recall performance Comparison betweenmethods

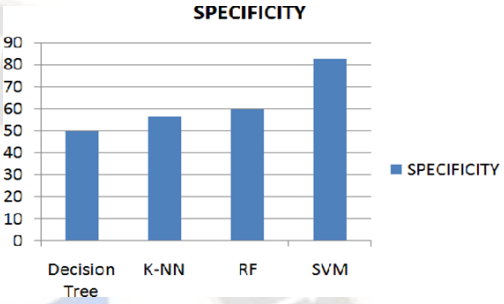

Fig. 5: Specificity performance Comparisonbetween methods