RESEARCH ARTICLE OPEN ACCESS

HEART DISEASE IDENTIFICATION USING MACHINE LEARNING METHODOLOGIES

Ms. N. Zahira Jahan, M.C.A., M.Phil.,[1], Mr. S. Vigneshkumar[2],

[1]Associate Processor, Department in Computer Applications,

Nandha Engineering College, (Autonomous), Tamil Nadu, India

[2]Final M.C.A, Department in Computer Applications,

Nandha Engineering College, (Autonomous), Tamil Nadu, India

[1]zahirajahan1977@gmail.com, [2]vignesh6120@gmail.com

Abstract— In day-to-day life, there are numerous factors which affect a mortal heart. Numerous problems are being at a rapid-fire pace and new heart conditions are fleetly being identified. In moment's world of stress Heart, being an essential organ in a mortal body which pumps blood through the body for the blood rotation is essential and its health is to be conserved for a healthy living. The main provocation of doing this design is to present a heart complaint vaticination model for the vaticination of circumstance of heart complaint. Further, this exploration work is aimed towards relating the stylish bracket algorithm for relating the possibility of heart complaint in a case. The identification of the possibility of heart complaint in a person is complicated task for medical interpreters because it requires times of experience and violent medical tests to be conducted. In this work, three data mining bracket algorithms like KNN bracket, SVM bracket, Naïve Bayes and Random Forest are addressed and used to develop a vaticination system in order to dissect and prognosticate the possibility of heart complaint. The main idea of this significant exploration work is to identify the algorithms suitable for providing maximum accuracy when classification of normal and abnormal person is carried out. Therefore prevention of the loss of lives at an earlier stage is being possible. It is sure that Random Forest algorithm performs better when compared to other algorithms for heart complaint prediction. The design is designed using R Language3.4.4 with R Studio.

Keywords— Data mining, Prediction model Classification algorithms, Feature selection, Heart disease prediction

I. INTRODUCTION

There may also be several inheritable factors through which a type of heart complaint is passed down from generations. According to the World Health Organization, every time more than 12 million deaths are being worldwide due to the colorful types of heart conditions which is also known by the term cardiovascular complaint.

The term Heart complaint includes numerous conditions that are different and specifically affect the heart and the highways of a mortal being. Indeed youthful aged people around their 20-30 times of lifetime are getting affected by heart conditions.

The increase in the possibility of heart complaint among youthful may be due to the bad eating habits, lack of sleep, restless nature, depression and multitudinous other factors similar as rotundity, poor diet, family history, high blood pressure, high blood cholesterol, idle geste, family history, smoking and hypertension. The opinion of the heart conditions is a veritably important and is itself the most complicated task in the medical field.

All the mentioned factors are taken into consideration when assaying and understanding the cases by the croaker through homemade check-ups at regular intervals of time. The symptoms of heart complaint greatly depend upon which of the discomfort felt by an existent. Some symptoms aren't generally linked by the common people. Still, common symptoms include casket pain, breathlessness, and heart pulsations. The casket pain common to numerous types of heart complaint is known as angina, or angina pectoris, and occurs when a part of the heart doesn't admit enough oxygen. Angina may be started by stressful events or physical exertion and typically lasts under 10 twinkles.

Heart attacks can also do as a result of different types of heart complaint. The signs of a heart attack are analogous to angina except that they can do during rest and tend to be more severe. The symptoms of a heart attack can occasionally act indigestion. Heartburn and a stomach pang can do, as well as a heavy feeling in the casket. Other symptoms of a heart attack include pain that travels through the body, for illustration from the casket to the arms, neck, back, tummy, or jaw, flightiness and dizzy sensations, gushing sweating, nausea and vomiting. Heart failure is also an outgrowth of heart complaint, and breathlessness can do when the heart becomes too weak to circulate blood.

Some heart conditions do with no symptoms at each, especially in aged grown-ups and individualities with diabetes. The term' natural heart complaint'covers a range of conditions, but the general symptoms include sweating, high situations of fatigue, fast twinkle and breathing,

breathlessness, casket pain. Still, these symptoms might not develop until a person is aged than 13 times.

In these types of cases, the opinion becomes an intricate task taking great experience and high skill. A threat of a heart attack or the possibility of the heart complaint if linked beforehand, can help the cases take preventives and take nonsupervisory measures. Lately, the healthcare assiduity has been generating huge quantities of data about cases and their complaint opinion reports are being especially taken for the analysis of heart attacks worldwide. When the data about heart complaint is huge, the machine literacy ways can be enforced for the analysis.

Data Mining is a task of rooting the vital decision making information from a philanthropy of once records for unborn analysis or vaticination. The information may be hidden and isn't identifiable without the use of data mining.

The classification is one data mining fashion through which the unborn outgrowth or prognostications can be made grounded on the literal data that's available. The medical data mining made a possible result to integrate the bracket ways and give motorized training on the dataset that further leads to exploring the retired patterns in the medical data sets which is used for the analysis of the case's unborn state. Therefore, by using medical data booby-trapping it's possible to give perceptivity on a case's history and is suitable to give clinical support through the analysis. For clinical analysis of the cases, these patterns are veritably important essential. In simple English, the medical data mining uses bracket algorithms that are a vital part for relating the possibility of heart attack before the circumstance.

The classification algorithms can be trained and tested to make the prognostications that determine the person's nature of being affected by heart complaint. In this exploration work, the supervised machine learning conception is employed for making the prognostications. A relative analysis of the three data mining bracket algorithms videlicet Random Forest, Decision Tree and Naïve Bayes are used to make prognostications. The analysis is done at several situations of cross confirmation and several chance of chance split evaluation styles independently.

The StatLog dataset from UCI machine learning depository is employed for making heart complaint prognostications in this exploration work. The prognostications are made using the bracket model that's erected from the bracket algorithms when the heart complaint dataset is used for training. This final model can be used for analysis of any types of heart conditions.

II. RELATED WORKS

In this paper [1] the authors stated that the prognostic of life for cases with heart failure remains poor. By using data mining styles, the purpose of this study was to estimate the most important criteria for prognosticating patient survival and to outline cases to estimate their survival chances together with the most applicable fashion for health care. Styles Five hundred and thirty three cases who had suffered from cardiac arrest were included in the analysis.

They performed classical statistical analysis and data mining analysis using substantially Bayesian networks. RESULTS The mean age of the 533 cases was 63 [± 17] and the sample was composed of 390 men and 143 women. Cardiac arrest was observed at home for 411 cases, in a public place for 62 [12] cases and on a public trace for60 [11] cases. The belief network of the variables showed that the probability of remaining alive after heart failure is directly associated to five variables age, coitus, the original cardiac meter, the origin of the heart failure and technical reanimation ways employed.

Data booby-trapping styles could help clinicians to prognosticate the survival of cases and also acclimatize their practices consequently. This work could be carried out for each medical procedure or medical problem and it would come possible to make a decision tree fleetly with the data of a service or a croaker. The comparison between classic analysis and data mining analysis showed us the donation of the data mining system for sorting variables and snappily conclude on the significance or the impact of the data and variables on the criterion of the study. The main limit of the system is knowledge accession and the necessity to gather sufficient data to produce a applicable model.

Cardiac arrest is defined as a robotic unrecoverable arrest of the general rotation by cardiac inefficacity. It's honored with the absence of the femoral palpitation for further than 5 seconds. Without reanimation, cardiac arrest leads to unforeseen cardiac death. The public health impact of unforeseen cardiac death is heavy since the survival rate is estimated at between 1 and 20 for cardiac arrest cases. This represents to deaths a time in the United States and to deaths in France. The profile of the case is now well known since it generally concerns men from about 45 to 75 times.

Hospitalization must be optimal and fast. According to the type of cardiac attack, the procedure of supposition of responsibility can vary and some studies [11] show the interest of colorful ways over others, according to the cause of the cardiac arrest. Heart complaint is the number one cause of death in theU.S. Ac-cording to the American Heart Association, an estimated people in theU.S. will have a coronary attack each time.

Ninety five percent of unforeseen cardiac arrest victims die before reaching the sanitarium and heart complaint claims further lives each time than the following six leading causes of death combined (cancer, habitual lower respiratory conditions, accidents, diabetes mellitus, influenza and pneumonia). Nearly people in theU.S. who die from heart complaint each time are under the age of 65. These data show the interest for prognosticating the threat of death after heart failure and the need to dissect the events that passed during care to give prognostic information.

Classic statistical analyses have formerly been done and give some information about epidemiology of the heart failure and causes of the failure. This paper presents the use of a probability in a statistical approach in order to outline heart failure in a sample of cases and prognosticate the impact of some events in the care process.

They concluded that it seems that the use of the Bayesian network in medical analysis could be useful to explore data and to find hid- den connections between events or characteristics of the sample. It's a first approach for agitating suppositions between clinicians and statistical experts. The main limit of these tools is the necessity to have enough data to find chronicity in the connections.

In this paper [2] the authors stated that after a decade of abecedarian interdisciplinary exploration in machine literacy, the spadework in this field has been done; the 1990s should see the wide exploitation of knowledge discovery as an aid to assembling knowledge bases. The contributors to the AAAI Press book Knowledge Discovery in Databases were excited at the implicit benefits of this exploration. The editors hope that some of this excitement will communicate itself to AI Magazine compendiums of this composition.

It has been estimated that the quantum of information in the world doubles every 20 months. The size and number of databases presumably increases indeed briskly. In 1989, the total number of databases in the world was estimated at five million, although utmost of them are small DBASE III databases. The robotization of business conditioning produces an ever- adding sluice of data because indeed simple deals, similar as a telephone call, the use of a credit card, or a medical test, are generally recorded in a computer.

Scientific and government databases are also fleetly growing. The National Aeronautics and Space Administration formerly has much further data than it can dissect. Earth observation satellites, planned for the 1990s, are anticipated to induce one terabyte (1015 bytes) of data every day — further than all former operations combined. At a rate of one picture each alternate, it would take a person several times (working nights and weekends) just to look at the filmland generated in one day. In biology, the federally funded Human Genome design will store thousands of bytes for each of the several billion inheritable bases.

Closer to everyday lives, the 1990U.S. tale data of a million million bytes render patterns that in retired ways describe the cultures and mores of moment's United States. What are we supposed to do with this deluge of raw data? Easily, little of it'll ever be seen by mortal eyes.

Still, it'll have to be anatomized by computers, If it'll be understood at all. Although simple statistical ways for data analysis were developed long agone, advanced ways for intelligent data analysis aren't yet mature.

As a result, there's a growing gap between data generation and data understanding. At the same time, there's a growing consummation and anticipation that data, intelligently anatomized and presented, will be a precious resource to be used for a competitive advantage. The computer wisdom community is responding to both the scientific and practical challenges presented by the need to find the knowledge adrift in the deluge of data.

In assessing the eventuality of AI technologies, Michie (1990), a leading European expert on machine literacy, prognosticated that " the coming area that's going to explode is the use of machine literacy tools as a element of large-scale data analysis." A recent National Science Foundation factory on the future of database exploration ranked data mining among the most promising exploration motifs for the 1990s.

Some exploration styles are formerly well enough developed to have been made part of commercially available software. Several expert system shells use variations of ID3 for converting rules from exemplifications. Other systems use inductive, neural net, or inheritable literacy approaches to discover patterns in particular computer databases. Numerous forward- looking companies are using these and other tools to dissect their databases for intriguing and useful patterns.

American Airlines searches its frequent leaflet database to find its better guests, targeting them for specific marketing elevations. Farm Journal analyzes its subscriber database and uses advanced printing technology to custom-figure hundreds of editions acclimatized to particular groups. Several banks, using patterns discovered in loan and credit histories, have deduced better loan blessing and ruin vaticination styles. General Motors is using a database of machine trouble reports to decide individual expert systems for colorful models. Packaged- goods manufacturers are searching the supermarket scanner data to measure the goods of their elevations and to look for shopping patterns.

A combination of business and exploration interests has produced adding demands for, as well as increased exertion to give, tools and ways for discovery in databases. This book is the first to bring together leading-edge exploration from around the world on this content. It spans numerous different approaches to discovery, including inductive literacy, Bayesian statistics, semantic query optimization, knowledge accession for expert systems, information proposition, and fuzzy sets.

The book is aimed at those interested or involved in computer wisdom and the operation of data, to both inform and inspire farther exploration and operations. It'll be of particular interest to professionals working with databases and operation information systems and to those applying machine literacy to real- world problems.

What Is Knowledge Discovery? There's an immense diversity of current exploration on knowledge discovery in databases. To give a point of reference for this exploration, we begin then by defining and explaining applicable terms.

Description of Knowledge Discovery

Knowledge discovery is the nontrivial birth of implicit, preliminarily unknown, and potentially useful information from data. Given a set of data (data) F, a language L, and some measure of certainty C, we define a pattern as a statement S in L that describes connections among a subset FS of F with a certainty c, similar that S is simpler (in some sense) than the recitation of all data in FS.

A pattern that's intriguing (according to a stoner-assessed interest measure) and certain enough (again according to the stoner's criteria) is called knowledge. The affair of a program that monitors the set of data in a database and produces patterns in this sense is discovered knowledge. These delineations about the language, the certainty, and the simplicity and interestingness measures are designedly vague to cover a wide variety of approaches. Inclusively, these terms capture our view of the abecedarian characteristics of discovery in databases. In the following paragraphs, we epitomize the connotations of these terms and suggest their applicability to the problem of knowledge discovery in databases.

They noted that one of the foremost discovery processes was encountered by Jonathan Swift's Gulliver in his visit to the Academy of Labado. The " Design for perfecting academic Knowledge by practical and mechanical operations" was generating sequences of words by arbitrary permutations and " where they plant three or four Words that might make Part of a Judgment, they mandated them to... Scribes." This process, although promising to produce numerous intriguing rulings in the (veritably) long run, is rather hamstrung and was lately proved to be NP-hard.

III. METHODOLOGY

A. BIG DATA ANALYTICAL MODEL

Big data analytics support the conception of artificial intelligence at the heart of numerous new digital health platforms and perfection health tools. Immaculately, application of big data logical tools in cardiovascular care will restate into better care and issues at a lower cost.

The eventuality for more important prophetic models is a charming operation of big data analytics. Historically, prediction models have reckoned on a limited number of specified variables manually entered to estimate a' threat score'. Similar models generally warrant perfection they perform' nicely well at the population position, but not at the individual patient position.

And despite the actuality of dozens of threat models related to cardiovascular conditions, many are employed to make remedial opinions.

Big data analytics may yield more important analysis of issues ranging from mortality to case- reported issues to resource application, and therefore could be more clinically practicable. Machine literacy, for illustration, evaluates patterns associated with an outgrowth directly from the data, rather than from a pre-specified set of variables.

A full range of associations and relations among the data are assessed. Whereas traditional statistical models are done, machine literacy uses a training process whereby the model is iteratively given varied data sets to explore numerous combinations of prophetic features to optimize analysis.

Phenol-mapping, or deep phenol-typing, is another promising big data operation. Current complaint groups, or phenotypes, are squishy and miscellaneous Big data analytics can identify analogous case clusters, creating multiple phenotypes within each complaint reality. In proposition, further refined phenol-mapping of complaint countries and circles should help inform more customized-health opinions

Big data can support the combination of multiple data sources from large case populations to more estimate the implicit benefits of curatives similar as ICD's for individual cases. Indeed, big data logical styles are central to the success of perfection health, given the growing interest in incorporating data, which extensively increases the size and complexity of datasets. Similar datasets bear advanced logical platforms and styles that are the emblems of big data analytics.

Big data analysis can guide programs to address a certain case member by specific interventions. The success of the policy is critically dependent on the quality of the underpinning exploration and the quality (effectiveness) of the interventions. For numerous interventions (for case in the social/ internal health sphere) widely accepted styles for validating success are still lacking. There's several challenges result regarding Big Data and population heart complaint similar as

• Data protection regulation makes it delicate to dissect data from different heart complaint providers and services in combination

• A significant part of the population health records is unshaped heart complaint textbook

• There are interoperability, data quality and data integration limitations

.

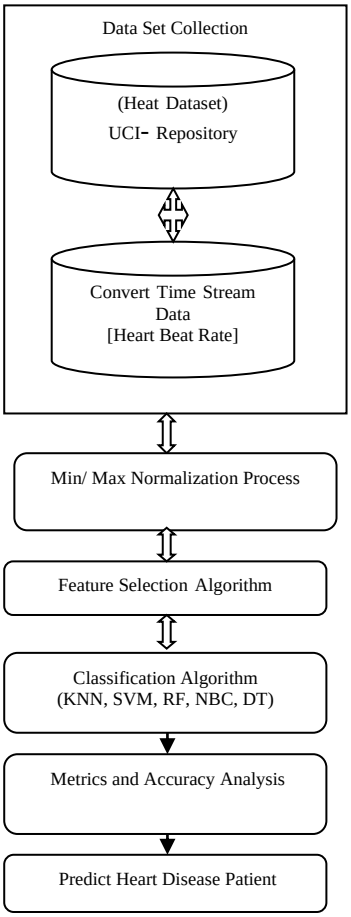

Fig 3.1 Architecture Diagram of Proposed System

B. NORMALIZATION MODEL

Data transformation such as Normalization is a data preprocessing tool used in data mining system. An attribute of a dataset is normalized by scaling its values so that they fall within a small-specified range, such as 0.0 to 1.0. Normalization is particularly useful for classification algorithms involving neural networks, or distance measurements such as nearest neighbor classification and clustering. There are many methods for data normalization includes min-max normalization, z-score normalization and normalization by decimal scaling.

Min-max normalization performs a linear transformation on the original heart dataset. Min-max normalization maps a value d of P to d′ in the range [new_min(p), new_max(p)].

Min max normalization preserves the relationship among the original heart dataset values**.** The table 4.1 Describe a sample normalized heat disease dataset model details shows,

Attribute	**Original Values**	**Normalized dataset**
Age	70.0	35.0
chest pain type	1.0	0.0
resting blood pressure	130.0	140.0
maximum heart rate achieved	109.0	79.0
exercise induced angina	0.0	1.0

Table 3.1 Normalized Heart Dataset

C. GREEDY FEATURE EXTRACTION MODEL

Feature selection is one of the dimension reduction ways which have been used to allow a better understanding of data and ameliorate the performance of other literacy tasks. Although the selection of applicable features has been considerably studied in supervised literacy, point selection with the absence of class markers is still a grueling task.

This paper develops a new system for unsupervised point selection, which efficiently selects features in a greedy manner. The paper first defines an effective criterion for unsupervised point selection which measures the reconstruction error of the data matrix grounded on the named subset of features. The paper also presents a new algorithm for greedily minimizing the reconstruction error grounded on the features named so far. The greedy algorithm is grounded on an effective recursive formula for calculating the reconstruction error.

The greedy algorithm selects replication the most representative point among the remaining features, and also eliminates the effect of the named features from the data matrix. This step makes it less likely for the algorithm to elect features that are analogous to preliminarily named features, which consequently reduces the redundancy between the named features. In addition, the use of the recursive criterion makes the algorithm computationally doable and memory effective compared to the state of the art styles for unsupervised (forward and backward) point selection.

D. CLASSIFICATION ANALYTICAL MODEL

Machine Literacy indicates how computers can learn or ameliorate their performance using data. System programs to generally plant to descry the styles and prove specialized predicated opinions on data. Machine knowledge is a fast growing discipline. Then, using classic problems in machine literacy that are largely affiliated to data mining.

• Supervised bracket literacy

Supervised bracket literacy model correspond of all data is labeled and algorithm learn to prognosticate the affair from training dataset.

E.g. SVM, Random Forest, Naive Bayes.

• Unsupervised bracket learning

Unsupervised bracket literacy is used for clustering grounded algorithm. In this session all the information is mot included and algorithm founds to essential structure from the input dataset.

E.g. K- means, KNN. Neural Networks

•Semi-supervised bracket Learning

Semi-supervised literacy is a combination of supervised literacy and unsupervised literacy. InSemi-supervised literacy some data is labeled and some data isn't labeled. In this approach, labeled training dataset are used to learn class models and unlabelled training dataset are used to define the boundaries between classes.

IV. RESULTS AND DISCUSSIONS

A. DATASET DESCRIPTION

Preparing the database-for carrying the result, this paper used Heart patient data sets from ILPD (Indian Heart Case) Data Set (table 4.1). Completely, heart dataset has 583 samples which holds 10 independent variables and one dependent variable. Independent Variables are Age, Gender, Total Bilirubin, Direct Bilirubin, Total Proteins, Albumin, SGPT (serum glutamic-pyruvic transaminase), SGOT (serum glutamic oxaloacetic transaminase), Alkaline Phosphatase and one dependent variable is Placing (class).

Attributes Type	Description	Gender Categorical
Age	age given in years	Real number
Sex	sex (Value 1 : male; Value 0 : female)	String
Cp	chest pain type(1: typical angina ; 2: atypical angina 3: non-anginal pain ;4: asymptomatic)	Real number
Trestbps	resting blood pressure (in mm Hg on admission to the hospital)	Real number
Chol	Cholestoral(Serum cholestoral) in mg/dl	Real number
Fbs	Fasting blood sugar in mg/dl (>120) Value 1 = true; Value 0 = false)	Real number
Restecg	Resting electrocardiographic results	Real number
Thalach	Heart rate achieved at maximum	Integer
Exang	Exercise induced angina (Value 1 : yes; Value 0 : no)	Integer
Oldpeak	ST depression originated by exercise relative to rest	Integer
Slope	Slope of the peak exercise ST segment (Value 1: upsloping ; Value 2: flat ; Value 3: downsloping)	
Ca	Major vessels (0-3) colored by flouroscopy	Integer
Thal	Result of thalium stress test (Value 3 = normal; Value 6 = fixed defect; Value 7 = reversable defect)	Integer
Num	status of heart disease (angiographic status) Value 0: < 50% diameter narrowing Value 1: > 50% diameter narrowing	Binary

Table 4.1 Dataset Attribute

Table 4.1 describes the attribute type, description and Gender Categorical values.

B. PERFORMANCES METRICS ANALYSIS

Here, the evaluation criteria for Heart complaint analysis model are represented. The table contains Mean Absolute error, Root Relative square Error, Root Relative Forecourt Error and Delicacy values of SVM, KNN, RF, DT and MLP bracket algorithm are considered to dissect the performance of the over mentioned.

METRICS	SVM	KNN	RF	DT	MLP
MAE	0.245	0.223	0. 237	0.217	0.322
RMSE	0.352	0.2 86	0.332	0.282	0.393
RAE	0.423	0.3 82	0.402	0.285	0.492
RRSE	0.378	0.406	0.398	0.302	0.418
Accuracy	0.891	0.908	0.894	0.952	0.926

Table 4.2 Performances Analysis

Table 4.2 describes an evaluation metrics for Heart disease prediction model. The table contains Mean Absolute error and Root Relative square Error of SVM, KNN, RF, DT and MLP classification algorithm details are shown

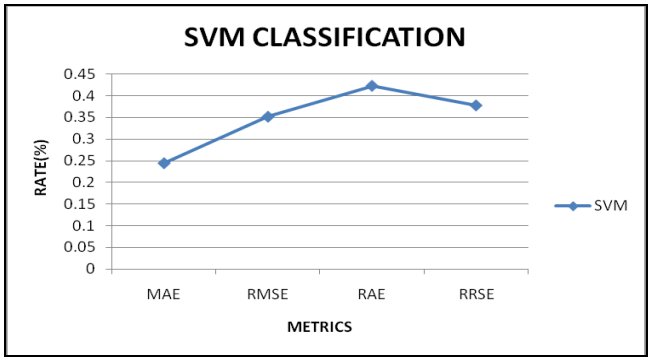

Fig 4.1 Performance of SVM Classification

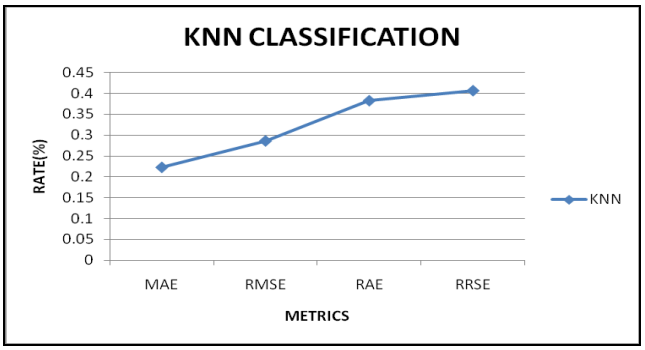

Fig 4.2 Performance of KNN Classification

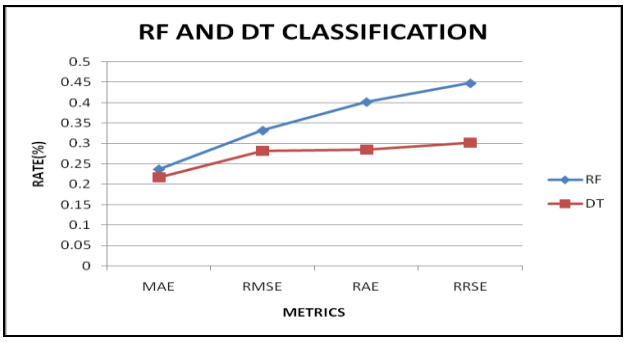

Fig 4.3 Performance of RF and DT Classification

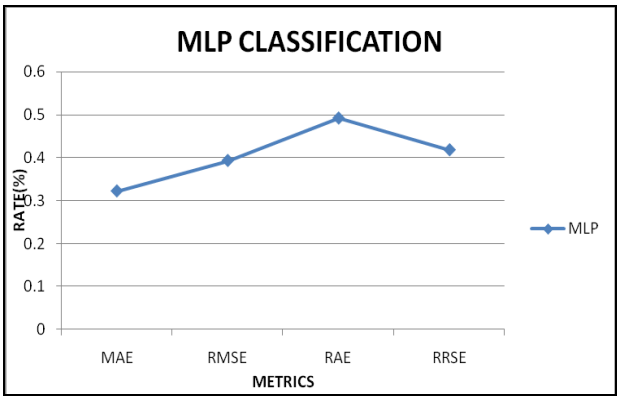

Fig 4.4 Performance of MLP Classification

Fig 4.1, Fig 4.2, Fig 4.3, Fig 4.4 describes the performance analysis of SVM, KNN, RF, DT and MLP classification algorithms respectively.

V. CONCLUSION

The proposed technique is producing an enhanced conception over the heart complaint prediction within new data mining ways; SVM, RF, NB, MLP and DT the weighted association classifier. DT bracket is the fashion to cluster the attributes from the case record. The SVM clustering and DT with weighted association classifier can enhance the bracket performance and delicacy of the heart complaint opinion. These all criteria is being vindicated by the experts and professional croakers of heart specialist. The proposed fashion works better than being system. In unborn work, we will look for the further advancements to get better result over the heart complaint prediction by adding the criteria and croakers suggestion within different types of medical term and also give the first-aid suggestion in attainability of heart specialist or experts.

REFERENCES

[1] Franck Le Duff, Cristian Munteanb, Marc Cuggiaa and Philippe Mabob, “Predicting Survival Causes After Out of Hospital Cardiac Arrest using Data Mining Method”, Studies in Health Technology and Informatics, Vol. 107, No. 2, pp. 1256-1259, 2004.

[2] W.J. Frawley and G. Piatetsky-Shapiro, “Knowledge Discovery in Databases: An Overview”, AI Magazine, Vol. 13, No. 3, pp. 57-70, 1996.

[3] Kiyong Noh, HeonGyu Lee, Ho-Sun Shon, Bum Ju Lee and Keun Ho Ryu, “Associative Classification Approach for Diagnosing Cardiovascular Disease”, Intelligent Computing in Signal Processing and Pattern Recognition, Vol. 345, pp. 721-727, 2006.

[4] Latha Parthiban and R. Subramanian, “Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm”, International Journal of Biological, Biomedical and Medical Sciences, Vol. 3, No. 3, pp. 1-8, 2008.

[5] Sellappan Palaniappan and Rafiah Awang, “Intelligent Heart Disease Prediction System using Data Mining Techniques”, International Journal of Computer Science and Network Security, Vol. 8, No. 8, pp. 1-6, 2008

[6] Shantakumar B. Patil and Y.S. Kumaraswamy, “Intelligent and Effective Heart Attack Prediction System using Data Mining and Artificial Neural Network”, European Journal of Scientific Research, Vol. 31, No. 4, pp. 642-656, 2009.

[7] Nidhi Singh and Divakar Singh, "Performance Evaluation of K-Means and Hierarchal Clustering in Terms of Accuracy and Running Time", Ph.D Dissertation, Department of Computer Science and Engineering, Barkatullah University Institute of Technology, 2012.

[8] Weiguo, F., Wallace, L., Rich, S., Zhongju, Z.: "Tapping the Power of Text Mining", Communication of the ACM. 49(9), 77-82, 2006.

[9] Jiawei Han M. K, Data Mining Concepts and Techniques, Morgan Kaufmann Publishers, An Imprint of Elsevier, 2006

[10] Huang Z, "Extensions to the k-means algorithm for clustering large data sets with categorical values," Data Mining and Knowledge Discovery, Vol.2, pp:283–304, 1998

[11] A comparison of antiarrhythmic-drug therapy with implant-able defibrillators in patients resuscitated from near-fatalventricular arrhythmias. The Antiarrhythmics versusImplantable Defibrillators (AVID) Investigators. N Engl JMed. 1997 Nov 27;337(22):1576-83.

[12] R.Wu, W.Peters, M.W.Morgan, "The Next Generation Clinical Decision Support: Linking Evidence to Best Practice", Journal of Healthcare Information Management. 16(4), pp. 50-55, 2002.

[13] Mary K.Obenshain, "Application of Data Mining Techniques to Healthcare Data", Infection Control and Hospital Epidemiology, vol. 25, no.8, pp. 690–695, Aug. 2004.

[14] G.Camps-Valls, L.Gomez-Chova, J.Calpe-Maravilla, J.D.MartinGuerrero, E.Soria-Olivas, L.Alonso-Chorda, J.Moreno, "Robust support vector method for hyperspectral data classification and knowledge discovery." Trans. Geosci. Rem. Sens. vol.42, no.7, pp.1530–1542, July.2004.

[15] Obenshain, M.K: "Application of Data Mining Techniques to Healthcare Data", Infection Control and Hospital Epidemiology, 25(8), 690–695, 2004.

[16] Wu, R., Peters, W., Morgan, M.W.: "The Next Generation Clinical Decision Support: Linking Evidence to Best Practice", Journal Healthcare Information Management. 16(4), 50-55, 2002.

[17] Charly, K.: "Data Mining for the Enterprise", 31st Annual Hawaii Int. Conf. on System Sciences, IEEE Computer, 7, 295-304, 1998.

[18] Blake, C.L., Mertz, C.J.: "UCI Machine Learning Databases", http://mlearn.ics.uci.edu/databases/heart-disease/, 2004.

[19] Mohd, H., Mohamed, S. H. S.: "Acceptance Model of Electronic Medical Record", Journal of Advancing Information and Management Studies. 2(1), 75-92, 2005.