ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

A Survey on Performance Analysis of Clinical Decision Support System

Vikram V. Kale, Dr. B. R. Bombade

M.Tech Student, Dept. of CSE, SGGSIE&T, Nanded, Maharashtra, India

Associate Professor, Dept. of CSE, SGGSIE&T, Nanded, Maharashtra, India

ABSTRACT: Accuracy plays a very important role in the field of medical disease diagnosis as it directly concern with the life of an individual. Disease improvement is one of the application where data mining techniques are showing effective outcomes. There are extensive research has been supervised on disease prediction and diagnosis using data mining techniques. However there is no argument on which classifier generates best result. A specific classifier generates better result on specific dataset than the other classifiers, but another classifiers could perform better on other datasets. This survey paper covers the performance analysis of various data mining techniques such as Data pre-processing includes Feature selection, Feature extraction, Normalization, etc and Data classification techniques used in Clinical Decision Support System to perform better prediction and diagnosis of various diseases. The comparative performance of data pre-processing techniques and performance of various classifiers on different medical datasets has been covered and shown in this survey paper.

KEYWORDS: Clinical Decision Support System, Data mining, Data pre-processing, Feature selection and feature extraction, Classification, Performance analysis.

I. INTRODUCTION

Clinical Decision Support System(CDSS)is a computerized expert system that is used to analysing and interpreting the Electronic Health Records (EHR) to assists the physicians, doctors and all health care providers to make an effective and correct clinical decisions. CDSS is also useful to prepare a diagnosis and to review a diagnosis as a mean of improving the final result. There are main two types of CDSS. First, CDSS which uses a knowledge based, applies rules to patient dataset using inference engine and display result to the end user. Second, CDSS which is without knowledge based depends on machine learning to analyse clinical data. DXplain and Iliad are the most evolving Diagnostic decision support system studied in this survey [4][1]. Data mining techniques is used to take correct medical decision in CDSS. Data mining includes data pre-processing techniques used to cleaning the medical dataset because quality of dataset is much more important to get exact classification results [17]. Feature selection method contains the attribute importance mining function. Attribute importance is a supervised function that ranks attributes according to their significance in predicting a target class. Feature extraction is an attribute reduction process and feature extraction actually transforms the attributes into smaller dimensions. The transformed attributes or features are linear combinations of the original attributes. Classification technique in data mining having goal to accurately predict the target class for each case in the data. Classification techniques contains various kinds of classifiers used that assigns item in a collection to the target class. In this study used three datasets named as PIMA Indian Diabetes Dataset, Cleveland Heart Disease Dataset and Hepatitis Disease Dataset. All of these datasets taken from UCI Machine Learning Repository datasets which is a collection of datasets, domain theories and data generator that are used for analysis of machine learning algorithm. First, PIMA Indian Diabetes dataset contains 768 instances and 9 features including 1 class attribute denoting that presence of Diabetes i.e '1' and absence of diabetes i.e. '0' [12]. Second, Cleveland Heart disease dataset taken from UCI Machine Learning Repository contains total 303 instances and 14 attributes [10]. The class attribute is divided into five classes, 0 corresponding to absence of heart disease and 1,2,3,4 corresponding to four different kinds of heart diseases. Third, Hepatitis Disease Dataset contains 155 instances and 20

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

attributes including class attribute [11]. Class attribute is nominal and contains values 'LIVE' and 'DIE'. 'LIVE' corresponding to presence of hepatitis disease and 'DIE' corresponding to absence of disease.

II. RELATED WORK

As advances in computational technology a number of research attempts have been directed in involving the machine learning algorithms to the design and develop a clinical decision support system for healthcare industry, especially in disease prediction and diagnosis. Most of the disease detection performs on Heart disease, Cancer disease, Diabetes disease because these are major chronic diseases directly impact on economic condition of nation.

A clinical expert system has been proposed in [3] which adopts decision tree approach for prediction of presence and absence of diabetes among the instances. The author used PIMA Indians Diabetes Data Set, which collects the information of patients with and without developing diabetes. Their research went through two phases. The first phase is data pre-processing including attribute identification and selection, handling missing values, and numerical discretization. The second phase is a diabetes prediction model construction using the decision tree method and they used Weka software throughout all the phases of their study.

Support Vector Machine (SVM), Naive Bayes classifier and Decision Tree are effective classifiers that holds impressing accuracy and efficiency in clinical expert system in terms of prediction of diseases [21]. In this study they reviewed the benefits of different pre-processing techniques such as Principal Component Analysis and Discretization on PIMA Indian Diabetes Dataset for decision support systems for predicting diabetes. Authors found out the accuracy variation which is comparison of these three classifiers using with or without data pre-processing techniques.

In [15] proposed the clinical decision support system using Support Vector Machine (SVM) and Artificial Neural Network (ANN) on Cleveland Heart Disease Dataset. In this paper for diagnosis heart disease the researcher occupied multilayered perceptron neural network(MLPNN). SVM classifiers used to classifies heart disease into two classes whether presence of heart disease or absence of heart disease with 80.41% of accuracy.

Authors in [23], used Logistic regression technique on Hepatitis disease dataset. Principal Component Analysis method used as a data pre-processing and then logistic regression is apply that obtained the accuracy of 89.60%.

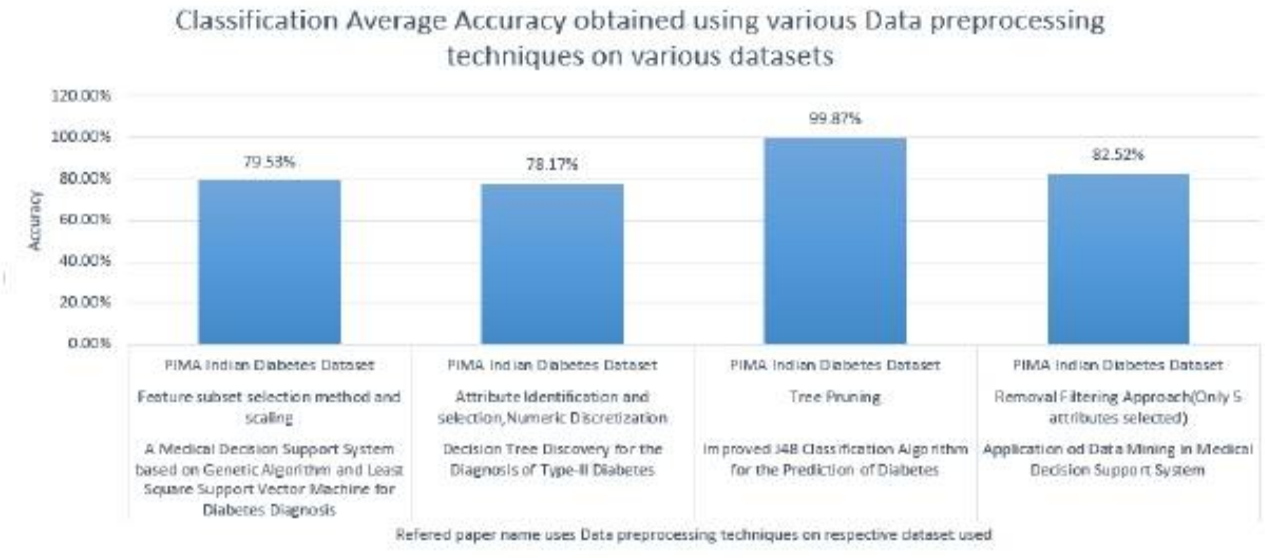

Figure 1. Feature selection accuracy using Naive Bayes classifier on different datasets

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

III. DATA PRE-PROCESSING

Data Preprocessing is the process of preparing the data in a particular way into a format that is suitable for actually applying the data mining algorithms. For the success of data mining and/or machine learning task the representation and quality of instance data is very important and foremost. In case of healthcare industry they collect huge amount of Electronic Healthcare data which is unfortunately are inferior and substandard that affects the effective decision making task and it is very difficult to discover hidden patterns. Data preprocessing techniques used in this situation. Data preprocessing technique follows eliminating redundant entries, removes outliers, missing values, missing entries in dataset, etc. It also follows feature selection and feature extraction methods which used give best classification results.

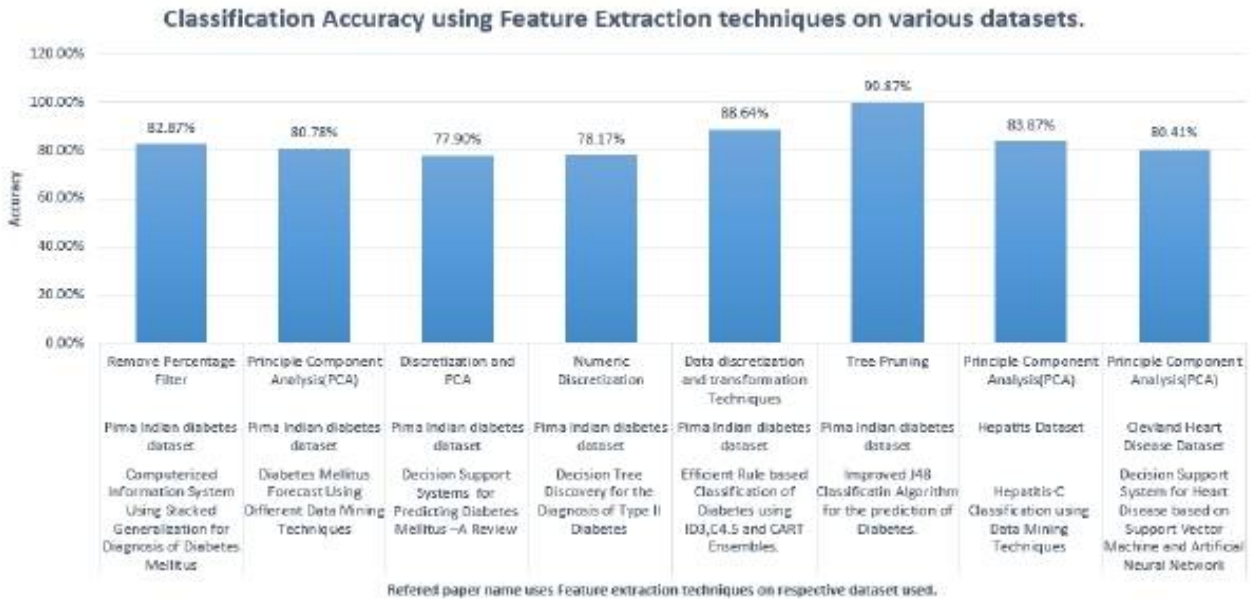

Figure 2. Feature extraction accuracy using various classifier on different datasets

Veena Vijayan V. et.al, used initial preprocessing techniques [20] in their research on PIMA Indian diabetes dataset. Mrudula Gudadhe et.al, used Cleveland Heart Disease dataset contains number of patients instance is 303among them 6 tuples having missing values so they ignore these tuples and then split the dataset into two parts for trainingand testing[15].Rakesh Motka et.al, usedPrincipal Component Analysis method as feature extraction on PIMA Indian Diabetes dataset. The method creates new set of variables from original variables called principals and all principal components are orthogonal to each other and there is no redundant information remains into the dataset [14]. Anjali C. et.al, used initial data pre-processing technique on PIMA Indian Diabetes dataset as well as their research follows Discretization and Principal Component Analysis techniques [21]. In Discretization technique the huge number of attributes are divided into small number of intervals and then labels are given to those attributes. This method mainly applies to attributes that are used in classification and association analysis to reduce the complexity. Aishwarya S. et.al, used the feature subset selection method and scaling which is most important data mining method because eliminating less important feature gives better classification result [16]. For their research PIMA Indian Diabetes dataset is used that contains the attributes with different range of values that directly could affect system stability. Min-max normalization method is used to avoid the complications in computation the numeric values are linearly transformed to fixed ranged. The dataset contains values of features that are normalized between ranges 0 to 1 by using the scaling formula [16]. Asma A. AlJarullah used the attribute identification and selection, handling missing values and numerical discretization techniques on PIMA Indian Diabetes dataset in her research for data pre-processing [3].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

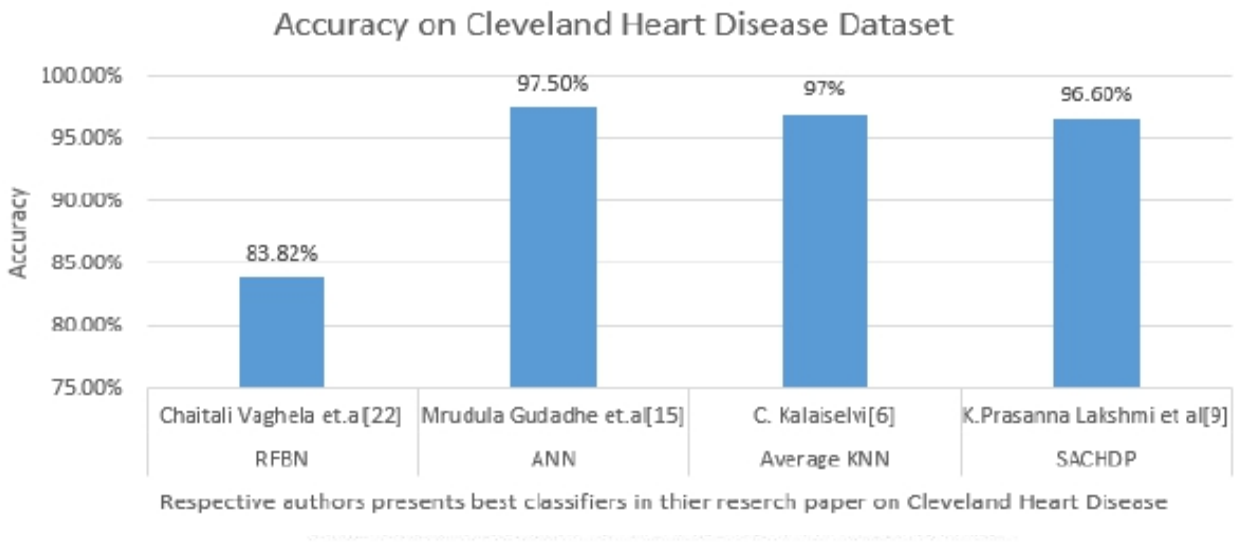

Figure 3. Accuracy on PIMA Indian Diabetes Disease Dataset

Gaganjot Kaur et.al, used additional feature of J48 algorithm such as accounting for missing values, decision tree pruning,continuous attribute values range, derivation of rules, etc. as data pre-processing techniques. The paper focuses on finding entropy and Information Gain for attribute selection from the dataset. The paper also follows tree pruning that helps to decrease classification errors [8]. Habib Shariff Mahmud et.al, used attribute removal filtering approach in his research on PIMA Indian Diabetes dataset [13]. In machine learning task feature selection and feature extraction considered to be most important step before performing classification, Figure (1) and Figure (2) shows the comparison of various data pre-processing techniques used in different survey papers on various datasets.

IV. CLASSIFICATION

The classification is the process of predicting certain outcome based on given input and classification is a data mining function that assigns items in a collection to target classes. The goal of classification is to accurately predict the target class for each case in the data. In this research study includes brief study of various classification techniques apply on datasets used in referred survey papers and comparison of this various papers in terms of accuracy is shown for each datasets used.

Chetali Vaghela et.al, proposed survey on various classification techniques used for clinical decision support system [22]. In their research they used Bayesian Belief Network, Neural Network, Decision Tree, Naive Bayes, Support Vector Machine, Fuzzy Based Approach, Genetic Algorithm, Rough set Approach, K-Nearest Neighbor classifiers on PIMA Indian Diabetes Dataset and Cleveland Heart Disease Dataset. On comparing the accuracy result of all respective classifiers they found out on Cleveland Heart Disease Dataset the Bayesian Network(83.49%) and Naive Bayes(83.49%) gives highest accuracy as compare to MLP(80.85%), RBNF(83.82%), Decision Tree(77.55%) and on PIMA Indian Diabetes Dataset Naive Bayes(76.30%) gives highest accuracy as compare to Bayesian Network(74.34%), Decision Tree(73.83%),MLP(75.39%) and RBNF(75.39%). The comparison of best classifiers used in paper is shown in figure (3).

Mrudula Gudhane et.al, proposed the clinical decision support system using Support Vector Machine(SVM) and Artificial Neural Network(ANN) on Cleveland Heart Disease Dataset[15]. In this paper for diagnosis heart disease the researcher occupied multilayered perceptron neural network(MLPNN). SVM classifiers used to classifies heart disease into two classes whether presence of heart disease or absence of heart disease with 80.41% of accuracy and ANN classifiers classifies the heart disease dataset into five classes namely H0,H1,H2,H3,H4 with 97.5% of accuracy as shown in figure (4).

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

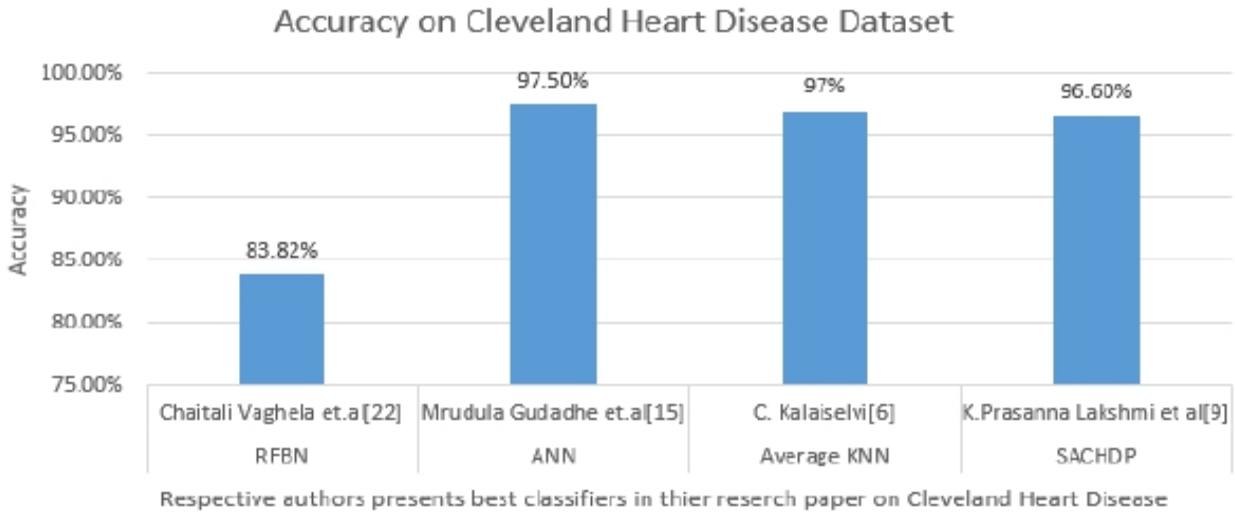

Figure 4. Accuracy on Cleveland Heart Disease Dataset

Veena Vijayan V. et.al, proposed the Computerized Information System using Stacked Generalization for Diagnosis of Diabetes Mellitus [20] for PIMA Indian Diabetes Dataset using Stacking Algorithms with Naive Bayes, Decision Tree, SVM, Decision Stump classifiers. The method Stacked Generalization is the ensemble learning technique in which classification is based on number of base classifiers and Meta classifiers. In this research after evaluation the stacking-Naive Bayes has the highest accuracy (82.32%) as compared to Stacking-Decision stump (81.21%), Staking-Decision Tree (77.90%) and Stacking-SVM (79%) and the best classifiers accuracy from respective paper is shown in figure (3).

Panighrahi Shrikanth et.al, proposed study of various Decision tree classification algorithms such as J48, ADTree, BFTree, LADTree, NBTree, RandomTree. The evaluation result indicates that the Random Tree (100%) classifier having highest accuracy as compared to LADTree(80.72%), BFTree(77.21%), NBTree(78.25%) and J48(84.11%) [18] And the best classifier accuracy as compared to other classifiers is as shown in figure (3).

K.Prasanna Lakshmi et.al[9], designed a decision support system called Stream Associative Classification Heart Disease Prediction(SACHDP) which helps to identify the risk score for predicting the heart disease and their experimental results show that SACHDP performance better when compared to other associative classification techniques with accuracy of 96.60% which is graphically viewed in figure(4). C.Kalaiselvi [6], used K-Nearest Neighbour algorithm for diagnosing heart disease and the experimental result of heart disease prediction gives the classification accuracy achieved is 96.5% with 13 attributes and 97% with 12 attributes which is much better than the existing approaches and the implementation part is done using MATLAB12 and the best classification accuracy shown in figure (4).

Rakesh Motka et.al, uses different data mining techniques for Diabetes Mellitus diagnosis using Neural network, Principle Component Analysis(PCA) with Neural network(NN), Artificial Fuzzy Interference system(ANFIS) and PCA with ANFIS[14]. The research evaluates comparison between these classifiers and get resultant accuracy as NN is 72.9% , ANFIS is 70.56%, PCA with ANFIS is 89.2% and PCA with NN is 90.49% and best accuracy result of respective classifier is as shown in figure(3).

Veena Vijayan V. et.al, proposed a Decision Support System that uses PIMA Indian diabetes dataset for classification [21]. The interpretation of their research found out that the Decision Tree(79.01&) and Naive Bayes(79.01%) having highest accuracy as compare to Support Vector Machine(75.69%) which is graphically viewed in figure(3).

Aishwarya S. et.al, proposed medical decision support system on the basis of Genetic Algorithm and Least Square Support Vector Machine in their research [16]. They used PIMA Indian Diabetes Dataset and classification accuracy of

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

proposed Least Linear SVM with genetic Algorithm is 81.33% and resultant accuracy is graphically viewed in figure(3).

Asma J. AlJarullah used Decision tree classification technique for diagnosis of Type-II Diabetes [3]. Decision tree was generated using classifier J48 decision tree algorithm on PIMA Indian Diabetes dataset and accuracy found out to be 78.17% which is shown comparatively to other classifiers from all reference papers is in figure(3).

Saba Bashir et.al, proposed medical decision support system which is based on efficient rule based classification techniques for diabetes diagnosis [5]. After experimental evaluation of all classification techniques the Bagging technique gives highest accuracy of 91.56% as compare to Adaboost (88.34%), Bayesian (89.33%) and Stacking (85.36%) and graphical view shown in figure (3).

Gaganjot Kaur et.al, proposed improved J48 classification algorithm that helps to improve the accuracy rate of data mining procedure and proposed J48 algorithm gives accuracy 99.87% on PIMA Indian Diabetes dataset [8] as shown in figure(3). Huda Yasin et.al, used Logistic regression technique on Hepatitis disease dataset [23]. Principal Component Analysis method used as a data pre-processing and then logistic regression is apply that obtained the accuracy of 89.60% and the best classifiers accuracy compared to other classifiers from all reference paper is shown in figure (5). A.H.Roslina et.al, proposed the prediction of Hepatitis disease by combining the feature selection method prior to classification processes such as they used Support Vector Machine and Wrapper Method [2]. After experimental evaluation they have got 74.55% of accuracy and the dataset used is Hepatitis dataset taken from UCI Repository. The best experimental evaluation with accuracy is graphically viewed in figure (5).

G.Sathyadevi, developed intelligent medical decision support systems and in their research they used decision trees C4.5 algorithm, ID3 algorithm and CART algorithm to classify Hepatitis disease and compare the accuracy [7]. CART algorithm always generates a binary decision tree. That means the decision tree generated by CART algorithm has exactly two or no child. But the decision tree which is generated by other two algorithms may have two or more child. Also, in respect of accuracy CART algorithm performs better than the other two algorithm with accuracy of 83.18% with respect to the accuracy of ID3 (64.8%) algorithm and C4.5 (71.4%) algorithm as shown in figure (5).

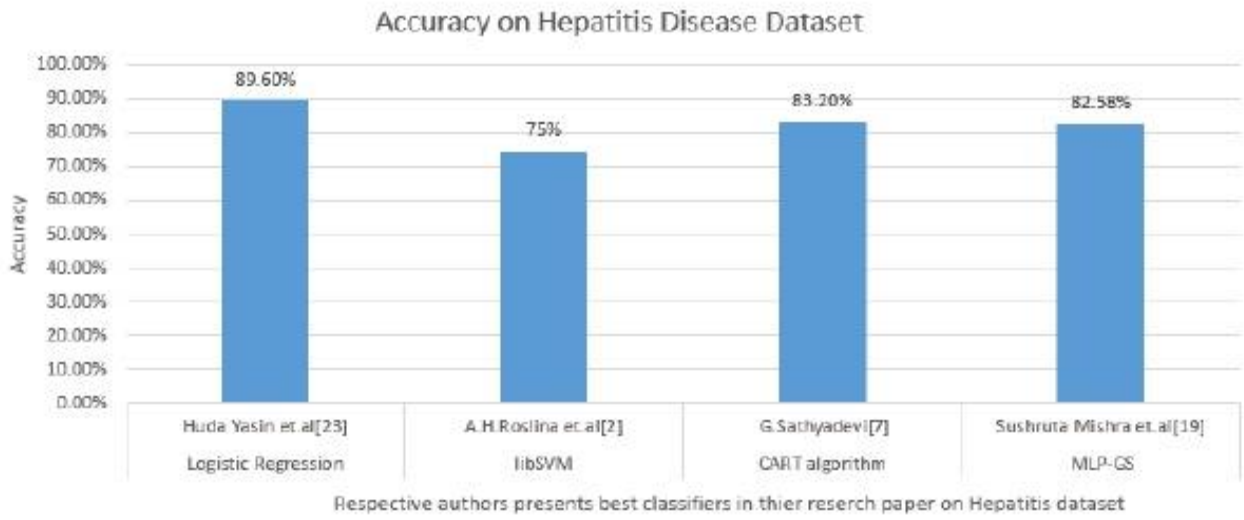

Figure 5. Respective authors presents best classifiers in their research paper

Sushruta Mishra et.al, developed a new hybrid system that can support the physician decision concerning the hepatitis disease treatment [19]. In their study, they developed new model with combination of Genetic search Algorithm(GS) and Multilayer Perceptron(MLP) named as MLP-GS is used to diagnose hepatitis with highest accuracy result is 82.58% and the best classifiers accuracy compared to other classifiers from all reference paper is shown in figure (5).

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

Habib Shariff Mahmud et.al, proposed decision support system for predicting diabetes disease using Decision Tree, Naive Bayes and Multilayer Perceptron. In their research found out that decision tree algorithms have the highest percentage of accuracy of classification with 82.03%, followed by Nave Bayes with 77.60% then multilayer perceptron with accuracy of 76.82%. But in terms of percentage split, when the percentage split is 66%, Nave Bayes outperformed the others with the accuracy of almost80%, followed by multilayer perceptron with 78.54% then Decision Tree with 75.86% and best classifier accuracy result as compare to other classifiers from various reference paper on Hepatitis dataset is as shown in figure (3).

In our survey paper mainly focuses on performance analysis of these various classification algorithms or classification techniques from all the reference papers we have used for this survey. In concern with Classification, this is the process of predicting certain outcomes based on given input datasets. In order to predict the outcome the classification algorithms process a training set containing set of attributes and their respective outcomes. These classification algorithm tries to discover the relationship between the attribute makes it possible to predict the outcomes. The algorithms given a datasets which is not seen before called predictive set which contains the same set of attributes except for the prediction attributes not yet known. The figure (3) represents various classifiers gives best classification accuracy results on PIMA Indian Diabetes Datasets. figure (4) represents various classifiers gives best classification accuracy results on Cleveland Heart Disease Dataset and figure (5) represents various classifiers gives best classification accuracy results on Hepatitis dataset.

V. PERFORMANCE ANALYSIS

In our study we have analyzed several research paper based on Clinical Decision Support System regarding to three datasets named as PIMA Indian Diabetes Dataset, Cleveland Heart Disease Dataset and Hepatitis Disease Dataset. There are various data pre-processing techniques such as feature selection, feature extraction, tree pruning, Principal Component Analysis, Discretization, etc. and classification techniques used in respective research paper are analyzed. On the basis of analysis of each paper used in our research we have presented the best accuracy result of each of best classifiers on respective datasets used from each of paper. Detailed study and analysis of selected research papers on three datasets used in our paper the various classifiers behaves well and increases their accuracy

VI. CONCLUSION

Data mining techniques are well suited for decision taking and prediction in medical diseases diagnosis. This paper provides overview of the performance analysis of various data mining techniques and their respective results and comparisons. Using various data preprocessing techniques that plays effective role and increases the classification accuracy on all three datasets used for this study using different classifiers as well as combination of various classifiers gives efficient average accuracy on all datasets.

REFERENCES

1 University of UtahSchool of medicine's dept. of medical informatics. http://www.openclinical.org/aispiliad.html.
2 A.H.Roslina and A.Noraziah. "Prediction Of Hepatitis Prognosis Using Support Vector Machines And Wrapper Method ".IEEE Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD),pp.2209–2211, 2010.
3 Asma A. AlJarullah. ,"Decision Tree Discovery for Diagnosis of Type II Diabetes.", IEEE International Conference onInnovations in Information Technology (IIT), pp.303–307, 2011.
4 Hupp JA, Hoffer EP, Barnett GO and Cimino JJ, "DXplain. An evolving diagnostic decision support system". Journal of American Medical Association, Vol.258, Issue.1, pp.67–74,1987.
5 Saba Bashir, Usman Qamar, Farah Hassan Khan, and M.Younus Javed., "An Efficient Rule-Based Classification of Diabetes Using ID3,C4.5 and CART Ensembles", IEEE 12th International Conference onFrontiers of Information Technology (FIT),Vol.978,Issue.1,pp.226–231, 2014.
6 C.Kalaiselvi.,"Diagnosing of Heart Diseases using Average K-Nearest Neighbor Algorithm of Data Mining", IEEE 3rd International Conference onComputing for Sustainable Global Development (INDIACom), Vol.978, Issue.9, pp.3099–3103,2016.
7 G.Sathyadevi. "Application of CART algorithm in Hepatitis Disease Diagnosis", IEEE International Conference on Recent Trends in Information Technology (ICRTIT),Vol.978, Issue.1, pp.1283–1287,2011.
8 Gaganjot Kaur and Amit Chhabra. "Improved J48 Classification Algorithm for the prediction of Diabetes",International Journal of Computer Application,Vol.98, Issue.22, pp.13-17, 2014.
9 K.Prasanna Lakshmi and Dr. C.R.K.Reddy. "Fast RuleBased Heart Disease Prediction using Associative Classification Mining", IEEE International Conference onComputer, Communication and Control (IC4), Vol.978, 2015.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

10 UCI Machine learning Repository. Breast cancer dataset. “https://archive.ics.uci.edu/ml/datasets/”.
11 UCI Machine learning Repository. Hepatitis disease dataset. “https://archive.ics.uci.edu/ml/datasets/”.
12 *UCI Machine learning Repository. Pima indian diabetes dataset. “https://archive.ics.uci.edu/ml/datasets/”*.
13 Habib Shariff Mahmud, “Application of Data Mining in Medical Decision Support System”. International Journal of Information System and Engineering, Vol.1, Issue.1,pp.1-14, 2015.
14 Rakesh Motka, Viral Parmar, Balbildra Kumar, and A.R.Verma. “Diabetes Mellitus Forecast Using Different Data Mining Techniques”, IEEE 4th International Conference on Computer and Communication Technology (ICCCT), Vol.978, pp.99–103, 2013.
15 Kapil Wankhade Mrudula Gudadhe and Snehlata Dongre. “Decision Support System for Heart Disease Based on Support Vector Machine and Artificial Neural Network”, IEEE International Conference on Computer and Communication Technology (ICCCT),Vol.978,pp.741–745, 2010.
16 Aishwarya S. and Anto S. “A Medical Decision Support System Based on Genetic Algorithm and Least Square Support Vector Machine for Diabetes Disease Diagnosis”, International Journal of EngineeringScience& Research Technology, Vol.3, Issue.4,pp.4042–4046, 2014.
17 D.Kanellopoulos S.B.Kotasiantis and P.E.Pintelas. “Data Preprocessing for Supervised Learning”, International Journal of Computer Science, Vol.1, Issue.1, pp.111–117, 2006.
18 Panigrahi Shrikanth and Dharmaih Deverpalli. “A Critical Study of Classification Algorithm Using Diabetes Diagnosis”, IEEE 6th International Conference Advanced Computing (IACC), Vol.978,pp.245–249, 2016.
19 Brojo Kishore Mishra Sushruta Mishra and Hrudaya Kumar Thripathy. “A Neuro-Genetic Model to Predict Hepatitis Disease Risk”, IEEE International Conference onComputational Intelligence and Computing Research (ICCIC), Vol.978,2015.
20 Veena Vijayan V. and Anjali C. “Computerized Information System Using Stacked Generalizatio n for Diagnosis of Diabetes Mellitus.”. IEEE Recent Advances in Intelligent Computational Systems (RAICS), Vol.978, pp.173–178, 2015.
21 Veena Vijayan V. and Anjali C. “Decision Support System for Predicting Diabetes Mellitus-A Review.”, Global Conference on Communication Technologies (GCCT), pp.98–102, 2015.
22 Chaitali Vaghela, Nikita Bhatt, and Darshana Mistry. “A Survey on Various Classification Techniques for Clinical Decision Support System.”,International Journal on Computer Application, Vol.116, Issue.23, pp.14-17, 2015.
23 Huda Yasin, Tahseen A. Jilan, and Madiha Danish. “Hepatitis-C Classification using Data Mining Techniques.”, International Journal of Computer Applications, Vol.24, Issue.3, pp.1-6,2011

BIOGRAPHY

Vikram Vitthalrao Kaleis a Final year M.Tech student in the Computer Science and Information Technology Department, SGGS Institute of Engineering and Technology, Vishnupuri, Nanded, Maharashtra, India. His research interests are Data mining and machine learning.

Dr. B. R. Bombadeis Associate Professor in the Computer Science and Information Technology Department, SGGS Institute of Engineering and Technology, Vishnupuri, Nanded, Maharashtra, India. His research interests are Data mining and pattern recognition.