Performance Comparison of Classification Algorithms on Medical Datasets

Venkata Ramana Bendi and Raja Sarath Kumar Boddu

January 6, 2020

Performance Comparison of Classification Algorithms on Medical Datasets

Venkata Ramana Bendi [1,] Raja Sarath Kumar Boddu [2]

[1] Department of Information Technology, Aditya Institute of Tech & Magt, Tekkali, India.

[2] Department of CSE, Lenora College of Engineering, India.

Abstract:

This paper studies selected classification algorithms on medical datasets. The selected medical datasets are Breast Cancer Data, Chronic Kidney Disease, Cryotherapy, Hepatitis, Immunotherapy, Indian Liver Patient Dataset (ILPD), Liver Disorders, and Liver disorders dataset. ILPD and Liver disorders, pima diabetes, risk factors cervical cancer and Statlog (Heart) Data Set dataset are taken from the UCI repository. The classification algorithms considered here are Naive Bayes, J48, Multilayer perceptron, JRip, IBK and bagging classifiers .

Keywords: *Classifiers, Medical Datasets, Performance Evaluation*

I. INTRODUCTION:

Machine learning methods are widely used for medical diagnosis. Machine learning algorithms were designed to analyze medical datasets and it is currently well suited for analyzing medical data. Machine learning promises to help physicians make perfect diagnoses and help him to choose the best medications for their patients. In this paper, Naive Bayes, J48, Multilayer perceptron, JRip, IBK and bagging classifiers are taken from bayes, Decision tree, Neural network, Rule based, KNN and Meta classifiers category respectively. The performance of these classifiers are evaluated based on accuracy, sensitivity, precision, specificity and ROC Area. In this analysis binary class medical datasets are taken from University of California at Irvine (UCI) Machine Learning Repository [12][16] and 10-fold cross-validation has been used.

II. RELATED WORK:

Yaqiang Wang et al [1] employed Naive Bayes classifier and Support Vector Machine classifier and proposed a novel framework of automatic diagnosis of TCM utilizing raw free-text clinical records for clinical practice. H.S.Hota [2] In this research work ,various intelligent techniques including supervised Artificial Neural Network (ANN) ,unsupervised Artificial Neural Network ,Statistical and decision tree based have been applied to classify data related to breast cancer health care obtained from UCI repository site. Sina Bahramirad et al [3] proposed eleven data mining classification algorithms and applied on two real liver patient datasets and the performance of all classifiers are compared against each other in terms of accuracy, precision, and recall. Munaza Ramzan [4] implemented data mining classification algorithms like J48, naive-bayes and random-forest on medical datasets for better predictions and supports in decision making in diagnosing Cancer, Cardiovascular diseases and Diabetes. Kun- Hong Liu and De-Shuang Huang [7] addressed the microarray dataset based cancer classification using rotation forest. Principal component analysis (PCA) was applied to feature transformation in the original rotation forest. In this paper Independent component analysis (ICA) was applied on breast cancer dataset and prostate dataset to validate the efficiency of rotation forest. Akin Ozcift and Arif Gultenb [9] proposed rotation forest (RF) a new ensemble classifiers of 30 machine learning algorithms to evaluate their classification performances using Parkinson's, diabetes and heart diseases.Bendi Venkata Ramana et al. [10] compared popular Classification Algorithms for evaluating their classification performance in terms of Accuracy, Precision, Sensitivity and Specificity in classifying liver patients dataset. Accuracy, Precision, Sensitivity and Specificity are better for the AP Liver Dataset compared to UCI liver datasets with all the selected algorithms. Bendi Venkata Ramana et al. [11] proposed ANOVA and MANOVA for population comparision between ILPD data set and UCI data set. The results indicates that there exists more significant difference in the groups with all the possible attribute combinations except analysis on SGPT between non liver patients of UCI and INDIA data sets. Bendi Venkata Ramana et al. [13] proposed Bayesian Classification for diagnosis of liver diseases. The Bayesian Classification is combined with Bagging and Boosting for better accuracy. This accuracy can be further improved with huge amount of data. Ayse Cufoglu et al. [17] compared four different classification algorithms which are Naïve Bayesian (NB), Instance-Based Learner (IB1), Bayesian networks (BN) and Lazy Learning of Bayesian Rules (LBR) classifiers and The simulation results show that, the NBTree has the highest classification accuracy performance with the lowest error rate. Rong-Ho Lin [18] proposed an intelligent model for the diagnosis of liver diseases which integrates classification and regression tree (CART) and case-based reasoning (CBR) techniques. CART is used to extract rules from health examination data to show whether the patient suffers from liver disease whereas CBR is developed to diagnose the type of liver disease.

III. MEDICAL DATASETS:

The selected medical datasets were taken from University of California at Irvine (UCI) Machine Learning repository [16]. The medical datasets have been specified with their set of attributes and instances and they are presented in table 1.

Table 1: UCI Medical Datasets

Medical Datasets	Breast Cancer Data	Chronic Kidney Disease	Cryotherapy	Hepatitis	Immunotherapy	Indian Liver Patient Dataset (ILPD)	Liver Disorders	Pima diabetes	Risk factors cervical cancer	Statlog (Heart) Data Set
Attributes	11	25	7	20	8	11	7	9	36	14
Instances	699	400	90	155	90	583	345	768	858	270

IV. RESULTS AND DISCUSSION:

In this study selected medical datasets were taken from University of California at Irvine (UCI) repository. Selected classifiers Naive Bayes, C 4.5, Multi layer perception, Rule based, KNN and bagging are considered for performance evaluation based on accuracy, sensitivity, precision, specificity and ROC Area. The performance evaluation is depicted based on confusion matrix. A confusion matrix is a technique for summarizing the performance of a classification algorithm and it shows the ways in which your classification model is confused when it makes predictions. Example of confusion matrix for a binary classifier presented in figure 1.

	Predicted: NO	**Predicted: YES**
Actual: NO	TN	FP
Actual: YES	FN	TP

Figure 1. Confusion Matrix for a binary classifier

- True positives (TP) : Predicted yes and they do have the disease.
- True negatives (TN) : Predicted no and they don't have the disease.
- False positives (FP) : Predicted yes, but they don't actually have the disease.
- False negatives (FN) : Predicted no, but they actually do have the disease.

The performance of classifiers is based on accuracy, sensitivity, precision, specificity and ROC Area.

Accuracy: The accuracy of a classifier is the percentage of the test set tuples that are correctly classified by the classifier.

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

Sensitivity: Sensitivity is also referred as True positive rate i.e the proportion of positive tuples that are correctly identified.

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

Precision: precision is defined as the proportion of the true positives against all the positive results (both true positives and false positives)

$$\text{Precision} = \frac{TP}{TP + FN}$$

Specificity: Specificity is the True negative rate that is the proportion of negative tuples that are correctly identified.

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

Where TP means True Positives, TN means True Negatives, FP means False Positives and FN means False Negatives.

ROC: The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The true-positive rate is also known as sensitivity, recall or probability of detection. The false-positive rate is also known as the fall-out (1 − specificity). The

Figure1shows three ROC curves representing excellent, good, and worthless tests plotted on the same graph. The accuracy of the test depends on how well the test separates the group being tested into those with and without the disease in question. Accuracy is measured by the area under the ROC curve. An area of 1 represents a perfect test and an area of .5 represents a worthless test presented in figure 2.

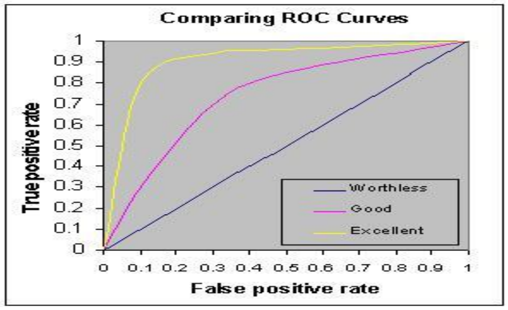

Figure 2. ROC Curve

Table 2: Accuracy of selected classifiers for medical datasets

Accuracy										
Diseases → Classifiers ↓	Breast Cancer Data	Chronic Kidney Disease	Cryotherapy	Hepatitis	Immuno therapy	Indian Liver Patient Dataset (ILPD)	Liver Disorders	Pima diabetes	Risk factors cervical cancer	Statlog (Heart) Data Set
Bagging	95.85	98.75	88.89	64.52	84.44	69.3	69.57	75.78	96.15	80
IBK	95.14	95.75	90	66.45	70	64.49	62.9	70.18	94.41	75.19
J48	94.56	99	93.33	58.06	82.22	68.78	68.7	73.83	95.1	76.67
JRip	96.28	97.75	87.78	63.23	82.22	66.38	66.67	76.04	96.15	80.74
MP	95.85	99.75	87.78	62.58	80	68.95	71.59	75.39	94.76	77.41
NB	95.99	95	83.33	71.61	76.67	55.75	55.36	76.3	88.69	83.7

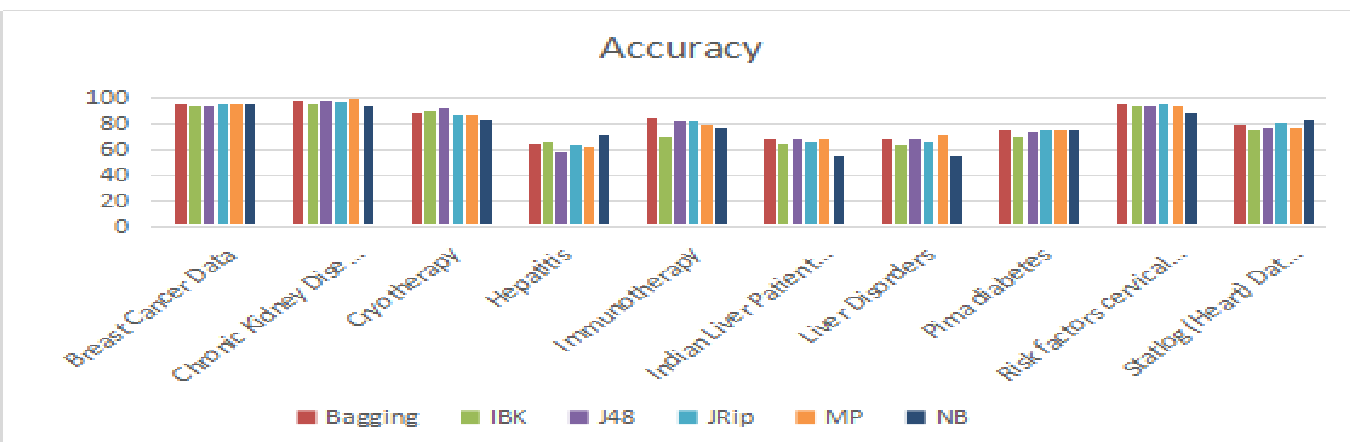

Figure 3. Accuracy of classifiers on medical datasets

Table 3: Sensitivity of selected classifiers for medical datasets

Sensitivity										
Diseases→ Classifiers ↓	Breast Cancer Data	Chronic Kidney Disease	Cryotherapy	Hepatitis	Immunotherapy	Indian Liver Patient Dataset (ILPD)	Liver Disorders	Pima diabetes	Risk factors cervical cancer	Statlog (Heart) Data Set
Bagging	95.44	99.33	83.33	51.43	47.37	28.74	79	58.58	81.82	82
IBK	92.12	100	89.58	60	21.05	47.9	67.5	52.99	47.27	76.67
J48	92.53	98	89.58	47.14	47.37	33.53	80	59.7	67.27	79.33
JRip	96.68	96	77.08	48.57	47.37	17.96	74.5	58.21	87.27	85.33
MP	95.02	100	85.42	61.43	36.84	28.14	82	60.82	54.55	78
NB	97.51	100	89.58	57.14	21.05	95.21	40	61.19	74.55	87.33

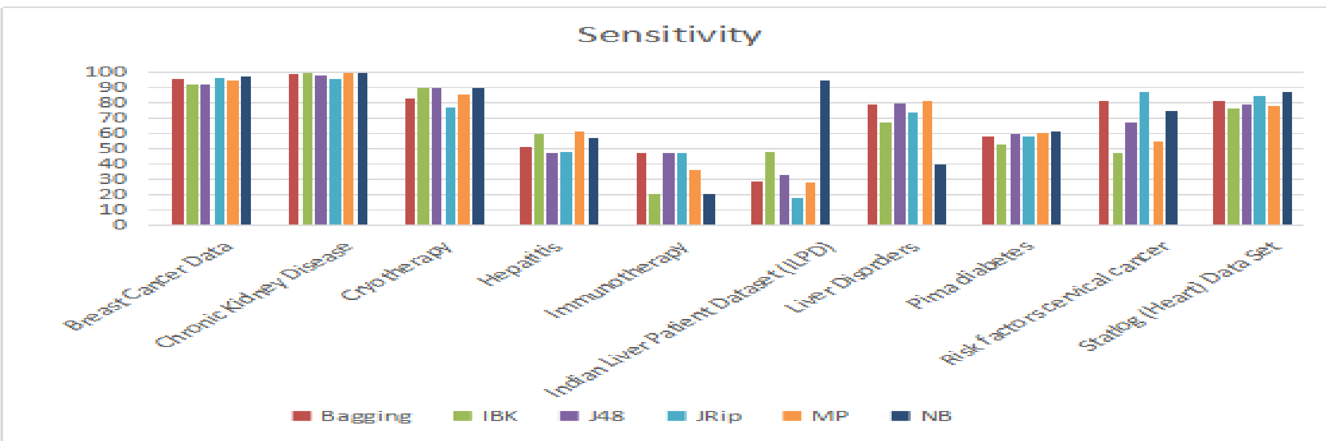

Figure 4. Sensitivity of classifiers on medical datasets

Table 4: Precision of selected classifiers for medical datasets

Precision										
Diseases→ Classifiers ↓	Breast Cancer Data	Chronic Kidney Disease	Cryotherapy	Hepatitis	Immunotherapy	Indian Liver Patient Dataset (ILPD)	Liver Disorders	Pima diabetes	Risk factors cervical cancer	Statlog (Heart) Data Set
Bagging	92.74	97.39	95.24	63.16	69.23	44.44	71.49	67.67	66.18	82
IBK	93.67	89.82	91.49	63.64	25	40	68.18	57.96	57.78	78.23
J48	91.77	99.32	97.73	54.1	60	44.09	70.18	63.24	60.66	78.81
JRip	92.83	97.96	100	61.82	60	33.71	69.95	68.42	64.86	81.01
MP	93.09	99.34	91.11	58.11	53.85	43.52	72.57	65.99	60	80.69
NB	91.44	88.24	81.13	74.07	40	38.88	70.18	67.77	33.06	83.97

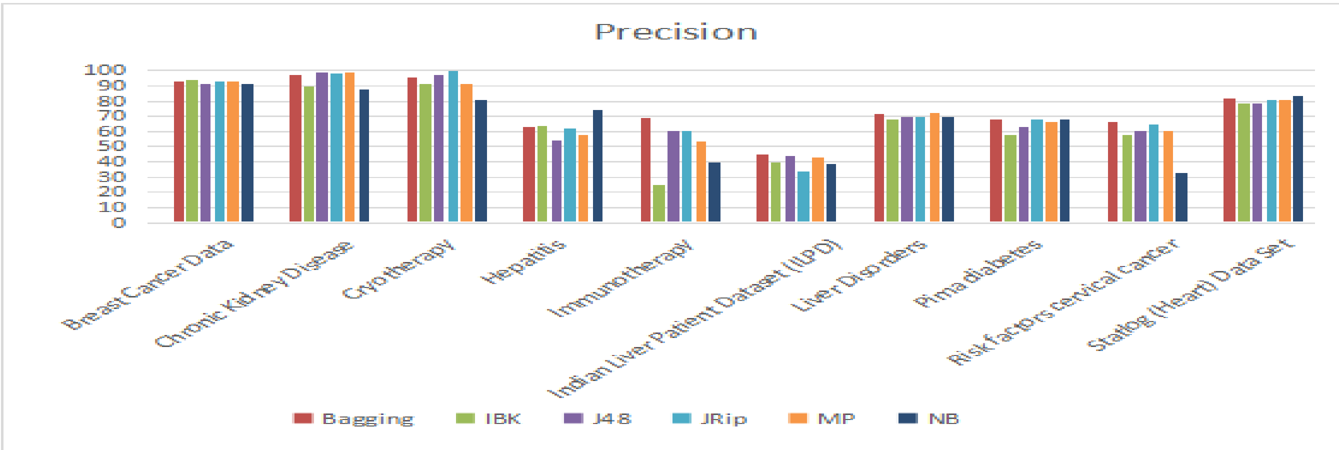

Figure 5. Precision of classifiers on medical datasets

Table 5: Specificity of selected classifiers for medical datasets

Specificity										
Diseases → Classifiers ↓	Breast Cancer Data	Chronic Kidney Disease	Cryotherapy	Hepatitis	Immunotherapy	Indian Liver Patient Dataset (ILPD)	Liver Disorders	Pima diabetes	Risk factors cervical cancer	Statlog (Heart) Data Set
Bagging	96.07	98.4	95.24	75.29	94.37	85.58	56.55	85	97.14	77.5
IBK	96.72	93.2	90.48	71.76	83.1	71.15	56.55	79.4	97.63	73.33
J48	95.63	99.6	97.62	67.06	91.55	82.93	53.1	81.4	97.01	73.33
JRip	96.07	98.8	100	75.29	91.55	85.82	55.86	85.6	96.76	75
MP	96.29	99.6	90.48	63.53	91.55	85.34	57.24	83.2	97.51	76.67
NB	95.2	92	76.19	83.53	91.55	39.9	76.55	84.4	89.66	79.17

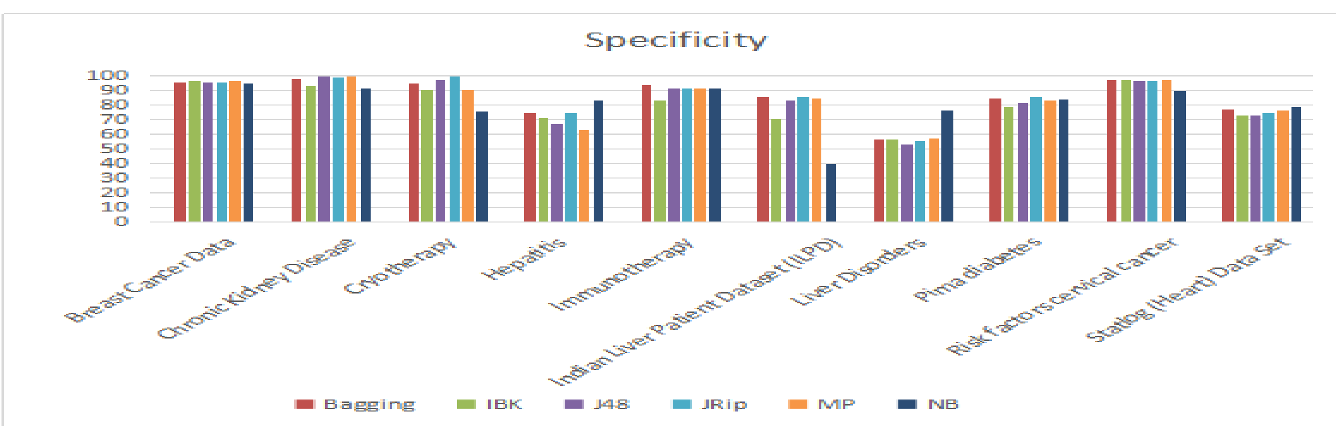

Figure 6. Specificity of classifiers on medical datasets

Table 6: ROC Area of selected classifiers for medical datasets

ROC Area										
Diseases→ Classifiers ↓	Breast Cancer Data	Chronic Kidney Disease	Cryotherapy	Hepatitis	Immunotherapy	Indian Liver Patient Dataset (ILPD)	Liver Disorders	Pima diabetes	Risk factors Cervical cancer	Statlog (Heart) Data Set
Bagging	0.989	0.999	0.932	0.723	0.773	0.707	0.745	0.812	0.919	0.873
IBK	0.945	0.966	0.902	0.678	0.521	0.572	0.630	0.650	0.738	0.750
J48	0.955	0.999	0.923	0.607	0.662	0.673	0.665	0.751	0.846	0.744
JRip	0.965	0.976	0.913	0.596	0.643	0.535	0.673	0.739	0.888	0.809
MP	0.989	1.000	0.926	0.642	0.759	0.708	0.742	0.793	0.904	0.853
NB	0.986	1.000	0.935	0.759	0.701	0.726	0.640	0.819	0.875	0.898

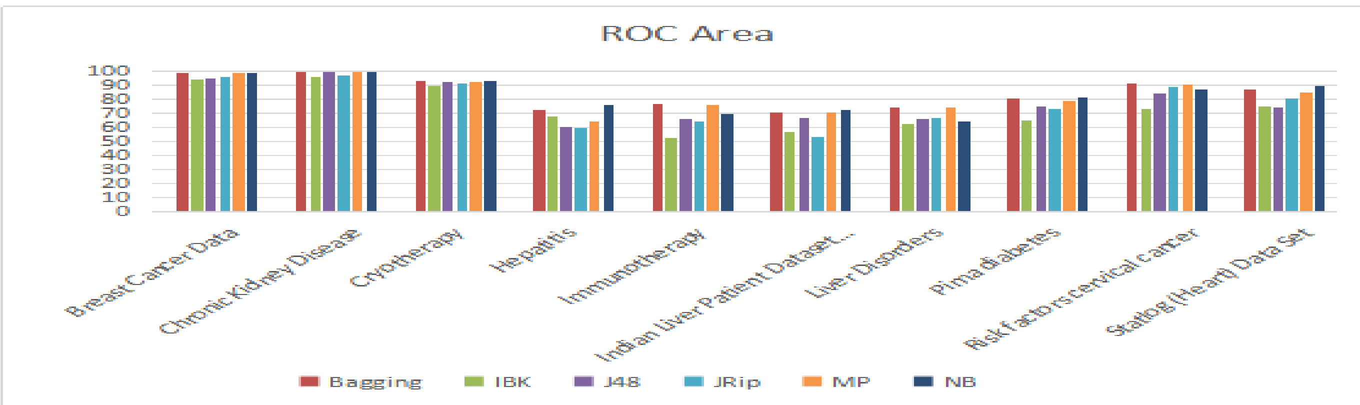

Figure 7. ROC Area of classifiers on medical datasets

The results reported in Table 2, Table 3, Table 4, Table 5 & Table 6 are the Accuracy, sensitivity, precision, specificity and ROC area of selected classifiers for medical datasets respectively. The same results are represented graphically in figure 3, figure 4, figure 5, figure 6 and figure 7 respectively. In this analysis Multilayer perception classifier has the highest accuracy i.e 99.75 % on Chronic Kidney Disease dataset. Sensitivity is 100 % for both Multilayer perception classifier and Naive bayes classifier. Rule based classifier has highest precision and Specificity i.e 100 % on Cryotherapy. ROC area is highest i.e 1 for both Multilayer perception classifier and Naive bayes classifier on Chronic Kidney Disease dataset.

V. CONCLUSIONS:

In this study, selected classification algorithms were considered for evaluating their classification performance in terms of accuracy, sensitivity, precision, specificity and ROC area. The accuracy is high for multilayer perceptron on Chronic Kidney Disease dataset. Sensitivity is high for both Multilayer perception classifier and Naive bayes classifier. Precision and Specificity are high for Rule based classifier on Cryotherapy dataset. ROC area is high for both Multilayer perception classifier and Naive bayes classifier on Chronic Kidney Disease dataset. This may be due to Chronic Kidney Disease dataset has highest number of attributes among all the medical datasets and Multilayer perception classifier is one of the neural network classifier which gives highest classifier performance.

VI.REFERENCES:

[1]. Yaqiang Wang et al, A framework and its empirical study of automatic diagnosis of traditional chinese medicine utilizing raw free-text clinical records, journal of biomedical informatics, 45(2012) 210-223.

[2]. H.S.Hota, Diagnosis of Breast Cancer Using Intelligent Techniques, International Journal of Emerging Science and Engineering , ISSN: 2319–6378, Volume-1, Issue-3, January 2013 .

[3]. Sina Bahramirad et al, Classification of Liver Disease Diagnosis: A Comparative Study, ISBN: 978-1-4673-5256-7, 42-46, 2013, IEEE.

[4]. Munaza Ramzan, Comparing and Evaluating the Performance of WEKA Classifiers on Critical Diseases, ISBN: 978-1-4673-6984-8, 2016, IEEE.

[5]. Theresa Princy. R and J. Thomas, Human Heart Disease Prediction System using Data Mining Technique, International Conference on Circuit, Power and Computing Technologies [ICCPCT], 2016.

[6]. Mira Kania Sabariah et al, Early Detection of Type II Diabetes Mellitus with Random Forest and Classification and Regression Tree (CART), International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA), 238-242, 2014.

[7]. Kun-Hong Liu and De-Shuang Huang: "Cancer classification using Rotation forest". In Proceedings of the Computers in biology and medicine, 38, pages 601-610, 2008.

[8]. Juan J. Rodrı´Guez, Member, Ludmila I. Kuncheva and Carlos J. Alonso:" Rotation Forest: A New Classifier Ensemble Method". In Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, No. 10, pages 1619-1630, 2006.

[9]. Akin Ozcift and Arif Gulten:" Classifier Ensemble Construction With Rotation Forest to Improve Medical Diagnosis Performance Of Machine Learning Algorithms". In Proceedings of the Computer Methods and Programs in Biomedicine , pages 443-451, 2011.

[10]. B. V. Ramana, M. S. P. Babu, and N. B. Venkateswarlu, "A critical study of selected classification algorithms for liver disease diagnosis," Int. J. Database Manage. Syst., vol. 3, no. 2, pp. 101–114, 2011.

[11]. Bendi Venkata Ramana, Prof. M. S. Prasad Babu and Prof. N. B. Venkateswarlu, "A Critical Comparative Study of Liver Patients From USA and INDIA: An Exploratory Analysis", International Journal of Computer Science Issues, Vol.9, issue 3, May 2012.

[12]. Weka-3.9 : data mining with open source machine learning software © 1999-2016, The University of Waikato, Hamilton, New Zealand.

[13]. Bendi Venkata Ramana, Prof. M.Surendra Prasad Babu and Prof. N. B. Venkateswarlu: "A Critical Evaluation of Bayesian Classifier For Liver Diagnosis Using Bagging and Boosting Methods", International Journal of Engineering Science and Technology (IJEST), Vol.3, No.4, pages 3422-3426, April 2011.

[14]. Gang Zheng et al, Study of Classification Rules on Weighted Coronary Heart Disease Data, Proceedings of The Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, August 2005, 1845-1850.

[15]. B. V. Ramana and Prof. M. S. Prasad Babu, Liver Classification using Modified Rotation Forest, International Journal of Engineering Research and Development, Vol 1, Issue 6, pp. 17- 24, 2012.

[16] UCI repository of machine learning databases available from ftp://ftp.ics.uci.edu/pub/machine-learningdatabases/, last accessed: 16th May 2018.

[17] Ayse Cufoglu, Mahi Lohi and Kambiz Madani:" A Comparative Study of Selected Classifiers With Classification Accuracy In User Profiling". In Proceedings of the IEEE International Conference on Computer Science and Information Engineering, pages 708-712, 2008.

[18] Rong-Ho Lin: "An intelligent model for liver disease diagnosis". In Proceedings of the Artificial intelligence in medicine , 47, pages 53-62, 2009.