Research Paper Computer Science

# Receivers Operating Characteristics for Heart Diseases Dataset


| | |
|---|---|
| **R.KARTHIKEYAN** | Assistant Professor, Department of Computer Science and Engineering. Bharath University, Selaiyur, Chennai-600073, India |
| **A.KUMARAVEL** | Professor and Dean, Department of Computer Science and Engineering.Bharath University, Selaiyur, Chennai-600073, India |
| **V.KHANAA** | Professor, and Dean Department of Information Techonology. Bharath University, Selaiyur, Chennai-600073 India |


**ABSTRACT**

Performance measure in the context like diagnosis of heart disease requires more attention by the data analytic community. Summative measures vary sometimes drastically due to variations is the distribution of instances in the underlying dataset. Here we address the problem of visualizing the sensitivity versus specificity in the graph in spite of the presence of appreciable accuracy. In this paper we establish the value of area under the ROC curve maximum around 0.9063 and the maximum accuracy 75.667% iterating over the familiar set of classifiers.

**KEYWORDS**

## 1. Introduction

Absolute mean error, root mean square, etc are the summative measures showing the performance deviation where as measures like accuracy, Area under cure gives strength to the recommendations for the models generated .In this paper ROC measures are considered for recognizing patterns in the heart disease dataset.

The preprocessed dataset is considered for attributes reduction. Almost one tenths of the attributes are identified as significant by simple 'Subset evaluation'method with 'Best first' search technique. Classification is performed iteratively over the classes of tree classifier, Rules classifier, Meta classifier, Bayes classifier and finally the ROC are plotted for classifiers which yield maximum accuracy.

## 2. Dataset Collection and Data Preparation

In this section, we dwell the collection of data and format in which the data has to be presented for mining experiments following the iterative steps in Figure 1. We use java based implementation namely Weka tool from University of Waikato, New Zealand.

The datasets for these experiments are from [4].The original data format has been slightly modified and extended in order to get relational format.

### 2.1 Dataset description:

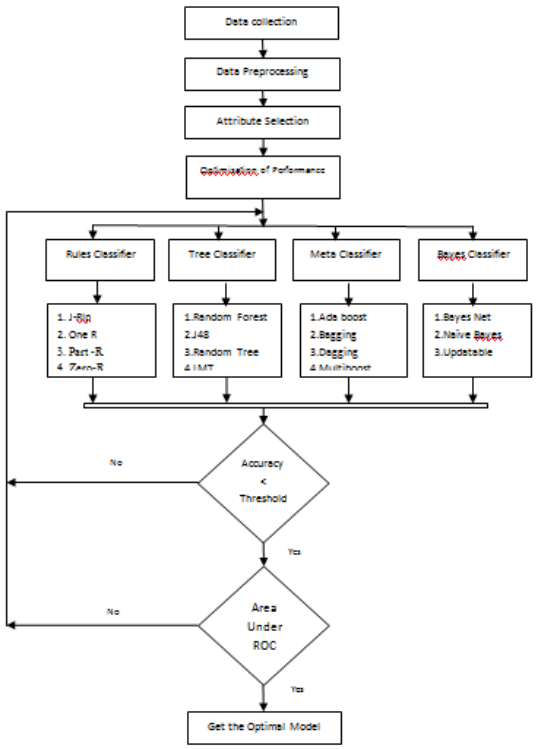
The database of diabetes describes a set of 279 attributes[12] as shown in the below list 2.2. This dataset carries 16 categories of classes. The number of instances in this database is 452.

| Class code | Class | Number of instances |
|---|---|---|
| 01 | Normal | 245 |
| 02 | Ischemic changes (Coronary Artery Disease) | 15 |
| 03 | Old Anterior Myocardial Infarction | 15 |
| 04 | Old Inferior Myocardial Infarction | 15 |
| 05 | Sinus tachycardy | 13 |
| 06 | Sinus bradycardy | 25 |
| 07 | Ventricular Premature Contraction (PVC) | 3 |
| 08 | Supraventricular Premature Contraction | 07 |
| 09 | Left bundle branch block | 09 |
| 10 | Right bundle branch block | 50 |
| 11 | 1. degree AtrioVentricular block | 0 |
| 12 | 2. degree AV block | 0 |
| 13 | degree AV block | 0 |
| 14 | Left ventricule hypertrophy | 4 |
| 15 | Atrial Fibrillation or Flutter | 5 |
| 16 | Others | 22 |

## 3. Methods Description

Here we select standard set of methods [13] for predicting from the data set described above. We consider three types of classifiers for our study such as tree based, Bayes approach based, and Meta level based classifiers. The following sections describe briefly the methods for classifier and results of such methods are tabulated further. Then final results are interpreted

**3.Methods for the proposed frame Work**
**Figure.1 Identifying the Optimal classifier in the proposed framework**




**3.1 Tree Classifiers:**
Supervised Learning is performed conducted using tree classifiers .We select four types of tree classifiers as shown below.

**3.1.1 J48**
The first number is the total number of instances (weight of instances) reaching the leaf. The second number is the number (weight) of those instances that are misclassified.

If your data has missing attribute values then you will end up with fractional instances at the leafs. When splitting on an attribute where some of the training instances have missing values, J48 will divide a training instance with a missing value for the split attribute up into fractional parts proportional to the frequencies of the observed non-missing values. This is discussed in the Witten & Frank Data Mining book as well as Ross Quinlan's original publications on C4.5.

**3.1.3 Random Forest**
Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The generalization error for forests converges a.s. to a limit as the number of trees in the forest becomes large. 3.1.3 Random Tree

Class for constructing a tree that considers K randomly chosen attributes at each node. Performs no pruning. Also has an option to allow estimation of class probabilities based on a hold-out set back fitting.

**3.1.4 LMT**
Classifier for building 'logistic model trees', which are classification trees with logistic regression functions at the leaves. The algorithm can deal with binary and multi-class target variables, numeric and nominal attributes and missing values.

**3.2 Bayes Classifiers**
These types of classifiers includes probability measure for the class values and comes under supervised learning.

**3.2.1 Bayes Net**
Bayes Network learning using various search algorithms and quality measures. Base class for a Bayes Network classifier. Provides data structures and facilities common to Bayes Network learning algorithms like K2 and B.

**3.2.2 Naïve Bayes**
Class for a Naive Bayes classifier using estimator classes. Numeric estimator precision values are chosen based on analysis of the training data. For this reason, the classifier is not an Updateable Classifier you need the Updateable Classifier functionality, use the Naïve Bayes Updateable classifier.

**3.2.3 Naive BayesUpdateable**
Class for a Naive Bayes classifier using estimator classes. This is the updateable version of Naïve Bayes. This classifier will use a default precision of 0.1 for numeric attributes when build Classifier is called with zero training instances.

**3.3 Meta Classifiers**
Most of the time, the aggregation of more than one classifier has better performance. Such combinational methods are shown below.

**3.3.1 Adaboost**
Class for boosting a nominal class classifier using the Adaboost M1 method. Only nominal class problems can be tackled. Often dramatically improves performance, but sometimes over fits.

**3.3.2 Bagging**
Class for bagging a classifier to reduce variance. Can do classification and regression depending on the base learner. Generate B bootstrap samples of the training data: random sampling with replacement. Train a classifier or a regression function using each bootstrap sample for classification: majority vote on the classification results. For regression: average on the predicted values.

**3.3.3 Dagging**
This meta classifier creates a number of disjoint, stratified folds out of the data and feeds each chunk of data to a copy of the supplied base classifier. Predictions are made via averaging, since all the generated base classifiers are put into the Vote meta classifier. Useful for base classifiers that are quadratic or worse in time behavior, regarding number of instances in the training data.

**3.3.4 Multi Boost**
MultiBoosting is an extension to the highly successful AdaBoost technique for forming decision committees. It is able to harness both AdaBoost's high bias and variance reduction with wagging's superior variance reduction. Using C4.5 as the base learning algorithm, Multi-boosting is demonstrated to produce decision committees with lower error than either AdaBoost or wagging significantly more often than the reverse over a large representative cross-section of UCI data sets. It offers the further advantage over AdaBoost of suiting parallel execution.

**3.4 Rules Classifiers**
**3.4.1 J-Rip**
This class implements a propositional rule learner, Repeated Incremental Pruning to Produce Error Reduction (RIPPER), which was proposed by William W. Cohen as an optimized version of IREP.

**3.4.2 One-R**
Class for building and using a 1R classifier; in other words, uses the minimum-error attribute for prediction, discretizing numeric attributes.

**3.4.3 PART**
Class for generating a PART decision list. Uses separate-and-conquer. Builds a partial C4.5 decision tree in each iteration and makes the "best" leaf into a rule.

**3.4.4 Zero-R**
Class for building and using a 0-R classifier. Predicts the mean for a numeric classor the mode for a nominal class.

**4.1 Performance Measure:**
Accuracy: probably the most widely used performance metric in Machine Learning. It is defined as the proportion of correct predictions the classifier makes relative to the size of the dataset. If a classifier has continuous outputs (e.g. neural nets), a threshold is set and everything above this threshold is predicted to be a positive.

Root-mean-squared-error (RMSE): widely used in regression, it measures how much predictions deviate from the true targets. 1RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{N}\sum(\text{Pred}(c) - \text{True}(C))^2}$$

Lift: often used in marketing analysis, Lift measures how much better a classifier is at predicting positives than a baseline classifier that randomly predicts positives (at the same rate observed for positives in the data). The definition is:

$$LIFT = \frac{\%\text{of true positives above the threshold}}{\%\text{of dataset above the threshold}}$$

Usually the threshold is set so that a fixed percentage of the dataset is classified as positive. For example, suppose a marketing agent wants to send advertising to potential clients, but can only afford to send ads to 10% of the population. A classifier is trained to predict how likely a client is to respond to the advertisement, and the ads are sent to the 10% of the population predicted most likely to respond. A classifier with optimal lift will get as many clients as possible that will respond to the advertisement in this set.

Precision and recall : These measures are widely used in Information Retrieval. Precision is the fraction of examples predicted as positive that are actually positive. Recall is the fraction of the true positives that are predicted as positives. These measures are trivially maximized by not predicting anything, or predicting everything, respectively, as positive. Because of this these measures often are used together. There are different ways to combine these measures as described by the next 4 metrics. Precision

Recall F-score: for a given threshold, the F-score is the harmonic mean of the precision and recall at that threshold. Precision at a recall level: as the name suggests, set the threshold such that you have a given recall and the precision for this threshold is computed. Precision recall break-even point: is defined as the precision at the point (threshold value) where precision and recall are equal.

Average precision: usually is computed as the average of the precisions at eleven evenly spaced recall levels.

**4.2 Receiver Operating Curves (ROC)**
Central to constructing, deploying, and using classification models is the question of model performance assessment. Traditionally this is accomplished by using metrics derived from the confusion matrix or contingency table. However, it has been recognized that (a) a scalar is a poor summary for the performance of a model in particular when deploying non-parametric models such as artificial neural networks or decision trees and (b) some performance metrics derived from the confusion matrix are sensitive to data anomalies such as class skew. Recently it has been observed that Receiver Operating Characteristic (ROC) curves [1,2] visually convey the same information as the confusion matrix in a much more intuitive and robust fashion. ROC curves are two-dimensional graphs that visually depict the performance and performance trade-off of a classification model. ROC curves were originally designed as tools in communication theory to visually determine optimal operating points for signal discriminators.

Evaluation Measures for Data Mining Tasks False Positive Rate = ROC graphs are constructed by plotting the true positive rate against the false positive. A number of regions of interest in a ROC graph can be identified. The diagonal line from the bottom left corner to the top right corner denotes random classifier performance, that is, a classification model mapped onto this line produces as many false positive responses as it produces true positive responses. To the left bottom of the random performance line there is the conservative performance region. Classifiers in this region commit few false positive errors. In the extreme case, denoted by point in the bottom left corner, a conservative classification model will classify all instances as negative. In this way it will not commit any false positives but it will also not produce any true positives.

The region of classifiers with liberal performance occupies ROC graphs are constructed by plotting the true positive rate against the false positive. A number of regions of interest can be identified in a ROC graph. The diagonal line from the bottom left corner to the top right corner denotes random classifier performance, that is, a classification model mapped onto this line produces as many false positive responses as it produces true positive responses. To the left bottom of the random performance line is the conservative performance region. Classifiers in this region commit few false positive errors. In the extreme case, denoted by point in the bottom left corner, a conservative classification model will classify all instances as negative.

In this way it will not commit any false positives but it will also not produce any true positives. The region of classifiers with liberal performance occupies the top of the graph. These classifiers have a good true positive rate but also commit substantial numbers of false positive errors. Again, in the extreme case denoted by the point in the top right corner, we have classification models that classify every instance as positive. In that way, the classifier will not miss any true positives but it will also commit a very large number of false positives. Classifiers that fall in the region to the right of the random performance line have a performance worse than random performance, that is, they consistently produce more false positive responses than true positive responses. However, because ROC graphs are symmetric along the random performance line, inverting the responses of a classifier in the "worse than random performance" region will turn it into a well performing classifier in one of the regions above the random performance line. Finally, the point in the top left corner denotes perfect classification: 100% true positive rate and 0% false positive rate.

**Table 1: Comparing performance of Reduced and Original set of attributes with Tree-Classifiers**

| Trees | No .of Attributes | Accuracy | ROC Curve |
|---|---|---|---|
| Random Forest | 280(Original No of Attribute) | 64.0177 | 0.8138 |
| | 25(Reduced Attribute) | 73.2892 | 0.8913 |
| J48 | 280(Original No of Attribute) | 66.4459 | 0.7576 |
| | 25(Reduced Attribute) | 69.9779 | 0.7815 |
| Random Tree | 280(Original No of Attribute) | 53.5398 | 0.6648 |
| | 25(Reduced Attribute) | 65.4867 | 0.7941 |
| LMT | 280(Original No of Attribute) | 69.0265 | 0.7851 |
| | 25(Reduced Attribute) | 73.0088 | 0.9063 |

**Table 2: Comparing performance of Reduced and Original set of attributes with Rules-Classifiers**

| Rules | No .of Attributes | Accuracy | ROC Curve |
|---|---|---|---|
| J-Rip | 280(Original No of Attribute) | 71.4602 | 0.8147 |
| | 25(Reduced Attribute) | 71.6814 | 0.8198 |
| One R | 280(Original No of Attribute) | 58.4071 | 0.5779 |
| | 25(Reduced Attribute) | 58.4071 | 0.5779 |
| PART | 280(Original No of Attribute) | 64.1593 | 0.7731 |
| | 25(Reduced Attribute) | 65.0442 | 0.7731 |
| Zero R | 280(Original No of Attribute) | 54.2035 | 0.4898 |
| | 25(Reduced Attribute) | 54.2035 | 0.4898 |

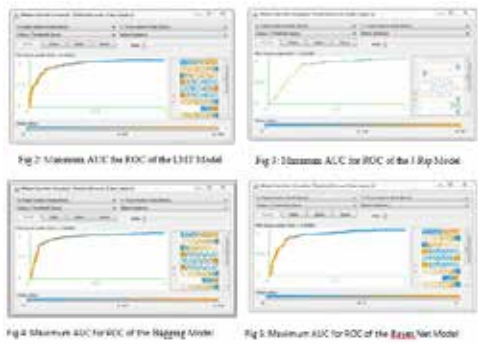
**Table 3: Comparing performance of Reduced and Original set of attributes with Meta-Classifiers**

| Meta | No .of Attributes | Accuracy | ROC Curve |
|---|---|---|---|
| Ada boost | 280(Original No of Attribute) | 55.3097 | 0.6366 |
| | 25(Reduced Attribute) | 55.3097 | 0.6366 |
| Bagging | 280(Original No of Attribute) | 73.2301 | 0.8922 |
| | 25(Reduced Attribute) | 75.6637 | 0.8968 |
| Dagging | 280(Original No of Attribute) | 60.8407 | 0.8244 |
| | 25(Reduced Attribute) | 62.6106 | 0.7883 |
| Multiboost | 280(Original No of Attribute) | 55.531 | 0.6388 |
| | 25(Reduced Attribute) | 55.531 | 0.6388 |

**Table 4 Comparing performance of Reduced and Original set of attributes with Bayes -Classifiers**

| Bayes | No .of Attributes | Accuracy | ROC Curve |
|---|---|---|---|
| Bayes Net | 280(Original No of Attribute) | 71.6814 | 0.8756 |
| | 25(Reduced Attribute) | 75.6637 | 0.8908 |
| Naïve Bayes | 280(Original No of Attribute) | 61.5044 | 0.8313 |
| | 25(Reduced Attribute) | 69.9115 | 0.8828 |
| Naïve Bayes updatable | 280(Original No of Attribute) | 61.5044 | 0.8313 |
| | 25(Reduced Attribute) | 69.9115 | 0.8628 |

he above four tables 1-4 show the accuracy and corresponding area under the ROC curve. It enables us to observe the fluctuations and correlations of these two measures. Based on this we present the final results.



**5.1 Results:**
The experiment clearly shows the value of area under the ROC curve maximum around 0.9063 and the maximum accuracy 75.667% iterating over the familiar set of classifiers.

**5.2 Conclusion:**
Evaluations of classifiers through measures like various errors and ratios sometimes do not guarantee the success in selecting the optimal classifier. We prove here the area under ROC curves yield fruitful combination in the context of arrhythmia dataset and the same argument can be extended to various datasets.

**ACKNOWLEDGEMENTS:**
The authors would like to thank the management of Bharath University for the support and encouragement for this research work.

**REFERENCES**

1. Hanley JA, McNeil BJ. The meaning and use of the area under the Receiver Operating Characteristic (ROC) curve. Radiology 1982 143 29-36. | 2. Metz CE. Semin, Basic principles of ROC analysis, Nuclear Med, 1978, VIII(4) 283-298. | 3. Companion slides for the text by Dr..H.Dunham, Data Mining, Introductory and Advanced Topics, Prentice Hall, 2002. | 4. Uci reprository https://archive.ics.uci.edu/ml/datasets.html downloaded on 13th Aug 2014. | 5. A.Kumaravel, Pradeepa.R, Efficientmolecule reduction for drug design by intelligent search methods Int J Pharm Bio Sci 2013 Apr; 4(2): (B) 1023 – 1029 | 6.Toward optimal feature selection using ranking methods and classification Algorithms Jasmina Novaković, Perica Strbac, Dusan Bulatović Yugoslav Journal of Operations Research 2011, Number 1, 119-135 | 7. https://www.waset.org/journals/waset/v68/v68-21.pdf world academy of science, engineering and technology, 2012.(Accessed on 9th july 2014) | 8 A.Kumaravel and D.Udhayakumarapandian, Consruction Of Meta Classifiers For Apple Scab Infections Int J Pharm Bio Sci 2013 Oct; 4(4): (B) 1207 – 1213. | 9. A.Gelman, Y. S. Su, M.Yajima, J. Hill, M.Pittau, J. Kerman, and T. Zheng, "arm:Data Analysis Using Regression and Multilevel/Hierarchical Models," R package version 1.5-02.://CRAN.Rproject.org/package=arm,2012. | 10. A.Kumaravel and D.Udhayakumarapandian, A Novel Subset Selection For Classification Of Diabetes Dataset By Iterative Methods Int J Pharm Bio Sci 2014 July ; 5 (3) : (B) 1 - 8 | 11. L. Breiman, " RandomForests,"inMachineLearning, vol. 45, pp. 5-32, 2001. | 12. https://archive.ics.uci.edu/ml/datasets/Arrhythmia (Accessed on 3rd july 2014) | 13. Dietterich, T. G., Jain, A., Lathrop, R.,Lozano-Perez, T. (1994). A comparison of dynamic reposing and tangent distance for drug activity prediction.Advances in Neural Information Processing Systems, 6. SanMateo, CA: Morgan Kaufmann. 216--223. | 14. A.Stensvand, T. Amundsen, L. Semb, D.M.Gadoury, and R.C. Seem. 1997.Ascospore release and infection of apple leaves by conidia and ascospores of Venturia inaequalis at low temperatures.Phytopathology 87:1046-1053. | 15. Biggs, A. R. "Apple Scab" in Jones, A. L.,and Aldwinckle, H. S. (editors).1990.Compendium of apple and pear diseases.APS Press, St. Paul, MN. | 16. http://weka.sourceforge.net/doc/weka/ (Downloaded on 10th Aug 2014. | 17. Sourceboutwekahttp://www.cs.waikato.ac.nz/ml/weka/downloaded on 10 Aug 14 | 18.Comparative study of attribute selection using gain ratio and correlation based feature selection Asha Gowda Karegowda1, A. S. Manjunath2 & M.A.Jayaram3 International Journal of Information Technology and Knowledge Management July-December 2010, Volume 2, No. 2, pp. 271-277. | 19.Toward optimal feature selection using ranking methods and classification Algorithms Jasmina Novaković, Perica Strbac, Dusan Bulatović Yugoslav Journal of Operations Research 2011, Number 1, 119-135 |