Accuracy-Rejection Curves (ARCs) for Comparison of Classification Methods with Reject Option

Malik-Sajjad Nadeem[1,2,5,⋆], Jean-Daniel Zucker[2,3,6], Blaise Hanczar[4]

(1)LIM&Bio, UFR de Sante, Medecine et Biologie Humaine (SMBH) - Leonard de Vinci, Universite Paris 13, 74 rue Marcel Cachin, 93017 Bobigny Cedex, FRANCE.
(2)INSERM U872 Equipe 7, Centre Recherches des Cordeliers, 15 rue de l'Ecole de Medecine, Universite Paris 6, 75005 Paris, FRANCE.
(3)UPMC University Paris 06, UMRS 872, NUTRIOMIQUE, CRC, 75006, Paris, FRANCE.
(4)CRIP5, Universite Paris Descartes, 45 rue des Saint-Peres, 75006 Paris, FRANCE.
(5)Department of Computer Sciences & IT, University of Azad Jammu & Kashmir Muzaffarabad, 13100 Muzaffarabad, Azad Jammu & Kashmir, PAKISTAN.
(6)Institut de Recherche pour le Developpement(IRD) IRD, UMI 209, UMMISCO, IRD France Nord, F-93143, Bondy, FRANCE.
msajjadnadeem@gmail.com
hanczar_blaise@yahoo.fr
jdzucker@gmail.com

Abstract. Data extracted from microarray chips are considered to be an important source for providing insight about different diseases. Several studies (including ROC graphs) based on microarray data have been reported for comparison of supervised machine learning approaches. These comparisons rely on the classification schemes where all the samples are discriminated no matter how much the classifier is confident on classification. In health care domain, it is better to abstain when the confidence on classification is not sufficiently high enough instead of classifying all examples with pretty low confidence. In our approach, we proposed to compare the classifiers' performance in the scenario of reject option by considering different reject areas. Based on Accuracy-Rejection tradeoff we proposed four types of Accuracy-Rejection Curves (ARCs). Empirical results based on pure artificial data and data synthesized from real patients' data for binary classification problem depict the efficacy of proposed comparison.

Key words: Classifier comparison; reject option; microarray;

1 Introduction

Microarray classification is a topic of great interest in now-a-days medical and bioinformatics research. Microarrays simultaneously measure the mRNA expres-

⋆ Corresponding author. E-mail addresses: msajjadnadeem@gmail.com, msajjadnadeem@yahoo.com. I am a Ph.D candidate in the filed of biomedical informatics.

sion level of thousands of genes in a cell mixture at certain times and in different environmental conditions. One of the main characteristic of this kind of data is the huge disproportion between the number of examples (generally 10 to 100 microarrays by experiment) and number of features (several thousands of genes). Microarrays are used in many fields of medical research. Among the most prominent and useful applications is the prediction of a biological parameters based on the gene-expression profile. For example, by comparing the expression profiles of different tissue types we can predict different types of tumors with different outcomes and hence assist in the selection of a therapeutic treatment [6, 1, 17].

A large number of methods, from machine learning, have been successfully applied to classify microarrays, diagonal linear discriminant analysis (DLDA) and k-nearest neighbors [6], Support Vector Machine [8], Random Forests [2]. Even if these methods produce classifiers with a good accuracy, very often they are still insufficiently accurate to be used in medical applications. A diagnostic or a choice of therapeutic strategy must be based on a very high confidence classifier. While classifying, if the performance is not up to a desired limit, it is often helpful to introduce a reject option in order to increase the classification accuracy. The principle is to refrain from taking decision for samples whose decision is less confident in order to reduce error probabilities so as to meet the performance level. The performance of a classifier with reject option is based on both its accuracy and rejection rate.

Considering reject option is of great importance in practice, especially in the case of medical application where it would be interesting to compare classifiers with reject option in order to determine the best one. According to our knowledge, there is no comparison study including classifiers with reject option in the literature. A general assumption is that the comparison of classifiers is the same with and without reject option. In this paper, we test this assumption and show that it is wrong. We point out that the rejection has different impact on the accuracy of different classifiers, and the best classifier depends also on the quantity of rejection. Our experiments present comparisons of different classifiers with and without reject option and investigate the effects of reject option on a prediction model's performance using diverse synthetic data sets. Our results show that the use of reject option considerably enhances the performance of classifiers in terms of accuracy.

The rest of the paper is organized as follows: in section 2, the theory and concepts of rejection are presented. Section 3 consists of the theory of comparison of classifiers with reject option. Section 4 describes the synthetic data sets generation and experimental design. Section 5 gives insights of results and discussion and finally paper ends with conclusion.

2 Rejection

Chow [3] introduces the concept of reject option in this following way. Consider a classification problem with two classes, $C = \{1, -1\}$, where an example is

characterized by a feature vector $x \epsilon R_p$ and a label $y \epsilon C$. The posterior probability is defined by the Bayes's formula:

$$p(C_i|x) = \frac{p(x|C_i)p(C_i)}{p(x)} = \frac{p(x|C_i)p(C_i)}{\sum_{i=1}^{2} p(x|C_i)p(C_i)}, \tag{1}$$

where $p(C_i)$ is the prior probability of class C_i, $p(x|C_i)$ is the conditional probability of x given C_i and $p(x)$ is the probability of x. A classifier is a function $f : R^p -> C$ which divides the feature space into two regions, R_1 , R_2, one for each predicted class, such that $x \epsilon Ri$ means that $f(x) = Ci$. The performance of a classifier is measured by its error rate,

$$\epsilon[f] = p(f(x) \neq y) = \sum_{i=1}^{2} \int_{R_i} \sum_{j=1;j\neq i}^{2} p(x|C_j)p(C_j)dx \tag{2}$$

which is the probability of making an incorrect classification. The accuracy of a classifier is defined as the probability of making a correct decision.

$$a[f] = 1 - \epsilon[f], \tag{3}$$

The classifier minimizing the error is called the Bayes classifier. It predicts the class having the highest posterior probability:

$$f_{Bayes}(x) = argmax_{C_i}(p(C_i|x)), \tag{4}$$

It is not possible to obtain a better accuracy than with the Bayes classifier given the ture posterior probabilities are known.

If the accuracy of the Bayes classifier is not sufficient for the task at hand, then one can take the approach not to classify all examples, but only those for which the posterior probability is sufficiently high. Based on this principle, Chow [4] presented an optimal classifier with reject option. A rejection region R_{reject} is defined in the feature space and all examples belonging to this region are rejected by the classifier. An example x is accepted only if the probability that x belongs to C_i is higher than or equal to a given probability threshold t:

$$f(x) = \begin{cases} argmax_{C_i}(p(C_i|x)) \;\; if \;\; max_{C_i}(p(C_i|x)) >= t \\ reject \;\; if \;\; p(C_i|x) < \forall_i \end{cases} \tag{5}$$

The classifier rejects an example if the prediction is not sufficiently reliable. The rejection rate is the probability that the classifier rejects the example,

$$p(reject) = \int_{R_{reject}} p(x)dx = p(max(p(C_i|x)) <= t), \tag{6}$$

In classification with reject option, we can define two types of error. The error, $\epsilon[f]$, is the probability of making an incorrect classification. The conditional error,

$$\epsilon^{cond}[f] = p(f(x) \neq y|accept) \tag{7}$$

is the probability of making an incorrect classification, given the classifier has accepted the example. The acceptance rate is the probability that the classifier accepts an example. So, we have the following basic properties:

$$p(accept) + p(reject) = 1 \tag{8}$$

$$p(f(x) = y) + p(f(x) \neq y) + p(reject) = 1 \tag{9}$$

$$p(f(x) = y|accept) + p(f(x) \neq y|accept) = 1 \tag{10}$$

There is a general relation between the error and rejection rate: According to Chow [4] the error rate decreases monotonically while the rejection rate increases. Based on this relation, Chow proposes an optimal error versus reject tradeoff.

In Chow's theory, an optimal classifier can be found only if the true posterior probabilities are known. This is rarely the case in practice. Fumera et al. [7] show that Chow's rule does not perform well if a significant error in probability estimation is present. In this case, they claim that defining different thresholds for each class gives better results. The classification rule becomes:

$$f(x) = \begin{cases} argmax_{C_i}(p(C_i|x)) \;\; if \;\; max_{C_i}(p(C_i|x)) >= t \\ reject \;\; if \;\; p(C_i|x) < t_i \forall_i \end{cases} \tag{11}$$

Although this kind of classifier is popular in the machine learning community, it is rarely used in microarray-based classification.

In classifier with rejection option, the key parameters are the thresholds t_i that define the reject areas. Several strategies have been proposed to find an optimal reject rule. Landgrebe et al. [13] define 3D ROC curves for a classifier, where the axes represent the true positive rate, the false positive rate rejected by the classifier and the false positive rate accepted by the classifier. The optimal thresholds are chosen by maximizing the volume under the 2D surface. Dubuisson and Masson [5] propose a rejection rule for problems where the classes are not well known. They include two rejection options: an ambiguity reject when an example is situated in the area between several classes and a distance reject for examples far from the samples of known classes. The approach presented by Hanczar and Dougherty [9] is to control the conditional error rate of the classifier and is applied to microarray based classification.

In our work, we do not deal with the problem of optimal tradeoff between error and rejection. In our approach, we used different rejection areas and computed resulting accuracies. We varied the size of rejection window from 0% to 100% by an increment of 0.2% resulting in 500 rejection windows. To represent the results we plotted the rejection windows against obtained accuracies.

3 Comparing Classifiers with Reject Option

The performances of classifiers are measured by their accuracy to predict the true class. This accuracy is estimated by re-sampling procedure like cross-validation

or bootstrap. A natural question is which is the best classifier for microarray based classification? Unfortunately the answer is not easy. Several comparative studies have been published. Man et al. [15] claim that SVM and PLS-DA have the best accuracy. Dudoit et al. [6] show that simple methods like diagonal linear discriminant analysis DLDA and k-nearest neighbors produce good results, whereas Statnikov et al. [16] conclude the superiority of SVM. More confident conclusion is probably given by both Lee et al. [14] and Huang et al. [11], there is no classifier uniformly better than the other. Actually, the performance of a classifier depends heavily on the data. For each classification task, a comparison study should be done to determine the best classifier.

In case of classification with reject option, the accuracy depends on the error rate. More we reject, better the accuracy. In this paper we propose a classifiers' comparison method in the scenario of reject option. The idea is to watch the accuracies of the classifiers in the function of their reject rate. Based on this idea we define 4 different situations:

1. Case1: A classifier (say Cls_1) initially performs verse than another classifier (say Cls_2). By opting reject option, Cls_1 outperforms Cls_2. Name this crossing over as $T1$ type Accuracy-Rejection Curve(ARC).
2. Case2: Without selecting to reject or rejeting to some extent both the classifiers Cls_1 and Cls_2 perform approximatley same but with more and more rejection, one of the classifier increases its performace more rapidly than other. Call this diversion as $T2$ type ARC.
3. Case3: If Cls_1 and Cls_2 are very much distinct in their performance without rejection but the reject option does affect identically to both of them. Name these curves as $T3$ type ARCs.
4. Case4: If Cls_1 and Cls_2 are very much distinct in their performance without rejection then consideration of reject option to certain limit makes the two classifiers identical in the performance. We will call this merge as $T4$ type ARC.

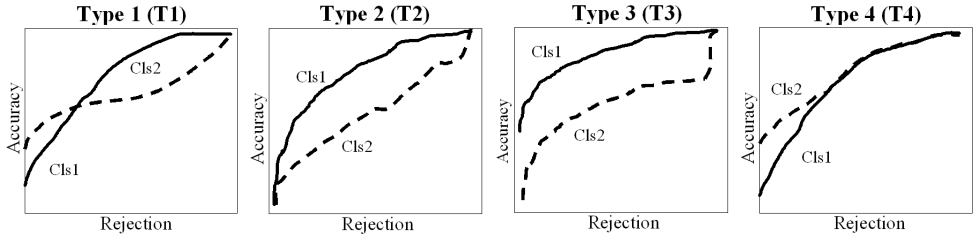

Fig. 1. Illustration of the 4 cases of possible Accuracy-Rejection Curves (ARCs).

Receiver operating characteristics (ROC) graphs [18, 19] are two-dimensional graphs in which true positive rate is plotted on the Y axis and false positive rate is plotted on the X axis. A ROC graph depicts relative tradeoffs between benefits (true positives) and costs (false positives) without considering reject

option. In this paper we propose that the performances of classifiers can also be represented by 2-dimensional Accuracy-Rejection Curves (ARCs) where the axes are their accuracies and rejection rates. Figure 1 illustrates the 4 different cases that we defined.

4 Experimentation

In this section we will first discuss the two types of data generation mechanisms: First, pure artificial data; Second, synthetic data based on patients' data from published works i.e. Alon et al., Shipp et al., and Golub et al. Then experiment design is discussed.

4.1 Data

We have based our experimentations on artificial data. The main interest of using synthetic is the possibility to compute accurately the error and rejection of each classifier. In high dimensional settings, where the number of samples remains very small as compared to the number of attributes of the samples, if we use real data, we have to use resampling methods like cross-validation or bootstrap in order to estimate the error rate. But, it have been shown that these methods are not reliable in high dimension [12, 1]. The main reasons of this problem come from the high variance of the estimators and the lack of correlation between true and estimated errors [10].

We consider 2 class classification problems where each class follows a Gaussian distribution. The classes are equally likely and the class-conditional densities are defined as: $N(\mu_1; \sigma_1 \sum)$, and $N(\mu_2; \sigma_2 \sum)$, where $\mu_1 = (-1, -1, -1, ...)$, and $\mu_2 = (1, 1, 1, ...)$. The covariance matrix of each class is defined by $\sigma_i \sum$ where $\sum$ has a block structure. That means, we define in $\sum$ B blocks, each feature is associated to a unique block. The correlation between two features in the same block is ρ, the correlation between two features from different blocks is 0. In varying the parameters of our model $(\mu_1, \mu_2, B, \rho, \sigma_1, \sigma_2)$ we can construct different kinds of classification problems (linear or non linear, with or without correlated features). For pure artificial data, we choose the parameters of the model. For artificial data from real data, the parameters of the model are estimated from real data using EM algorithm. We have used three real microarray datasets from cancer: colon (Alon et al.), lymphoid malignancy (Shipp et al.) and leukemia (Golub et al.).

For each classification problem, we generate data with 20 features, called noise free features. In real microarrays most of the genes are irrelevant for the classification task in hand. So to have a more realistic aspect, 380 irrelevant or noise features $d_{irrF} = 380$ are added to artificial datasets. A noise feature follows the same Gaussian distribution for the two classes $N(\mu; \sigma)$. The generated data contain N examples, 400 features where 380 are noise features and 20 are noise free features.

The different settings and description of the parameters can be found on the companion website `http://bioinfo.nutriomics.org/~sajjad/ARC/`.

4.2 Experimental Design:

We used following decorum in our experimental design.

1. Generate class-labelled train data n_{tr}containing 50, 100 or 200 examples and a total of $D = D_{nf} + D_n$ features.
2. Generate test data n_{ts}containing 10000 examples and a total of $D = D_{nf} + D_n$ features.
3. Find 20 or 40 best features by using t-test feature selection method on DT_r and reduce train data by selecting only $d_{sel} = 20$ best features from train data set.
4. Reduce test data by driving the same best features from test dataset DT_s.
5. Apply a classification rule to build a classifier Cls from DT_r according to most widely used classification rules for microarray analysis including Support Vector Machine Linear kernel (SVM-Linear); Support Vector Machine Radial kernel (SVM-Radial); Linear Discriminant Analysis (LDA); Quadratic Discriminant Analysis (QDA); random Forest (RF).
6. Compute true error rate/rejection rates of the underlying model.
7. Repeat step 6 for all sizes of rejection windows $R_{win} = \{0.002, 0.004, 0.006, ...100.000\}$
8. All steps 1-7 iterated 100 times.
9. Final result is averaged from all iterations.

We randomly generated 100 different data sets in each case and then these 100 replications are used for classification using classification rule (SVM-Linear, SVM-Radial, LDA, QDA, etc).

Our experimentations are based on two kinds of data: pure artificial data generated from Gaussian models, synthetic data generated from real microarray data and Gaussian models from microarray studies: colon cancer data (Alon et al.), lymphoid malignancy (Shipp et al.) and acute myeloid leukemia" (AML) and acute lymphoblastic leukemia (ALL) (Golub et al.).

5 Results and Discussion

The experiments use both pure synthetic data and synthetic data based on real microarray patient data. The experiments on synthetic data permit very accurate estimations of the error and rejection rates.

In each of the following figures we plot average rejection versus average accuracy for all classification rules $R = 5$ and for one of the data sets. Here we present some typical results while leaving the complete results on the companion website `http://bioinfo.nutriomics.org/~sajjad/ARC/`.

In the plots solid lines represent the rejection accuracy curve of SVM with Radial kernel, dashed lines show SVM with Linear kernel, dotted lines are of LDA, dashed-dotted lines are of QDA, and filled-circle lines represent RF.

We obtained Figure:2A by simulating the data where the problem is linear with non-correlated features; 1 Gaussian per class, train data contains 50 examples and test data have 10000 examples. Here we notice that SVM-Radial

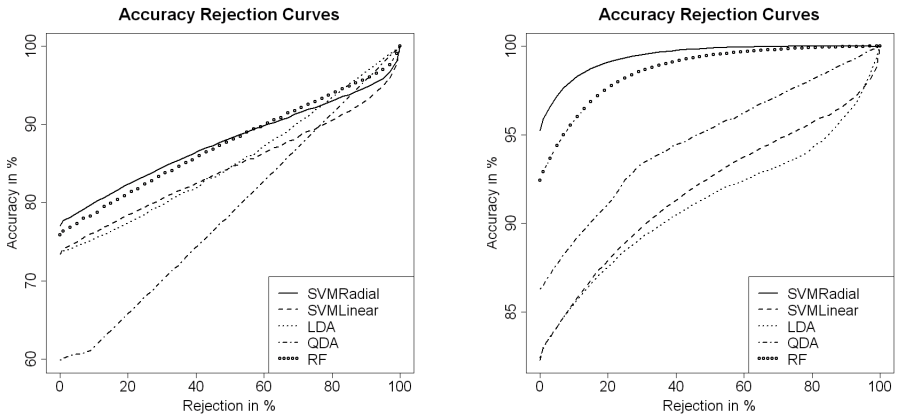

Fig. 2. A(left):Rejection verses Accuracy curve on linear, non-correlated data with 1 Gaussian per class where train set = 50 examples and test set= 10000 examples. B(right):Rejection verses Accuracy curve on non-linear, correlated data with 1 Gaussian per class where train dataset =100 examples and test dataset = 10000 examples.

without rejection (0% rejection) produces around 87% accuracy and RF without rejection (0% rejection) results 85% accuracy. By opting to reject around 50% RF becomes better classifier than SVM-Radial. Also an interesting point here in Figure:2A is that with 45% rejection rate both LDA and SVM-Linear behave similarly as for as accuracy is concerned. But after 45% rejection, LDA outperforms SVM-Linear. Figure:2A depicts that initially without rejection LDA and SVM-Linear have almost identical accuracies. While rejecting on 3% and more samples SVM-Linear performs better than LDA.

Figure:2B results from data possessing non-linear data with correlated features; 1 Gaussian per class where we have 100 examples as train data and 10000 examples as test data. Here, LDA and SVM-Linear produce similar accuracies starting from without rejection (0% rejection) to 18% rejection but from 19% rejection SVM-Linear starts performing much better than LDA.

Figure:3 contains the plot of data synthesized from colon cancer patient dataset with 5 Gaussians per class and with train data equal to 200 examples and test data consist of 10000 examples. In Figure:3, while comparing LDA and SVM-Radial, we found the situation where curves of LDA and SVM-Radial cut each other making LDA better than SVM-Radial. Also this figure does show that on evaluating the performances of QDA and SVM-Radial, QDA outperforms SVM-Radial on having reject option.

Each of our result reflects that as we reject more, we get more and more accuracy. Not all the classification rules used here respond identically to reject option. Our study shows that some respond more quickly and we get more accurate classification than that of others.

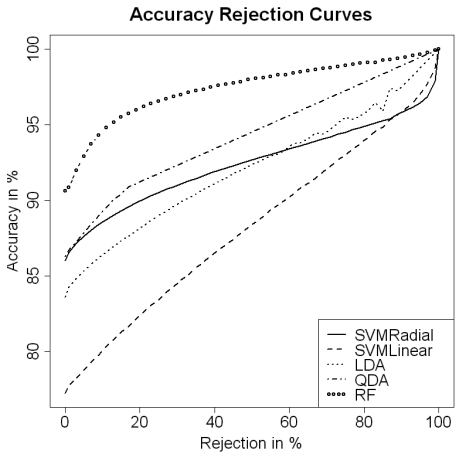

Fig. 3. Rejection verses Accuracy curve on Synthetic data from colon cancer patient dataset with 5 Gaussians per class where train dataset = 200 examples and test dataset= 10000 examples.

Table 1. Summary of Curves obtained in experimentation based on Pure synthetic data where $T1$, $T2$, $T3$ are cases illustrated in Figure:1

		No. of Gaussians			
		1		2	
Block size	Train Samples	$\sigma_1 = \sigma_2$	$\sigma_2 = \sigma_1/2$	$\sigma_1 = \sigma_2$	$\sigma_2 = \sigma_1/2$
1	50	T1, T2	T3	T2	T2
	100	T1	T2	T1, T2	T1
	200	T1, T2	T2	T1, T2	T1, T2
2	50	T1, T2	T2, T3	T1, T2	T2
	100	T1, T2	T3	T1, T2	T1
	200	T1	T2	T1, T2	T2
4	50	T1, T2	T1, T2	T1, T3	T2
	100	T1, T2	T2	T1, T2	T1
	200	T2	T2, T3	T1, T2	T2, T3
5	50	T1, T2	T2	T1, T2	T1
	100	T1, T2	T2	T1, T2	T2
	200	T1, T2	T2, T3	T1, T2	T2
10	50	T2	T2	T1, T3	T1, T2
	100	T1, T2	T2	T1, T2	T2
	200	T1, T2	T1, T3	T1, T2	T2
No Block (Non-Correl)	50	T1, T2	T2, T3	T1, T2	T2
	100	T3	T2	T1, T2	T1
	200	T2, T3	T2	T2	T1, T2

By analysing three presented figures in this section, we have interesting results where different classification rules respond differently different at different rejection rates. Empirical results show that most of the times one or more classification rules outperform the other(s).

On the basis of above presented results and discussion we can have three out of four types of ARCs proposed in the section Comparing Classifiers with Reject Option.

The identification of these types of curves is advantageous in several ways. First: during the selection of suitable classifier for a classification problem if $T1$ curves are available during selection process then the classifier which outperforms the others should be given priority. Second: In case of $T2$ curves, we may reject upto desired limit and then the classifier with high performance may be utilized. Third: When $T3$ curves are there then at a given rejection extent, the classifier with higher performance should be selected for use for that specific dataset for which the comparison was made.

Table 2. Curves obtained in experimentation based on synthetic data from real microarray data.

		No. of Gaussians		
Data	Train Samples	1	2	5
Golub	100	T3	T3	T3
	200	T2	T3	T3
Alon	100	T2	T2	T1
	200	T1, T2	T3	T1, T2
Shipp	100	T3	T3	*xxx
	200	T3	T3	*xxx

*xxx = Don't have results.

In Tables 1 and 2 we summarize our all the 90 experiments based on the above mentioned categories of curves.

While experimenting with pure artificial data we noticed that in 72 experiments we have 40 times the situation when one or more classifier outperforms the other (by crossing over of curves i.e. category $T1$). Also an interesting point is that we have 59 situations where without or with some rejection, two or more classifiers perform almost identically. But with more or less rejection, one of the classifier improves its prediction capability more promptly than the other (category $T2$). Here we have only 12 cases where $T3$ type curves are present in the results. Please refer Table 1.

In total of 90 experiments we found 43 times when one or more classifier outperforms the other through $T1$ curves. We also experienced 64 $T2$ curves. Please refer Table 1 and 2. The presence of more $T2$ and $T3$ curves reflects that the use of reject option in comparison of classifiers is extremely fruitful and in most of the cases aids in more optimal classifier selection. The presence of 22 $T3$

shows that sometimes rejection does not affect very much on the performance of the classifiers and there remains no significant change in the performances of two classifiers as compared to each other.

6 Conclusion

We have studied classifiers performance with reject option. The accuracy of a classifier depends highly on its rejection rate. We introduce the accuracy - rejection rate curves (ARCs) that allow to accurately represent the performance of classifiers. We see that it is necessary to watch both accuracy and rejection rate to compare two classifiers. On the basis of our empirical results we categorize the classifiers comparison into four types. First: the crossing over $T1$ type of ARCs where one of the classifier starts performing better than the other using reject option as in Figures:2A and 3. Second: $T2$ type of ARCs in which one of the two classifiers boosts its performance more rapidly than the other. For example Figure:3. Third: $T3$ ARCs are produced when there is no significant change in the performances of the two classifiers with reject option as compared to each other as in Figure:2A and B, and Figure:3. Fourth: sometimes two classifiers without rejection give different accuracies but with some rejection both of them start producing almost the same accuracies($T4$ type ARCs).

We made classifiers comparisons on a high number of experiment based on artificial data for 500 different reject areas ranging from 0.2% to 100% reject rates. We use different settings of parameters for pure synthetic data to construct different kinds of classification problems (linear and non-linear, correlated and non-correlated features with train sets). For synthetic data from real patients' data, model's parameters depend on the real data. In our results the presense of large number of T1 and T2 types of ARCs shows that ARCs are of interest while comparing classifiers' performances. Small number of T3 type ARCs reflects that there are some possibilities of no significinat change in performance of a classifier while using reject option but the chances remain very little. In our results we don't have any clear T4 type of curves but they may be of interest if present while experimenting with real microarray data sets. Obtaining optimal reject area is still an open question and needs further exploration. In function of the rejection rate, the conclusion of the comparison can be different.

7 Acknowledgements

We would like to acknowledge the Government of France, High Education Commission (HEC) Pakistan, Societe Francaise d'Exportation des Ressources Educatives (SFERE) France, and University of Azad Jammu & Kashmir, Muzaffarabad AJ&K, Pakistan for providing support for this research.

References

1. Braga-Neto, U.M., Dougherty, E.R.: Is cross-validation valid for small-sample microarray classification? Bioinformatics. 20(3), 374–380 (2004)

2. Breiman, L.: Random Forests. Machine Learning. 45, 5–32 (2001)
3. Chow, C.K.: An Optimum Character Recognition System using Decision Functions. IRE Trans. on Electronic Computers. EC-6, 247–254 (1957)
4. Chow, C.K.: On Optimum Error and Reject trade-off. IEEE Trans. on Information Theory. IT-16(1), 41–46 (1970)
5. Dubuisson, B., Masson, M.: A Statistical Decision Rule with incomplete Knowledge about Classes. Pattern Recognition. 26(1), 155–165 (1993)
6. Dudoit, S., Fridlyand, J., Speed, T.: Comparison of Discrimination Methods for the Classification of Tumors using Gene Expression Data. Journal of the American Statistical Association. 97, 77–87 (2002)
7. Fumera, G. and Roli, F. and Giacinto, G.: Analysis of Error-Reject Tradeoff in linearly combined Multiple Classifiers. Pattern Recognition. 33(12), 2099–2101 (2000)
8. Furey, T.S., Christianini, N., Duffy, N., Bednarski, D.W., Schummer, M., Haussler, D.: Support Vector Machine Classification and validation of Cancer Tissue Samples using Microarray Expression Data. Bioinformatics. 16(10), 906–914 (2000)
9. Hanczar, B., Dougherty, E.R.: Classification with Reject Option in Gene Expression Data. Bioinformatics. 24 no. 17, 1889–1895 (2008)
10. Hanczar, B., Hua, J., Dougherty, E.R.: Decorrelation of the True and Estimated Classifier Errors in High-Dimensional Settings. EURASIP Journal on Bioinformatics and Systems Biology. 2007, 12 pages (2007)
11. Huang, X., Pan, W., Grindle, S., Han, X., Chen, Y., Park, S.J., Miller, L.W., Hall, J.: A comparative study of Discriminating Human Heart Failure Etiology using Gene Expression profiles. BMC Bioinformatics. 6, 205 (2005)
12. Isaksson, A., Wallman, M., Gransson, H., Gustafsson, M.G.: Cross-validation and Bootstrapping are unreliable in small Sample Classification. Pattern Recognition Letters. 29(14), 1960–1965 (2008)
13. Landgrebe, T.C.W., Tax, D.M.J., Paclk, P., Duin, R.P.W.: The interaction between Classification and Reject Performance for Distance-based Reject-option Classifiers. Pattern Recognition Letters Pages. 27(8), 908–917 (2006)
14. Lee, J.W., Lee, J.B., Park, M., Songa, S.H.: An extensive Comparison of recent Classification tools applied to Microarray Data. Computational Statistics & Data Analysis. 48, 869–885 (2005)
15. Man, M.Z., Dyson, G., Johnson, K., Liao, B.: Evaluating Methods for Classifying Expression Data. Journal of Biopharmaceutical Statistics. 14, 1065–1084 (2004)
16. Statnikov, A., Aliferis, C.F., Tsamardinos, I., Hardin, D., Levy, S.: A comprehensive evaluation of Multicategory Classification methods for Microarray Gene Expression Cancer diagnosis. Bioinformatics. 21, 631–643 (2005)
17. Wang, L., Chu, F., Xie, W.: Accurate Cancer Classification Using Expression of Very Few Genes. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 4(1), 40–53 (2007)
18. Egan, J.P.: Signal detection theory and ROC analysis, Series in Cognition and Perception. Academic Press, New York. (1975)
19. Swets, J.A., Dawes, R.M., Monahan, J.: Better decisions through science. Scientific American. 283, 82-87 (2000)