DOI: 10.15415/mjis.2016.51006

Some Aspects on the Utility of Distance Measures in Comparing Two MROC Curves

SAMEERA G AND VISHNU VARDHAN R

Department of Statistics, Ramanujam School of Mathematical Sciences, Pondicherry University, Puducherry, India

Email: rvvcrr@gmail.com

Received: May 05, 2016| Revised: May 23, 2016| Accepted: August 03, 2016

Published online: September 05, 2016

Abstract Receiver Operating Characteristic (ROC) curve is a widely used and accepted tool to assess the performance of a classifier or procedure. Apart from this, comparing the procedures or ROC curves is also of interest. A multivariate extension of ROC (MROC) curve that considers a linear combination of several markers for classification was proposed by Sameera, Vishnu Vardhan and Sarma [13]. In the present paper, some inferential procedures are given to compare two MROC curves by means of distance measures based on scores of MROC curve and summary measures such as mean vectors and dispersion matrices. Real and Simulated data sets are used to demonstrate the above proposed inferential aspects.

Keywords: Bhattacharya Distance, Mahalanobis Distance, Mean vectors and Dispersion matrices and Multivariate Receiver Operating Characteristic curve.

1. INTRODUCTION

Receiver Operating Characteristic (ROC) curve is a classification tool which is widely used in the field of diagnostic medicine in present day. This tool helps in classifying the individuals/subjects into one of the two groups, healthy and diseased, by identifying a threshold value, which provides maximum accuracy. This model is univariate in nature, i.e., it considers a single marker for classification. However, in real situations it might not always be possible to judge the status of an individual basing on one marker alone. Hence, there is a need to have an ROC model that accommodates multiple markers at hand. A multivariate extension of this model was developed by Su and Liu [8] where data is assumed to follow multivariate normal distribution and two cases were considered, i.e., one with equal covariance matrices and the other

Mathematical Journal of Interdisciplinary Sciences
Vol-5, No-1,
September 2016
pp. 61–80

with unequal covariance matrices. Further, Liu, Schisterman and Zhu [1] and Gao, Xiong, Yan and Zhang [3] provided a modified version of the above model which helps in obtaining better accuracy and maximizes sensitivity. Recently, a Multivariate Receiver Operating Characteristic (MROC) curve was proposed by Sameera, Vishnu Vardhan and Sarma [13] by considering minimax procedure to obtain the linear combination of markers and provides an optimal cutpoint. Further, it is shown that the MROC model and the best linear combination provided by Sameera, Vishnu Vardhan and Sarma [13] provides mathematical ease and better results than the previous models.

Comparing models has become a necessity and gained its prevalence over the years as it helps in identifying a better one among the existing. A brief review about the existing comparison procedures on ROC curves is discussed. Comparison of curves can be done by comparing their accuracy measure Area Under the Curve (AUC) or their intrinsic measures Sensitivity (S_n) and Specificity (S_p). The seminal paper by Greenhouse and Mantel [5] focuses on comparison of curves by comparing their sensitivities at a fixed level of specificity. Gourevitch and Galanter [4] used detectability index to propose a large sample test for testing whether two observed data points belong to the same d' function or not, further extended to k observed data points by Marascuilo [10]. Metz and Kronman [12] proposed a χ^2 statistic to test the equality of curves by testing the parameters of the curve. The comparison of two ROC curves was also done by comparing their area's [6, 7]. Further, McClish [11] proposed F test and studentized range test to compare more than two ROC curves using their AUC's. Vardhan, Sameera, Chandrasekharan and Thulasi [14] proposed testing procedures for comparing MROC curves using AUC and comparing the curves at a particular point.

So far in the literature, the usual method of comparing two curves is done by using the summary measures such as AUC, pAUC, etcetera. In classification, distance measures also play an important role in identifying the similarity between two distributions/ populations. In terms of ROC curve, the distance measure helps in identifying the distance between healthy and diseased populations; larger the distance better the classification. Henceforth, the present paper addresses the utility of distance measures to compare two curves. In the field of classification and separation techniques, Mahalanobis distance is greatly used for explaining the magnitude between two populations using group centroids, depicting the extent of correct classification. The distance measures used in this paper are Mahalanobis and Bhattacharya distances. Conventionally, Mahalanobis distance is usually applied when the populations have equal variances/covariances and Bhattacharya Distance will be taken into account when the variances/ covariances are unequal. The motivation behind considering these two measures is to address the practical situation where these measures are to be used properly, since the variances/ covariances of the populations may not always have equal structures. In other words, the need for two measures is that, when the property of homogeneity of covariance matrices is ignored, the distance value is either overestimated or underestimated thus hiding the actual information. These distances are computed using the scores obtained through the MROC curve and also the mean vectors and covariance matrices of two populations. The proposed procedures are explained with the help of two real datasets (Indian Liver Patients (ILP) dataset and Salmon fish dataset) and simulation studies.

2. PROPOSED METHODOLOGY

Let us suppose X and Y are two multivariate normal random vectors of healthy(H) and diseased(D) populations with mean vectors μ_H, μ_D and covariance matrices Σ_H and Σ_D respectively i.e., $X \sim MVN(\mu_H, \Sigma_H)$ and $Y \sim MVN(\mu_D, \Sigma_D)$. Let n_H and n_D be the sample sizes of X and Y respectively. The expression for MROC curve given by Sameera, Vishnu Vardhan and Sarma [13] is

$$y(x) = \Phi\left(\frac{b'(\mu_D - \mu_H) - \sqrt{b'\Sigma_H b}\,\Phi^{-1}(1-x)}{\sqrt{b'\Sigma_D b}}\right) \tag{1}$$

where x is the false positive rate and b is the vector of linear combination of the markers and is given as $b = \left[t\Sigma_D + (1-t)\Sigma_H\right]^{-1}(\mu_D - \mu_H)$; $0 < t < 1$. The scores for each subject can be obtained as $U_H = b'X$ and for healthy and diseased populations respectively. The obtained scores can then be compared to optimal cut point to identify the status of the subject.

The accuracy measure, Area under the MROC (AUC) curve explains the extent of correct classification and is given as

$$AUC = \Phi\left(\frac{b'(\mu_D - \mu_H)}{\left[b'(\Sigma_D + \Sigma_H)b\right]^{1/2}}\right) \tag{2}$$

The intrinsic measures of MROC curve sensitivity and specificity, the probabilities that diseased and healthy individuals respectively are identified accurately are given as

$$S_n = \Phi\left(\frac{b'\mu_D - c}{\left(b'\Sigma_D b\right)^{1/2}}\right) \tag{3}$$

$$S_p = \Phi\left(\frac{c - b'\mu_H}{\left(b'\Sigma_H b\right)^{1/2}}\right) \tag{4}$$

The main objective of the paper is to provide comparison procedures based on the distance measures namely Mahalanobis and Bhattacharya distances. It is assumed that the scores of MROC curve follow normal distribution; hence, the distance measures are redefined using scores under normal distribution. The detailed methodology of comparison procedures that are proposed using these distance measures are given in subsequent subsections.

2.1 Comparison based on Mahalanobis Distance using Scores

The test scores for each subject is obtained using the linear combination of the MROC model. Further, it is assumed that the obtained scores follow normal distribution. If the test score of healthy and diseased populations tend to have equal variances, Mahalanobis distance can be used. The Mahalanobis distance based on scores is redefined under the setup of MROC model 'i' and is given by

$$D_U^{(i)} = \frac{\bar{U}_D^{(i)} - \bar{U}_H^{(i)}}{\sqrt{S_U^{2\,(i)}}}; i = 1,2 \tag{5}$$

where $\bar{U}_D = \frac{1}{n_D}\sum_{j=1}^{n_D} U_D$, $\bar{U}_H = \frac{1}{n_H}\sum_{j=1}^{n_H} U_H$ and

$$S_U^2 = \frac{\left(n_D - 1\right)\text{var}\left(U_D\right) + \left(n_H - 1\right)\text{var}\left(U_H\right)}{n_D + n_H - 2}.$$

Here, $D^{(1)}$ and $D^{(2)}$ are the Mahalanobis distances of two MROC curves under comparison. The null and alternative hypothesis to compare the distance measures is defined as

$$H_0 : D_U^{(1)} = D_U^{(2)} \sim H_1 : D_U^{(1)} \neq D_U^{(2)}$$

The test statistic used for testing the above hypothesis is

$$Z = \frac{D_U^{(1)} - D_U^{(2)}}{\mathrm{var}\left(D_U^{(1)}\right) + \mathrm{var}\left(D_U^{(2)}\right)} \sim N(0,1) \tag{6}$$

The curve with greater distance measure is said to be a better curve if the Z value obtained in equation (6) is significant.

2.2 Comparison based on Bhattacharya Distance using scores

The practical use of this procedure will come into existence when the scores of both populations have unequal variances. In such cases, Bhattacharya distance gives accurate information as it considers variances of both of populations instead of pooling them. Bhattacharya distance between the scores of healthy and diseased populations of an MROC curve 'i' can be obtained as

$$\begin{aligned} D_{BU}^{(i)} &= \frac{1}{4}\ln\left(\frac{1}{4}\left(\frac{\mathrm{var}\left(U_D^{(i)}\right)}{\mathrm{var}\left(U_H^{(i)}\right)} + \frac{\mathrm{var}\left(U_H^{(i)}\right)}{\mathrm{var}\left(U_D^{(i)}\right)} + 2\right)\right) \\ &+ \frac{1}{4}\left(\frac{\left(\bar{U}_D^{(i)} - \bar{U}_H^{(i)}\right)^2}{\mathrm{var}\left(U_D^{(i)}\right) + \mathrm{var}\left(U_H^{(i)}\right)}\right); i = 1,2 \end{aligned} \tag{7}$$

If $D_{BU}^{(1)}$ and $D_{BU}^{(2)}$ are Bhattacharya distances of two MROC curves to be compared, then the null and alternative hypothesis for testing two curves using their Bhattacharya distances is defined as

$$H_0 : D_{BU}^{(1)} = D_{BU}^{(2)} \sim H_1 : D_{BU}^{(1)} \neq D_{BU}^{(2)}$$

The test statistic used for testing the above null hypothesis against alternative hypothesis is

$$Z = \frac{D_{BU}^{(1)} - D_{BU}^{(2)}}{\operatorname{var}\left(D_{BU}^{(1)}\right) + \operatorname{var}\left(D_{BU}^{(2)}\right)} \sim N(0,1) \tag{8}$$

The variances in equation (8) are obtained using bootstrapping. A better curve is identified, if the Z value is greater than critical value otherwise the curves are said to have similar discriminating ability.

2.3 Comparison based on Mahalanobis Distance using Mean Vectors and Covariance Matrices

The conventional way of obtaining Mahalanobis distance is by using the mean vectors and pooled covariance matrix because the populations are assumed to have equal covariances. Mahalanobis distance between two populations, healthy and diseased for ith MROC curve is given by

$$D^{(i)} = \sqrt{\left(\bar{Y}^{(i)} - \bar{X}^{(i)}\right)' S^{(i)^{-1}} \left(\bar{Y}^{(i)} - \bar{X}^{(i)}\right)}\ ;\ \ i = 1,2 \tag{9}$$

Here, $\bar{X}^{(i)}$, $\bar{Y}^{(i)}$ and $S^{(i)}$ are mean vectors and pooled covariance matrix of ith MROC curve and $D^{(1)}$, $D^{(2)}$ are the Mahalanobis distances of two MROC curves. Then, the null and alternative hypothesis to compare the distance measures is defined as

$$H_0 : D^{(1)} = D^{(2)} \sim H_1 : D^{(1)} \neq D^{(2)}$$

The test statistic used for testing the above hypothesis is

$$Z = \frac{D^{(1)} - D^{(2)}}{\sqrt{Var\left(D^{(1)}\right) + Var\left(D^{(2)}\right)}} \sim N\left(0,1\right) \tag{10}$$

If the obtained Z value is greater than the critical value, the curves under comparison are said to differ from each other. The curve with greater distance has a better discriminating ability.

2.4 Comparison based on Bhattacharya Distance using Mean Vectors and Covariance Matrices

When the populations under study have unequal covariance matrices, it is better to use Bhattacharya distance instead of Mahalanobis distance to avoid loss of information. Bhattacharya distance between two multivariate normal populations; healthy and diseased is given by

$$D_B^{(i)} = \frac{1}{8}\left(\mu_D^{(i)} - \mu_H^{(i)}\right)^T \left(\frac{\Sigma_D^{(i)} + \Sigma_H^{(i)}}{2}\right)^{-1} \left(\mu_D^{(i)} - \mu_H^{(i)}\right) + \frac{1}{2}\ln\left(\frac{\det\left(\frac{\Sigma_D^{(i)} + \Sigma_H^{(i)}}{2}\right)}{\sqrt{\det\Sigma_D^{(i)}\det\Sigma_H^{(i)}}}\right); \quad i = 1,2 \tag{11}$$

Here, $D_B^{(1)}$ and $D_B^{(2)}$ are Bhattacharya distance of MROC curves to be compared. The null and alternative hypothesis for testing two curves using their Bhattacharya distances is defined as

$$H_0 : D_B^{(1)} = D_B^{(2)} \sim H_1 : D_B^{(1)} \neq D_B^{(2)}$$

The test statistic used for testing the above null hypothesis against alternative hypothesis is

$$Z = \frac{D_B^{(1)} - D_B^{(2)}}{\sqrt{Var\left(D_B^{(1)}\right) + Var\left(D_B^{(2)}\right)}} \sim N(0,1) \tag{12}$$

The curve with larger distance is said to be a better curve if the Z value is significant.

3. RESULTS AND DISCUSSIONS

The above inferential procedures are demonstrated with the help of real and simulated datasets. The simulated datasets are considered in such a way that they explain the various cases of MROC curves that are observed in real situations. These simulations are conducted at different sample sizes to observe the behavior of proposed models at small as well as large samples.

3.1 Real data

3.1.1 Indian Liver Patient Dataset

Indian Liver Patient (ILP) dataset [2] consists of ten variables; age, gender, Total Bilirubin (TB), Direct Bilirubin (DB), Total Proteins (TP), Albumin, A/G ratio, SGPT, SGOT and Alkphos. It has 441(75.65%) males and 142(24.35%) females. The comparison is done between males and females to check if parameters involved in identification of liver diseases vary based on gender. The MROC analysis is performed on both males and females using R and the results obtained are tabulated.

Table 1: MROC measures of ILP dataset.

Gender	AUC	Sensitivity	1-Specificity
Males	0.7541	0.6959	0.3041
Females	0.7232	0.6781	0.3219

ScoreMales = 0.0242*Age + 0.0063*TB + 0.2011*DB + 0.0012*Alkphos + 0.0029*SGPT – 0.0004*SGOT + 0.4145*TP – 0.8069*ALB + 0.2592*A/G

ScoreFemales = 0.0009*Age – 0.5375*TB + 1.1341*DB + 0.0012*Alkphos - 0.0056*SGPT + 0.0075*SGOT + 1.1394*TP – 1.8168*ALB + 1.5752*A/G

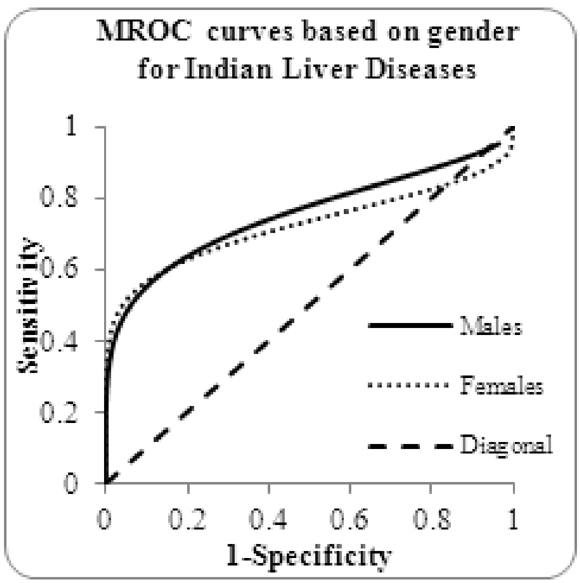

Figure 1: Indian Liver Patient Dataset.

Table 2: Estimates and Z values of Distance Measures on ILP Dataset.

Distance Measures		Males	Females	Z value
Based on Scores	Mahalanobis Distance (D_U)	0.9027	0.8702	0.2162
	Bhattacharya Distance (D_{BU})	0.2616	0.2918	0.2783
Based on Mean vectors & Covariance Matrices	Mahalanobis Distance (D)	0.9146	0.8889	0.1935
	Bhattacharya Distance (D_B)	3.7070	3.4080	0.4641

From table 1, the AUC's of males and females are almost equal indicating that their MROC curves have similar discriminating ability. MROC curves are drawn and it can be seen that they overlap each other. Further, the scores of males and females are computed using the linear combinations reported in Table 1.

The above figure depicts the MROC curves for Males and Females of ILP dataset

In order to test the homogeneity of covariance matrices of two populations and variances obtained from the scores, Box's M test and Levene's F test are used. The Box's M test value for females is 606.946 (0.000*) and for males is 1805.483 (0.000*) indicating that the covariance matrices are unequal. Further, the homogeneity of variances of obtained scores is tested for females and males, whose F values along with significance are 17.434 (0.000*) and 63.167 (0.000*) respectively. This leads to a conclusion that Bhattacharya distance is to be used for comparing the curves. The comparison between males and females is done using the distance measures to verify whether there is any difference in the identification of Liver diseases between them, this is achieved by comparing MROC curves The distances obtained using scores as well as mean vectors and covariance matrices along with their Z values are portrayed in Table 2.

In the case of distance measures using scores, the Mahalanobis measure overestimates the actual distance between the populations than Bhattacharya measure, due to the violation of homogeneity of variances. Further, when the homogeneity of covariance matrices is violated, the actual distance between populations is underestimated when Mahalanobis distance is used thus concluding that Bhattacharya distance should be used to avoid loss of information. The inference procedures indicate that there is no difference between the males and females with respect to identification of liver diseases. The Z values obtained for all the procedures are less than the standard value at 5% level of significance. This indicates that even though males are more prone to liver diseases when compared to females, the parameters used to identify the disease do not differ.

Table 3: MROC measures of ILP dataset.

Salmon	AUC	Sensitivity	1-Specificity
Canadian	0.5753	0.5534	0.4465
Alaskan	0.6402	0.6002	0.3997
ScoreCanadian = - 0.0121*Freshwater - 0.00398*Marine			
ScoreAlaskan = 0.0253*Freshwater + 0.0130*Marine			

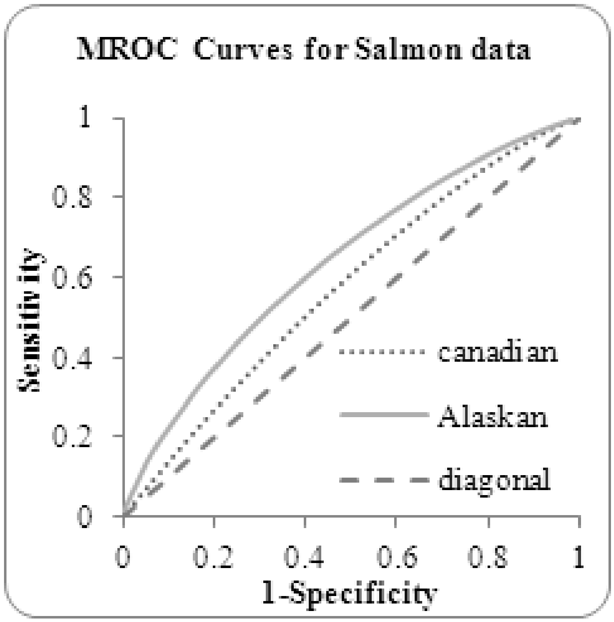

Figure 2: Salmon data.

3.1.2. Salmon Fish data

Salmon Fish data [9] contains data samples of 50 Canadian and Alaskan salmon fishes. This dataset consists of two variables, namely diameter of rings for the first year of growth in freshwater and marine water. Linear combinations are obtained for both Canadian and Alaskan salmon fishes in order to identify the gender of the fish using the diameter of rings observed in fresh and marine water.

Using the given information, gender identification in Canadian salmons can be done to the extent of 57.53% accurately, where as in Alaskan salmons it is observed to have an accuracy of 64.02%. This indicates the fact that good extent of classification is seen in Alaskan salmons than Canadian salmons. To

Table 4: Estimates and Z values of Distance Measures on ILP Dataset.

Distance Measures		Canada	Alaska	Z value
Based on Scores	Mahalanobis Distance (D_U)	0.4292	0.6001	0.4964
	Bhattacharya Distance (D_{BU})	0.0482	0.0646	0.2261
Based on Mean vectors & Covariance Matrices	Mahalanobis Distance (D)	0.4868	0.5805	0.2664
	Bhattacharya Distance (D_B)	0.0809	0.1083	0.3506

depict the same, MROC curves (Figure 2) are drawn for both types of Salmons and it is observed that the MROC curve for Alaskan salmon supersedes the curve of Canadian salmon indicating that the gender of the salmon fishes can be discriminated better using variation in diameter of rings in Alaskan salmons than Canadian salmons.

The above figure depicts the MROC curves for Alaskan and Canadian Salmons

Further, comparisons are obtained for both Canadian and Alaskan salmon fishes to see whether the extent of correct identification of gender is better in one type of fishes when compared to the other. This is addressed using distance measures. The homogeneity of covariance matrices is tested using Box's M test and the values for Alaskan and Canadian salmons are 3.805 (0.304^{NS}) and 1.040 (0.803^{NS}) respectively. The scores obtained through linear combinations are tested for homogeneity of variances using Levene's F Statistic and the results for Alaskan and Canadian salmons are 0.808 (0.373^{NS}) and 0.590 (0.446^{NS}) respectively. These results indicate that the distance measure to be used for comparison is the Mahalanobis distance, since the variances and covariance matrices are observed to be homogneous. The distance measures along with Z values are computed and placed in table 4.

The distance measure values are obtained based on scores as well as mean vectors and covariance matrices. From the results, it is noticed that the distances are underestimated when Bhattacharya measure is used. This means that the exact distance is not computed using Bhattacharya distance when the variances/ covariances are equal. Thus, it is shown that, when the variance/ covariance matrices are homogenous, Mahalanobis measure is to be used to observe the actual distance rather than Bhattacharya measure. Further, Z values observed in the table are not significant for all the four testing procedures indicating that the identification of gender of Canadian salmons is equivalent to that of Alaskan salmons.

Table 5: Mean Vectors and Covariance Matrices of Simulation Studies.

	μD	μH	ΣD	ΣH
1	$\begin{pmatrix}1.5046\\2.1597\\2.5511\end{pmatrix}$	$\begin{pmatrix}1.0337\\1.1615\\1.0921\end{pmatrix}$	$\begin{pmatrix}0.0331 & -0.0092 & 0.0894\\-0.0092 & 0.1821 & -0.0356\\0.0894 & -0.0356 & 0.3136\end{pmatrix}$	$\begin{pmatrix}0.0538 & 0.0129 & 0.1134\\0.0129 & 0.1679 & 0.0354\\0.1134 & 0.0354 & 0.2683\end{pmatrix}$
2	$\begin{pmatrix}1.1482\\1.3232\\1.4534\end{pmatrix}$	$\begin{pmatrix}1.0337\\1.1615\\1.0921\end{pmatrix}$	$\begin{pmatrix}0.0331 & -0.0092 & 0.0894\\-0.0092 & 0.1821 & -0.0356\\0.0894 & -0.0356 & 0.3136\end{pmatrix}$	$\begin{pmatrix}0.0538 & 0.0129 & 0.1134\\0.0129 & 0.1679 & 0.0354\\0.1134 & 0.0354 & 0.2683\end{pmatrix}$
3	$\begin{pmatrix}0.7944\\1.7645\\4.7523\end{pmatrix}$	$\begin{pmatrix}0.6971\\1.2340\\3.9064\end{pmatrix}$	$\begin{pmatrix}0.0060 & 0.0025 & 0.0530\\0.0025 & 0.1140 & 0.1401\\0.0530 & 0.1401 & 0.8590\end{pmatrix}$	$\begin{pmatrix}0.0144 & 0.0158 & 0.1814\\0.0158 & 0.0708 & 0.2744\\0.1814 & 0.2744 & 3.5279\end{pmatrix}$
4	$\begin{pmatrix}0.6420\\1.1027\\2.9357\end{pmatrix}$	$\begin{pmatrix}0.6971\\1.2340\\3.9064\end{pmatrix}$	$\begin{pmatrix}0.0060 & 0.0025 & 0.0530\\0.0025 & 0.1140 & 0.1401\\0.0530 & 0.1401 & 0.8590\end{pmatrix}$	$\begin{pmatrix}0.0144 & 0.0158 & 0.1814\\0.0158 & 0.0708 & 0.2744\\0.1814 & 0.2744 & 3.5279\end{pmatrix}$

3.2 Simulation Study

In this section, the sample size effect on the above proposed methods is demonstrated with the help of simulation studies. Four sets of multivariate normal random numbers are generated with mean vectors and covariance matrices given in table 5. Data is generated at various samples sizes 25, 50, 100, 150, 200 and 300.

The combinations in table 3 are considered in such a way that they represent typical forms of MROC curve. Simulation 1 represents a best curve, simulation 3 represents a better curve and simulations 2 and 4 represent moderate curves. Three combinations of these curves are considered to demonstrate

Table 6: Measures of MROC curve for four sets of simulations at six different sample sizes.

Simulation	Sample Size (nD = nH = n)	AUC	Sn	1-Sp	Linear combination
1	25	0.9987	0.9832	0.0168	-19.57*X1+6.20*X2+15.05* X3
	50	0.9983	0.9809	0.0191	-8.04*X1+5.37* X2+10.04* X3
	100	0.9975	0.9764	0.0236	-5.89*X1+5.82* X2+8.29* X3
	150	0.9951	0.9662	0.0338	-6.59*X1+5.14* X2+7.71* X3
	200	0.9957	0.9685	0.0315	-7.34*X1+5.40* X2+7.99* X3
	300	0.9954	0.9674	0.0326	-7.25*X1+5.47* X2+7.88* X3
2	25	0.7474	0.6824	0.3177	8.29*X1+1.25* X2-1.20* X3
	50	0.7250	0.6638	0.3362	-0.29*X1+0.94* X2+1.40* X3
	100	0.7066	0.6497	0.3503	-1.84*X1+1.09* X2+1.61* X3
	150	0.7096	0.6519	0.3480	-0.97*X1+1.12* X2+1.41* X3
	200	0.7069	0.6499	0.3501	-1.61*X1+1.12* X2+1.66* X3
	300	0.6982	0.6433	0.3567	-1.14*X1+1.01* X2+1.49* X3
3	25	0.9345	0.8593	0.1407	17.00*X1+7.39* X2-1.06* X3
	50	0.9217	0.8421	0.1578	13.78*X1+6.84* X2-1.24* X3
	100	0.9377	0.8615	0.1385	10.77*X1+8.29* X2-1.24* X3
	150	0.9288	0.8509	0.1491	13.35*X1+7.49* X2-1.17* X3
	200	0.9292	0.8508	0.1491	14.13*X1+7.36* X2-1.21* X3
	300	0.9269	0.8479	0.1521	14.50*X1+7.36* X2-1.14* X3

	25	0.7485	0.6925	0.3075	-1.41*X1-0.23* X2-0.73* X3
	50	0.6125	0.5801	0.4199	-2.18*X1-1.32* X2+0.03* X3
	100	0.6633	0.6177	0.3822	-0.74*X1-1.25* X2-0.23* X3
4	150	0.6415	0.6029	0.3971	0.67*X1-0.52* X2-0.32* X3
	200	0.6529	0.6114	0.3886	-0.38*X1-0.64* X2-0.27* X3
	300	0.6617	0.6181	0.3819	-1.30*X1-0.62* X2-0.25* X3

(a) The MROC curves in the graph indicate the case of almost parallel curves

(b) The MROC curves in the graph depict the curves where one is superior to the other

(c) The MROC curves in the graph depict curves that cross each other

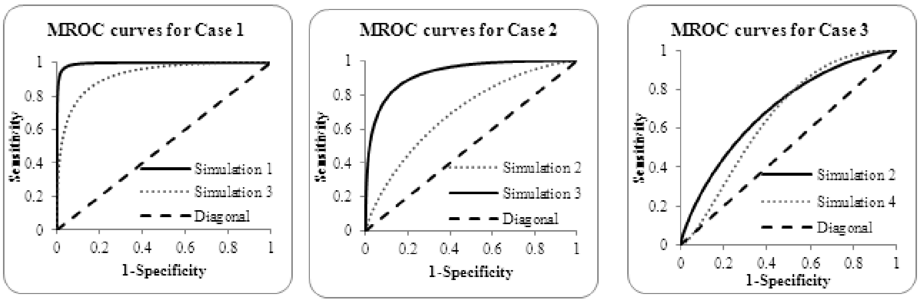

Figure 3: MROC curves of the considered combinations.

the comparison of MROC curves using proposed methods i.e, almost parallel curves (Simulation 1 and 3), One superior to the other (Simulation 2 and 3), and cross over curves (Simulation 2 and 4). The comparisons are made using all the proposed methods and the results are explained. The MROC model is applied to all the simulations at various sample sizes {25, 50, 100, 150, 200, 300} to obtain the linear combination required to obtain scores and the results are given in Table 6.

Case 1: Almost Parallel MROC curves

The MROC curves considered for comparison in this case are Simulations 1 and 3 (Fig 3a). The accuracies of Simulations 1 and 3 over all considered sample sizes is observed to be around 99% and 93% respectively, indicating that simulation 1 has a better capacity in discriminating the subjects into one of the two classes. This is further proved using the proposed testing procedures.

Table 7: Comparison between Simulation 1 and Simulation 3 (Almost parallel curves).

Sample Size	based on scores						based on Mean vectors and Covariance Matrices					
	Mahalanobis Distance			Bhattacharya Distance			Mahalanobis Distance			Bhattacharya Distance		
	$D_U^{(1)}$	$D_U^{(3)}$	Z Value	$D_{BU}^{(1)}$	$D_{BU}^{(3)}$	Z Value	$D^{(1)}$	$D^{(3)}$	Z Value	$D_B^{(1)}$	$D_B^{(3)}$	Z Value
25	4.4300	2.1606	3.3282	2.5035	0.6178	2.6138	4.6109	2.0982	3.5631	2.8688	0.9946	2.4402
50	4.2346	2.0772	4.4185	2.2686	0.5566	3.4863	4.1778	2.0171	4.4068	2.2863	0.8364	2.9681
100	4.0146	2.2524	5.5419	2.0268	0.6451	4.9653	3.9913	2.2246	5.5738	2.0648	0.9162	4.1546
150	3.6845	2.0986	6.1905	1.7037	0.5613	5.4400	3.6681	2.0806	6.2240	1.7739	0.8082	4.6240
200	3.7549	2.0887	8.5637	1.1643	0.5494	7.8888	3.7394	2.0746	8.6221	1.8369	0.8027	6.9828
300	3.7087	2.0589	9.5426	1.6064	0.5239	8.2647	3.6980	2.0508	9.5588	1.8129	0.7678	7.5005

The distance measures depicted in table 5 are obtained using the scores of the MROC models shown in Table 6. The distance observed in Simulation 1 is greater than that of Simulation 3 indicating that the overlapping area in simulation 1 is minimum when compared to simulation 3. This is further authenticated by using inference procedures that compare the distance measures. The Z values in Table 7 indicate a true difference between the measures at all sample sizes, thus proving that the difference between curves can also be identified at small and large sample sizes.

The computation of distance measures with mean vectors and covariance matrices is done on the same set of simulations 1 and 3. These results are reported in Table 7 and display similar information that is observed using distance measures through scores i.e., simulation 1 has a larger distance value than simulation 3 indicating that the extent of correct classification is better in simulation 1.

Table 8: Comparison between Simulation 2 and Simulation 3 (One superior to the other).

Sample Size	based on scores						based on Mean vectors and Covariance Matrices					
	Mahalanobis Distance			Bhattacharya Diistance			Mahalanobis Distance			Bhattacharya Distance		
	$D_U^{(2)}$	$D_U^{(3)}$	Z Value	$D_{BU}^{(2)}$	$D_{BU}^{(3)}$	Z Value	$D^{(2)}$	$D^{(3)}$	Z Value	$D_B^{(2)}$	$D_B^{(3)}$	Z Value
25	0.9962	2.1606	2.6678	0.1749	0.6178	2.1482	1.0369	2.0982	2.3732	0.3910	0.9946	2.3807
50	0.9821	2.0772	3.0442	0.1435	0.5566	2.6603	0.9782	2.0171	2.8974	0.2508	0.8364	3.5238
100	0.8246	2.2524	5.6761	0.0947	0.6451	4.7055	0.8147	2.2246	5.6528	0.1794	0.9162	6.0376
150	0.8126	2.0986	7.2500	0.0892	0.5613	5.9418	0.8098	2.0806	7.2119	0.1847	0.8082	7.1704
200	0.7885	2.0887	7.7808	0.0822	0.5494	6.5293	0.7845	2.0746	7.7524	0.1859	0.8027	7.8896
300	0.7436	2.0589	9.4013	0.0728	0.5239	8.1688	0.7421	2.0508	9.3737	0.1783	0.7678	9.0608

Case 2: One superior to the other

This case is considered to show the comparison between two MROC curves where one supersedes the other (Fig 3b). Simulations 2 and 3 are used to represent this case and the accuracies obtained through MROC model are 70% and 93% respectively. This shows that simulation 3 has a superior MROC curve when compared to simulation 2 and the same can be visualized in graph.

The above observation between simulations 2 and 3 is validated by conducting inferential procedures based on distance measures obtained using the scores of the MROC curve. On observing the distances of simulations 2 and 3 computed using scores (Table 8), it is clearly seen that there is a high discrepancy between them which in turn explains significant

Table 9: Comparison between Simulation 2 and Simulation 4 (Curves that cross each other).

Sample Size	based on Scores						based on Mean vectors and Covariance Matrices					
	Mahalanobis Distance			Bhattacharya Distance			Mahalanobis Distance			Bhattacharya Distance		
	$D_U^{(2)}$	$D_U^{(4)}$	Z Value	$D_{BU}^{(2)}$	$D_{BU}^{(4)}$	Z Value	$D^{(2)}$	$D^{(4)}$	Z Value	$D_B^{(2)}$	$D_B^{(4)}$	Z Value
25	0.9962	1.0591	0.1582	0.1749	0.2396	0.5346	1.0369	1.0513	0.0359	0.3910	0.4415	0.2971
50	0.9821	0.5207	1.4265	0.1435	0.0499	1.1984	0.9782	0.5138	1.4373	0.2508	0.1413	1.0726
100	0.8246	0.6545	0.7449	0.0947	0.0763	0.3483	0.8147	0.6551	0.7010	0.1794	0.2100	0.5158
150	0.8126	0.5454	1.7904	0.0892	0.0679	0.5779	0.8098	0.5452	1.7826	0.1847	0.1914	0.1518
200	0.7885	0.5802	1.4847	0.0822	0.0779	0.1209	0.7845	0.5807	1.4511	0.1859	0.2374	1.1915
300	0.7436	0.6020	1.1007	0.0728	0.0887	0.4819	0.7421	0.6026	1.0844	0.1783	0.2280	1.4160

Z values. Further, the proposed methods are observed to work effectively at small and large sample sizes in identifying the true difference between two MROC curves.

The distances obtained through mean vectors and covariance matrices also follow the same phenomenon as observed in the case of distances based on scores. The Z values obtained are significant indicating that simulation 3 has a better discriminating ability than simulation 2.

Case 3: Curves that cross each other

A case of MROC curves where it is difficult to identify a better curve is when they cross each other. Simulations 2 and 4 are the sets that relate to the phenomenon where the curves cross each other (Fig 3c). The accuracies observed for simulation 2 and simulation 4 after performing MROC analysis are around 70% and 66% respectively at all sample sizes.

The distances observed in simulation 2 and simulation 4 using scores (Table 9) of the MROC model are approximately equal portraying the fact that the overlapping area in both simulations is same. The Z values obtained through inference procedures are smaller than the table value at 5% level. This shows that simulation 2 and simulation 4 have similar discriminating capacity.

The distances observed in simulations 2 and 4 using mean vectors and covariance matrices (Table 9) are approximately equal thus concluding that the MROC curves do not differ from each other. The Z values obtained at all sample sizes are smaller than the standard value at 5% level showing that sample size does not affect the discriminating ability of the curves.

DISCUSSIONS

The main objective of the paper is to compare two MROC curves using distance measures based on the scores obtained through linear combination and mean vectors and covariance matrices. The importance of distance measures in classification led to the proposal of comparison procedures using two distance measures. Mahalanobis distance considers equal variances/covariances and is calculated for both the scores of the MROC curve as well as mean vectors and covariance matrices of the data. Another distance measure namely Bhattacharya distance that considers unequal variances/covariances between the populations is also used, since there are no restrictions on covariance matrices of healthy and diseased populations in MROC model. The main reason for considering two distance measures is that the distance between two populations is either overestimated or underestimated when the assumption of homogeneity of variances/covariance matrices is ignored.

The proposed methods are explained by using real datasets and simulation studies. Further, the effect of sample size on these methods is studied by applying the procedures at various sample sizes. The results indicate that all the proposed methods provide similar results and can be applied based on the nature of variances/covariance matrices. Further, it is noticed that the sample size does not affect the comparison procedures thus making them a valuable tool for comparisons. In ILP data, the MROC curves obtained for males and females do not differ from each other indicating a similar extent of correct classification for both genders. In Salmon data, the MROC curves obtained for Canadian and Alaskan salmons are proved to be indifferent when the gender of the fish is to be identified. The results further depict that the distances are underestimated when the property of homogeneity of variances/covariance matrices is neglected. Further, illustrations regarding three different scenarios (almost parallel curves, one superior to the other and curves that cross each other) are given with the support of simulation studies. This attempt is to enhance the knowledge of the researcher in making the use of distance measures in comparing curves rather than using the intrinsic and summary measures of ROC. The above three scenarios are handled using the following combinations. The difference between simulations 1 and 3 and 2 and 3 is observed at all sample sizes using all the proposed procedures to mimic the real situation where the curves (tests) are almost parallel and one superior to other. The MROC curves of simulations 2 and 4 are compared and proved to be indifferent indicating a similar level of classification in both the curves thus giving an insight about the case of curves that cross each other. The present paper reveals an interesting fact pertaining to the case of curves that cross each other. Along with the summary measures and intrinsic measures, distance measures have also not extracted the true difference between this type of curves. Hence, this case requires a special attention to distinguish the curves and to determine which test is to be used for the classification. However, it is shown that distance measures can also be used to explain the true difference between the two MROC curves apart from the conventional measures.

REFERENCES

[1] Aiyi Liu, Schisterman, E. F. & Yan Zhu (2005). On linear combinations of biomarkers to improve diagnostic accuracy. Statistics in Medicine, **24,** 37–47. http://dx.doi.org/10.1002/sim.1922

[2] Bendi Venkata Ramana, Prof. M. Surendra Prasad Babu & Prof. N. B. Venkateswarlu (2012). ILPD (Indian Liver Patient Dataset) Data Set, https://archive.ics.uci.edu/ml/datasets/ILPD+(Indian+Liver+Patient+ Datset).

[3] Feng Gao, Chengjie Xiong, Yan Yan &Zhengjun Zhang (2008). Estimating optimum linear combination of multiple correlated diagnostic tests at a fixed specificity with receiver operating characteristic curves. Journal of Data Science, **6,** 1–13.

[4] Gourevitch, V. & Galanter, E. (1967). A significance test for one parameter isosensitivity functions. Psychometrika, **32,** 25–33. http://dx.doi.org/10.1007/BF02289402

[5] Greenhouse, S.W. & Nathan Mantel (1950). The evaluation of diagnostic tests. Biometrics **6(4),** 399–412. http://dx.doi.org/10.2307/3001784

[6] Hanley, J. A. & McNeil, B. J. (1982). A meaning and use of the area under a receiver operating characteristics (roc) curves. Radiology, **143,** 29–36. http://dx.doi.org/10.1148/radiology.143.1.7063747

[7] Hanley, J. A. & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology, **148,** 839–843. http://dx.doi.org/10.1148/radiology.148.3.6878708

[8] John, Q. Su & Jun, S. Liu (1993). Linear combinations of multiple daignostic markers. Journal of American Statistical Association, **88(424),** 1350–1355. http://dx.doi.org/10.1080/01621459.1993.10476417

[9] Johnson, R.A. & Wichern, D.W. (2007). Applied Multivariate Statistical Analysis. Pearson Prentice Hall, 6/e.

[10] Marascuilo, L. A. (1970). Extension of the significance test for one parameter signal detection hypotheses. Psychometrika, **35,** 237–243. http://dx.doi.org/10.1007/BF02291265

[11] McClish D. K. (1987). Comparing the areas under more than two independent roc curves. Medical Decision Making, **7,** 149–155. http://dx.doi.org/10.1177/0272989X8700700305

[12] Metz, C. E. & Kronman, H. B. (1980). Statistical significance tests for binormal roc curves. Journal of Mathematical Psychology, 22, 243–243. http://dx.doi.org/10.1016/0022-2496(80)90020-6

[13] Sameera, G. Vishnu Vardhan, R & Sarma, KVS (2015). Binary Classification using Multivariate Receiver Operating Characteristic curve for Continuous Data. Journal of Biopharmaceutical Statistics, http://dx.doi.org/10.108010543406.2015.1052479.

[14] Vishnu Vardhan, R. Sameera, G. Chandrasekharan, P.A. & Thulasi Beere (2015). Inferential Procedures for Comparing the Accuracy and Intrinsic Measures of Multivariate Receiver Operating Characteristic (MROC) Curve. International Journal of Statistics in Medical Research, **4,** 87-93. http://dx.doi.org/10.6000/1929-6029.2015.04.01.10