

Inferential Procedures for Comparing the Accuracy and Intrinsic Measures of Multivariate Receiver Operating Characteristic (MROC) Curve

R. Vishnu Vardhan^{1,*}, G. Sameera¹, P.A. Chandrasekharan² and Thulasi Beere²

¹Department of Statistics, Pondicherry University, Puducherry - 605014, India

²Department of Gynecology, Sri Venkateswara Medical College, Tirupati - 517502, India

Abstract: A number of classification techniques are prevailing in literature. Of them, one of the most important techniques is the Receiver Operating Characteristic (ROC) curve. A multivariate extension of this technique is proposed in the recent years. This technique helps in classifying the objects/individuals into one of the two classes by considering two or more markers. The most important measure of an ROC curve is the Area Under the Curve (AUC) and it explains the accuracy and discriminating ability of the test under study. There are two intrinsic measures of ROC namely sensitivity (Sn) and specificity (Sp). Further, two ROC curves can be compared by comparing their measures. The practical application of the proposed inferential procedures is explained with the help of two real datasets namely, Indian Liver Patient (ILP) Dataset and Intra Uterine Growth Restricted Fetal Doppler Study (IUGRFDS) dataset. These inferential procedures are developed based on the measures of multivariate ROC (MROC) curve proposed by Sameera G, R Vishnu Vardhan and KVS Sarma [1].

Keywords: Multivariate Receiver Operating Characteristic Curve, Area Under the Curve, testing procedures.

1. INTRODUCTION

Receiver Operating Characteristic (ROC) curve is a well known classification technique especially used in the field of medicine and relatively has its importance in diversified fields such as Experimental Psychology (Green and Swets [2]), Radiology (Lusted [3]), Psychophysics (Swets [4]), Biometrics (Arun and Rohin [5]), Data Mining (Sinha and Jerrold [6]), Epidemiology (Cook [7]) and many more (Krzanowski and Hand [8]). The ROC curve is summarized using a number of indices of which the most important are sensitivity, specificity and accuracy of the curve. The ROC curve is a tradeoff between 1-Specificity (Sp) and Sensitivity (Sn) which are known as intrinsic measures. Here, Sensitivity and specificity are the probabilities of correct classification of diseased and healthy populations respectively. A test's ability is explained by its accuracy measure, Area Under the Curve (AUC) which lies between 0 and 1.

The univariate ROC model was developed by considering that the populations follow distributions such as binormal and bilogistic (Green and Swets [2]). In statistical theory, comparison of indicators/measures has a prominence and validates the significance of the measures considered. The comparison of two tests was done by comparing the measures of the curve by various authors over a period of time using different

statistics. Here, a brief review of the existing literature on the inferential aspects of comparing the measures of ROC curves is given. Diagnostic tests are vital for detection of a disease. The process of evaluating a diagnostic test and comparing two tests to identify a better one dates to the fifties. Greenhouse and Mantel [9] described a test to compare two sensitivities at a fixed level of specificity. The proposed t-statistic considers the difference of 10th percentile of the diseased population and 95th percentile of the healthy population to compare the curves. Gourevitch and Galanter [10] proposed a large sample test to test the significance of two observed data points by testing the hypothesis that they belong to the same d' function, which is generally interpreted as a detectability index. Marascuillo [11] extended the above large sample test for k observed data points. Further, confidence interval procedures were described to identify statistically significant data points. Metz and Kronman [12] proposed comparison strategies for comparing an obtained and estimated binormal ROC curve, two independent curves and among several independent ROC curves using a χ^2 statistic for comparing the parameters 'a' and 'b' where $a = \frac{\mu_D - \mu_H}{\sigma_D}$ and $b = \frac{\sigma_H}{\sigma_D}$

of the ROC curve by considering the maximum likelihood estimates of the parameters. Hanley and McNeil [13] compared two ROC curves by checking whether the difference between their areas is significant or not in order to identify the sample size which helps in acquiring a required sensitivity. Hanley and McNeil [14] induced correlation coefficient to compare two ROC curves obtained from same set of subjects. It was shown that the power of the test increases by considering the correlation between the areas that arises by studying same individual twice.

*Address correspondence to this author at the Department of Statistics, Pondicherry University, Puducherry – 605014, India;
E-mail: rvvccr@gmail.com

McClish [15] devised F test and Studentized Range (SR) test to compare more than two independent ROC curves and recommended F test for comparing areas although SR test performs better for balanced designs. A Monte Carlo study was performed to evaluate significance of the test and its power. McClish [16] proposed a combined AUC which is weighted average of areas under each study or strata and a Chi-square test for testing the equality of areas across studies or strata. The power of the test is also explored and the test can be applied for either parametric or non parametric areas of the curve. Metz, Herman and Roe [17] proposed a new algorithm, ROCKIT, which compares two diagnostic tests even if some subjects did not undergo both the tests. This algorithm converged for more than 99.8% of the datasets to which it was applied. The above said are various testing procedures on the parameters and measures of univariate ROC curve. However, the multivariate extension of the ROC model is developed in recent years.

Sameera G, R Vishnu Vardhan and KVS Sarma [1] proposed a Multivariate Receiver Operating Characteristic (MROC) model which is based on the minimax procedure to provide a linear combination of markers for classification. The present work focuses on developing some inferential aspects on measures of the MROC model. Comparison of two curves has its importance due to the fact that identification of a better test procedure among existing is always a necessity. To meet this situation, comparison procedures are proposed for the above said MROC curves. When curves are to be compared, it gives rise to two cases: independent and dependent. Here, independence refers to the conventional way of comparing two measures of two distinct populations where as dependence refers to the comparison of measures obtained from same set of subjects. The comparisons of accuracy measure AUC and intrinsic measure sensitivity are given for the above criteria. Further, a researcher might be interested in a fixed specificity value basing on the prior information about the study. There are cases when the curves differ at some points but are coinciding with each other at the other points. In these cases, the sensitivity obtained at that particular specificity is tested to identify the significance of the test.

2. METHODOLOGY

Let k be the number of markers considered in the study, X_i , $i = 1, 2, \dots, m$ be the marker values of i^{th} individual in healthy group, where $X_i = (X_{i1}, X_{i2}, \dots, X_{ik})$ and Y_j , $j = 1, 2, \dots, n$ be the marker values of j^{th}

individual in diseased group, where $Y_j = (Y_{j1}, Y_{j2}, \dots, Y_{jk})$. Let us suppose X and Y are two multivariate normal random vectors 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 $Y \sim MVN(\mu_D, \Sigma_D)$. The multivariate ROC (MROC) model given by Sameera G, R Vishnu Vardhan and KVS Sarma [1] is

$$y(c) = \Phi \left(\frac{b'(\mu_D - \mu_H) - (b' \Sigma_H b)^{1/2} \Phi^{-1}(1-x)}{(b' \Sigma_D b)^{1/2}} \right) \tag{1}$$

where

$$b = [t \Sigma_D + (1-t) \Sigma_H]^{-1} (\mu_D - \mu_H); 0 < t < 1$$

The AUC of MROC curve is given as

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

The variance of the above AUC can be obtained using bootstrapping. The need for bootstrap estimate arises due to the difficulty in obtaining variance expression for (2) as it involves mean vectors and covariance matrices. Let 'B' be the number of bootstraps obtained from the data with the sample sizes n_H and n_D respectively from healthy and diseased populations. Then the bootstrapped (bs) AUC estimate and its variance are given as

$$\widehat{AUC}_{bs} = \frac{1}{B} \sum_{b=1}^B AUC_b \tag{3}$$

$$Var(\widehat{AUC}_{bs}) = \frac{1}{B-1} \sum_{b=1}^B (AUC_b - \widehat{AUC}_{bs})^2 \tag{4}$$

The obtained \widehat{AUC} is to be verified whether the test has any discriminating ability or not. A test with AUC equal to half is considered useless as it does not classify at least 50% of the individuals correctly. For this test, the ROC curve coincides with the chance line and $TPR=FPR$. Hence, the null and alternative hypothesis is defined as $H_0 : \widehat{AUC}_{bs} = 0.5 \sim H_1 : \widehat{AUC}_{bs} \neq 0.5$. Then the test statistic is defined as

$$Z = \frac{\widehat{AUC}_{bs} - 0.5}{\sqrt{Var(\widehat{AUC}_{bs})}} \tag{5}$$

The comparisons of two AUC's of MROC curves are given in both independent and dependent case. The null and alternative hypothesis for comparing two curves is $H_0 : \widehat{AUC}_{bs(1)} = \widehat{AUC}_{bs(2)} \sim$

$H_1 : \widehat{AUC}_{bs(1)} \neq \widehat{AUC}_{bs(2)}$. Then the test statistic for independent curves is defined as

$$Z = \frac{\widehat{AUC}_{bs(1)} - \widehat{AUC}_{bs(2)}}{\sqrt{\text{Var}(\widehat{AUC}_{bs(1)}) + \text{Var}(\widehat{AUC}_{bs(2)})}} \quad (6)$$

and the test statistic for dependent curves is defined as

$$Z = \frac{\widehat{AUC}_{bs(1)} - \widehat{AUC}_{bs(2)}}{\sqrt{\text{Var}(\widehat{AUC}_{bs(1)}) + \text{Var}(\widehat{AUC}_{bs(2)}) - 2\text{Cov}(\widehat{AUC}_{bs(1)}, \widehat{AUC}_{bs(2)})}} \quad (7)$$

where variances can be computed using (4) and covariance is given as

$$\text{Cov}(\widehat{AUC}_{bs(1)}, \widehat{AUC}_{bs(2)}) = \frac{1}{B-1} \sum_{b=1}^B (AUC_{b(1)} - \widehat{AUC}_{bs(1)})(AUC_{b(2)} - \widehat{AUC}_{bs(2)}) \quad (8)$$

The values in equations (6) and (7) follow normal distribution with significance level α .

Similarly, comparing two ROC curves can also be done by comparing their sensitivities. The sensitivity of MROC curve is given as

$$S_n = \Phi\left(\frac{b' \mu_D - C}{(b' \sum_D b)^{1/2}}\right) \quad (9)$$

Then the bootstrapped estimate of Sensitivity and its variance are given as

$$\widehat{S}_{nbs} = \frac{1}{B} \sum_{b=1}^B S_{n_b} \quad (10)$$

$$\text{Var}(\widehat{S}_{nbs}) = \frac{1}{B} \sum_{b=1}^B (S_{n_b} - \widehat{S}_{nbs})^2 \quad (11)$$

Sensitivity explains the test's ability to correctly identify the diseased individuals. In order to make sure that the test can be used for identifying diseased individuals its sensitivity should be greater than 0.5. Then, the hypothesis for testing sensitivity of an MROC is defined as $H_0: S_n=0.5 \sim H_1: S_n \neq 0.5$. The test statistic that can be used for testing this hypothesis is

$$Z = \frac{\widehat{S}_{nbs} - 0.5}{\sqrt{\text{Var}(\widehat{S}_{nbs})}} \quad (12)$$

Two MROC curves can also be compared using their sensitivities. This comparison leads to the identification of that test which identifies the diseased individuals better than the other. The hypothesis defined for this kind of comparison is

$H_0 : \widehat{S}_{nbs(1)} = \widehat{S}_{nbs(2)} \sim H_1 : \widehat{S}_{nbs(1)} \neq \widehat{S}_{nbs(2)}$. The test statistic is defined as

$$Z = \frac{\widehat{S}_{nbs(1)} - \widehat{S}_{nbs(2)}}{\sqrt{\text{Var}(\widehat{S}_{nbs(1)}) + \text{Var}(\widehat{S}_{nbs(2)})}} \quad (13)$$

for independent case and

$$Z = \frac{\widehat{S}_{nbs(1)} - \widehat{S}_{nbs(2)}}{\sqrt{\text{Var}(\widehat{S}_{nbs(1)}) + \text{Var}(\widehat{S}_{nbs(2)}) - 2\text{Cov}(\widehat{S}_{nbs(1)}, \widehat{S}_{nbs(2)})}} \quad (14)$$

where variance values can be computed using (11) and the covariance is given as

$$\text{Cov}(\widehat{S}_{nbs(1)}, \widehat{S}_{nbs(2)}) = \frac{1}{B-1} \sum_{b=1}^B (S_{n_{b(1)}} - \widehat{S}_{nbs(1)})(S_{n_{b(2)}} - \widehat{S}_{nbs(2)}) \quad (15)$$

for dependent case. The above test statistic follows normal distribution at $\alpha\%$ level of significance.

Further, a researcher may be interested in testing sensitivity at a particular specificity. This helps in identifying a better test when a certain percentage of healthy individuals are identified correctly.

The specificity of an MROC curve is obtained and given as

$$Sp = \Phi\left(\frac{c - b' \mu_H}{(b' \sum_H b)^{1/2}}\right) \quad (16)$$

From the above expression c value can be obtained as

$$c = b' \mu_H + (b' \sum_H b)^{1/2} \Phi^{-1}(Sp) \quad (17)$$

The above equation is then substituted in sensitivity to obtain the sensitivity at a particular specificity. The expression for sensitivity is derived and given as

$$S_n = \Phi\left(\frac{b' \mu_D - b' \mu_H - (b' \sum_H b)^{1/2} \Phi^{-1}(Sp)}{(b' \sum_D b)^{1/2}}\right) \quad (18)$$

Once the sensitivity is obtained, testing procedures are equivalent to those of overall sensitivity i.e., the sensitivity at a particular specificity can be tested by using equations (12), (13) and (14).

3. RESULTS AND DISCUSSIONS

In this section, the practical application of the above proposed methodology is explained with the help of two

real datasets namely, Indian Liver Patient (ILP) Dataset [18] and Intra Uterine Growth Restricted Fetal Doppler Study (IUGRFDS) dataset. The description of the datasets are given in three subsequent subsections dedicated to demonstrate the inferential procedures of testing measures of single MROC curve, two MROC curves (independent case and dependent case). Along with these procedures, the sensitivities of MROC curves are tested at fixed levels of specificities ranging from 0.5 to 0.9 with an increment of 0.1. The sensitivities thus obtained are tested for normality and observed to be normally distributed.

3.1. Testing the Measures of Single MROC Curve

The ILP Dataset contains 10 variables that are age, gender, total Bilirubin, direct Bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT and Alkphos. It contains 416 diseased and 167 healthy patients. Selector is a class label used to classify the subjects into groups (liver patient or not).

The optimal cutoff point, the best linear combination and the AUC are derived from the expressions (1) and (2) and are given as follows: The optimal cutoff point obtained for the data is 2.2119 and the best linear combination is

$$0.0184*Age + 0.0035*TB + 0.1851*DB + 0.0013*Alkphos + 0.0028*Sgpt - 0.0003*Sgot + 0.5250*TP - 0.9622*ALB + 0.5290*A/G$$

The accuracy measure AUC and sensitivity are observed to be 0.7403 and 0.6878 respectively. In order to test the significance of measures of MROC curve, bootstrap samples are generated. The bootstrap estimates of AUC and sensitivity are observed to be 0.7503 and 0.6974 respectively. Further, using the test statistics' given in equations (5) and (12), the significance of AUC and sensitivity are tested and found to be highly significant with $Z_{AUC} = 11.7001$ and $Z_{S_n} = 11.2285$ indicating that the obtained MROC model can be used for classifying the subjects with 75.03% of accuracy, which is depicted in Figure 1.

The interest of an investigator lies in some fixed specificity basing on the prior information about the markers under study. In such cases, the sensitivity obtained at fixed specificity is to be tested. Conventionally, the fixed levels of specificity are considered between 0.5 and 1 with an increment of 0.1. The resulting test statistic values presented in Table 1 show that the sensitivities are significant when the fixed specificity lies between 0.5 and 0.8. On further

investigation, it is noticed that the sensitivities are significant for the specificity values between 0.5 and 0.8998.

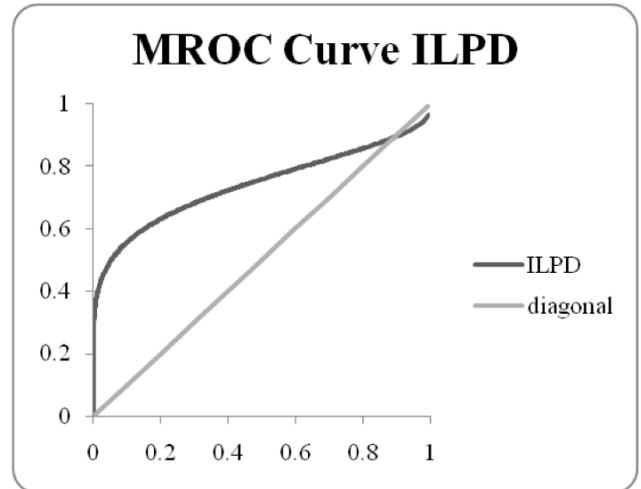


Figure 1: MROC curve for Indian Liver Patient Dataset.

Table 1: Bootstrap Estimates of MROC Curve at Fixed Specificities for Indian Liver Patient Dataset

Sp	\hat{S}_n	$Var(\hat{S}_n)$	Z Value
0.9	0.5752	0.0015	1.9537
0.8	0.6472	0.0008	5.0962
0.7	0.6961	0.0006	7.7648
0.6	0.7353	0.0006	9.5575
0.5	0.7692	0.0006	10.5965

However, in some cases information about markers might not be available to prefix the specificity. In such situations, one can consider Youden's index to identify the specificity value. In this particular dataset, the highest Youden's index value is 0.4472 observed at specificity 0.8 where 64.72% of diseased individuals are identified correctly.

3.2. Comparison of AUC's and Sensitivities of Two MROC Curves – Independent Case

For explaining the independent case, comparison is done between males and females to check whether the biomarkers considered in the study identify the disease similarly in both the genders or not. From the dataset it is noticed that the study constitutes 441(75.65%) males and 142(24.35%) females. For the male patients, the AUC is observed to be 0.7541 with an optimal cutoff 2.0266 and sensitivity 0.6959. The linear combination used for classification is

$$0.0242 \cdot \text{Age} + 0.0063 \cdot \text{TB} + 0.2011 \cdot \text{DB} + 0.0012 \cdot \text{Alkphos} + 0.0029 \cdot \text{Sgpt} - 0.0004 \cdot \text{Sgot} + 0.4145 \cdot \text{TP} - 0.8069 \cdot \text{ALB} + 0.2592 \cdot \text{A/G}$$

For the female patients, the accuracy value (AUC) is observed to be 0.7232 with an optimal cutoff value of 3.3666 and sensitivity 0.6781. The best linear combination is given as

$$0.0009 \cdot \text{Age} - 0.5375 \cdot \text{TB} + 1.1341 \cdot \text{DB} + 0.0012 \cdot \text{Alkphos} - 0.0056 \cdot \text{Sgpt} + 0.0075 \cdot \text{Sgot} + 1.1394 \cdot \text{TP} - 1.8168 \cdot \text{ALB} + 1.5752 \cdot \text{A/G}$$

The MROC curves for both males and females are shown in Figure 2.

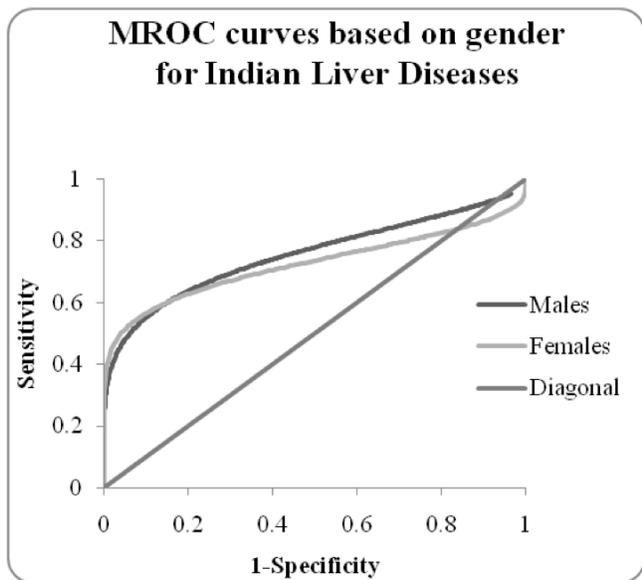


Figure 2: MROC curves for Independent case.

From Figure 2, we observe that the two MROC curves are very close to each other indicating little to no difference between males and females when the classification is done basing on the considered biomarkers. This is further proved by using testing procedure considering the AUC's and sensitivities of both the curves. The bootstrap estimates were used to obtain the Z value to show that there is no difference between male and female patients (Table 2). The bootstrap AUC's of male and female patients have a difference of 0.007 which is negligible thus indicating that any one of the combinations can be used for classification irrespective of the gender. Similarly, the difference between the sensitivities is 0.003 indicating that the markers to identify the status of individuals are same irrespective of the gender.

The MROC curves are further compared at fixed specificities. The results obtained are portrayed in

Table 2: Bootstrap Estimates of MROC Measures for Independent Case

Gender	\widehat{AUC}	$Var(\widehat{AUC})$	\widehat{S}_n	$Var(\widehat{S}_n)$
Male	0.7668	0.0005	0.7079	0.0004
Female	0.7599	0.0010	0.7051	0.0006
Z Value	0.174 (<1.96)		0.087 (<1.96)	

Table 3 and indicate that the sensitivities of male and female patients do not differ at any fixed specificity. This observation can be supported by Figure 2 in which the curves are very close to each other indicating no difference between them.

Table 3: Bootstrap Estimates of MROC Curves for Independent Case at Fixed Specificity

S_p	Male		Female		Z Value
	\widehat{S}_n	$Var(\widehat{S}_n)$	\widehat{S}_n	$Var(\widehat{S}_n)$	
0.9	0.5747	0.0021	0.5817	0.0023	-0.1066
0.8	0.6562	0.0012	0.6571	0.0014	-0.0173
0.7	0.7111	0.0009	0.7077	0.0013	0.0709
0.6	0.7543	0.0007	0.7477	0.0014	0.1440
0.5	0.7913	0.0006	0.7819	0.0016	0.2006

3.3. Comparison of AUC's and Sensitivities of Two MROC Curves - Dependent Case

To demonstrate the methodology of dependent case, there is a need to have a correlated data where different test procedures are applied on same set of subjects. The IUGRFDS dataset is collected from Sri Venkateswara Medical College which is a tertiary cum care hospital meets the above criterion where two procedures namely CPR and MCA are used to identify whether the blood flow from the womb of the mother to the baby is sufficient enough for its growth. The dataset consists of 82 samples in which 42(51.22%) are healthy and 40(48.78%) are diseased. Three indices are used to observe the growth of baby namely Pulsatility Index (PI), Resistivity Index (RI) and Systolic/Diastolic (S/D) ratio.

In order to meet the criterion defined, the best linear combination along with its optimal cutoff point are computed and listed in Table 4. Further, measures AUC and sensitivity are also computed and their significance is tested. The MROC curves for two procedures CPR and MCA are graphed in Figure 3.

Table 4: Measures of MROC for Dependent Case

Procedure	AUC	S_n	Optimal Cutoff	Best Linear Combination
CPR	0.7111	0.6531	1.6888	$0.9422 * PI - 1.6268 * RI + 1.8088 * S/D$
MCA	0.6273	0.5975	-0.9252	$0.0182 * PI - 4.5439 * RI + 0.6091 * S/D$

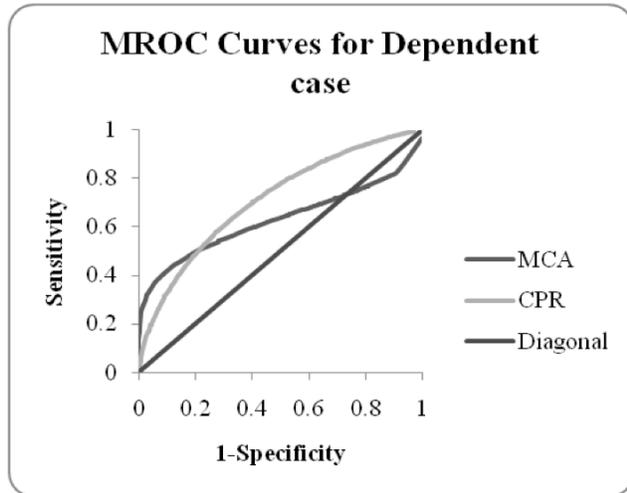


Figure 3: MROC curves for Dependent case.

The MROC curves in Figure 3 depict the case of curves that cross each other which makes it difficult to identify a better curve visually. Hence, the curves are compared by comparing their measures AUC and sensitivity. The bootstrap estimates (Table 5) are used to compare the curves to identify the better procedure. The AUC's and sensitivities of the curves are compared with each other and the results obtained indicate that though there is no difference between the two curves i.e., either procedure can be used for identifying the blood flow to the baby.

Table 5: Bootstrap Estimates of MROC Curves for Dependent Case

Estimates	CPR	MCA	Z value
\widehat{AUC}	0.7395	0.6758	0.7577
$Var(\widehat{AUC})$	0.0030	0.0029	
\widehat{S}_n	0.6785	0.6342	0.6478
$Var(\widehat{S}_n)$	0.0020	0.0019	

Further, the curves are compared at fixed specificities to see if they differ at that fixed point on the x-axis. Though the curves appear to be differing from each other, the results indicate that they do not differ leading to the conclusion that any procedure can be used basing on the convenience.

Table 6: Bootstrap Estimates of MROC Curves for Dependent Case at Fixed Specificity

Sp	CPR		MCA		Z Value
	\widehat{S}_n	$Var(\widehat{S}_n)$	\widehat{S}_n	$Var(\widehat{S}_n)$	
0.9	0.4017	0.0159	0.1671	0.0132	-0.1226
0.8	0.5493	0.0109	0.3749	0.0173	0.1851
0.7	0.6553	0.0075	0.5911	0.0129	0.5047
0.6	0.7373	0.0055	0.7507	0.0141	0.8112
0.5	0.8029	0.0042	0.8493	0.0105	1.0657

4. CONCLUSIONS

The objectives of the paper were to check whether an obtained curve is significant for future classification or not and to identify a better curve among two MROC curves. These are met by developing testing procedures for comparing the measures of MROC curve. Two datasets were used to explain the applications of proposed testing procedures. The results from testing the measures of single MROC curve indicate that the AUC and Sensitivity of the MROC curve are significant thus explaining that the obtained model can be used for classifying future individuals with an accuracy of 75.03%. The comparisons made at fixed specificity showed that the model is significant between 0.5 and 0.8998. Further, for identifying a fixed specificity when prior information about the study is unavailable, Youden's index was used and the optimal specificity is observed at 0.8 with sensitivity of 0.6472. For demonstrating the independent case, comparison between males and females of ILP dataset was made and the results show that gender does not influence the parameters to be considered for identifying liver disorders even though males are more prone to liver diseases compared to females. The curves were further compared at fixed specificities to check whether they differ at particular point but the results indicate no difference between them. In dependent case i.e., IUGRFDS dataset, the MROC curves depict the case of crossing over which makes it difficult to compare graphically. Even though the accuracy obtained in the case of CPR is higher than MCA, the testing procedures show that the

procedures: CPR and MCA do not differ from each other for identifying the blood flow from the womb of the mother to baby. This explains the need of using testing procedures to identify better tests because one would assume that CPR is a better procedure by looking at the accuracy and sensitivity while the results prove that the practitioner can use either CPR or MCA basing on financial convenience and availability of equipment. The curves were also tested at fixed specificities and the results indicate no difference among them leading to the conclusion that either can be used basing on the availability of facilities required to perform them. Thus, the testing procedures proposed in this paper help in authenticating a model used for classification or comparing two MROC models to identify a better one between them.

ACKNOWLEDGEMENT

The first author would like to acknowledge University Grants Commission (UGC) for funding [F.No. 42-1002/2013(SR) dated 22/03/2013 and 18/04/2013]

The second author wishes to acknowledge DST for providing sufficient funding for the research under DST-INSPIRE Programme, (0023-2013). We thank all the patients who participated in the study.

REFERENCES

- [1] Sameera G, Vishnu Vardhan R, Sarma KVS. Binary Classification using Multivariate Receiver Operating Characteristic Curve for Continuous Data. *J Biopharm Stat* 2014; Accepted for publication.
- [2] Green DM, Swets JA. *Signal Detection theory and Psychophysics*. Wiley, New York 1966.
- [3] Lusted LB. Signal detectability and medical decision making. *Science* 1971; 171: 1217-1219. <http://dx.doi.org/10.1126/science.171.3977.1217>
- [4] Swets JA. The Relative Operating Characteristic in psychology. *Science* 1973; 182: 990-1000. <http://dx.doi.org/10.1126/science.182.4116.990>
- [5] Ross A, Govindarajan R. Feature Level Fusion Using Hand and Face Biometrics. Appeared in Proc. of SPIE Conference on Biometric Technology for Human Identification II, (Orlando, USA) 2005; Vol. 5779: pp. 196-204.
- [6] Sinha AP, May JH. Evaluating and Tuning Predictive Data Mining Models Using Receiver Operating Characteristic Curves. *Journal of Management Information Systems* 2005; 21(3): 249-280.
- [7] Cook NR. ScD. Use and Misuse of the Receiver Operating Characteristic Curve in Risk Prediction. *Circulation* 2007; 115: 928-935. <http://dx.doi.org/10.1161/CIRCULATIONAHA.106.672402>
- [8] Krzanowski WJ, Hand DJ. *ROC curves for continuous data, Monographs on Statistics and Applied Probability*. CRC Press, Taylor and Francis Group. New York, NY 2009.
- [9] Greenhouse S, Mantel N. The evaluation of diagnostic tests. *Biometrics* 1950; 6: 399-412. <http://dx.doi.org/10.2307/3001784>
- [10] Galanter G. A significance test for one parameter isosensitivity functions. *Psychometrika* 1967; 32: 25-33. <http://dx.doi.org/10.1007/BF02289402>
- [11] Marascuilo LA. Extensions of significance test for one-parameter signal detection hypotheses. *Psychometrika* 1970; 35: 237-243. <http://dx.doi.org/10.1007/BF02291265>
- [12] Metz CE, Kronman HB. Statistical Significance Tests for Binormal ROC curves. *J Math Psychol* 1980; 22: 218-243. [http://dx.doi.org/10.1016/0022-2496\(80\)90020-6](http://dx.doi.org/10.1016/0022-2496(80)90020-6)
- [13] Hanley JA, Mc Neil BJ. A Meaning and Use of the area under a Receiver Operating Characteristics (ROC) Curves. *Radiology* 1982; 143: 29-36. <http://dx.doi.org/10.1148/radiology.143.1.7063747>
- [14] Hanley JA, Mc Neil BJ. A method of Comparing the Areas under Receiver Operating Characteristics Analysis derived from the same cases. *Radiology* 1983; 148: 839-843. <http://dx.doi.org/10.1148/radiology.148.3.6878708>
- [15] McClish DK. Comparing the Areas Under More Than Two Independent ROC Curves. *Medical Decision Making* 1987; 7: 149-155. <http://dx.doi.org/10.1177/0272989X8700700305>
- [16] McClish DK. Combining and Comparing Area Estimates across Studies or Strata. *Medical Decision Making* 1992; 12: 274-279. <http://dx.doi.org/10.1177/0272989X9201200405>
- [17] Metz CE, Herman BA, Roe CA. Statistical Comparison of Two ROC Curve Estimates Obtained from Partially-Paired Datasets. *Medical Decision Making* 1998; 18: 110-121. <http://dx.doi.org/10.1177/0272989X9801800118>
- [18] Ramana BV, Babu MSP, Venkateswarlu NB. ILPD (Indian Liver Patient Dataset) Data Set, [https://archive.ics.uci.edu/ml/datasets/ILPD+\(Indian+Liver+Patient+Dataset\)](https://archive.ics.uci.edu/ml/datasets/ILPD+(Indian+Liver+Patient+Dataset)); 2012.