Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2022;8(2):208-233

Original Article

Multi-Class Classification using Mixtures of Univariate and Multivariate ROC Curves

Siva Gajjalavari, Vishnu Vardhan Rudravaram*

Department of Statistics, Pondicherry University, Puducherry, India.

ARTICLE INFO

Received 14.02.2022
Revised 07.04.2022
Accepted 06.05.2022
Published 15.07.2022

Keywords:
Mixture models;
Bi-Normal ROC;
Multivariate ROC;
Multi-class classification;
Area under the Curve.

ABSTRACT

Introduction: Receiver Operating Characteristic (ROC) curve is one of the widely used supervised classification technique to allocate/classify the individuals and also instrumental in comparing diagnostic tests. Generally, to deal with classification problems we need to have knowledge on class labels. In most of the medical scenarios, most of data sets exhibit multi-model patterns in class labels which leads to multi-class classification problems.

The main aim of this study is to address on the issue of constructing ROC models when there exists multi-model patterns in the class labels further, to classify the individuals for better diagnosis and also to reduce the complexity of graphical representation of ROC curves in such classification problems.

Methods: A new version of univariate and multivariate ROC models are proposed in the framework of Finite Mixtures, due to the flexibility of identifying and modelling the subcomponents in the heterogeneous populations.

Results: Oral Glucose Tolerance Test and Disk Hernia datasets are used and simulation studies are also performed. Results show that the proposed models possess better accuracy when compared with Bi-Normal and MROC models with reasonable low 1-Specificity and higher Sensitivity. The ROC curves are depicted in a 2D space rather than higher dimension for multi-class classification problem.

Conclusion: It is suggested that before one proceeds to model ROC curves, it is better to take a look at the density patterns of the study variable(s), which in turn help in explaining the true information between the classes and also provides good amount of "true" accuracy.

Introduction

Over the past seven decades, the Receiver Operating Characteristic (ROC) curve has witnessed developments in both theory and application. Particularly, in the field of diagnostic medicine, ROC curve has been widely used for evaluating the test's

*.Corresponding Author: vrstatsgur@gmail.com

performance and also useful in comparing diagnostic tests by means of Area under the Curve (AUC) and Sensitivities.[1] ROC curve is a unit square plot between 1-specificity (False Positive Rate) and sensitivity (True Positive Rate) at various thresholds.

On literature review about modelling ROC curve, good number of articles were found, which were based on the assumption that the populations follow Normal and Non-Normal distributions.[2-9] The most widely used parametric ROC model is the Bi-Normal model,[2] which assumes that both the populations are distributed as Normal. For more details on various bi-distributional ROC models, one can refer to.[10]

In the recent past, multivariate extensions of ROC curve were proposed under the assumption that the populations follow multivariate normal.[11-15]

In general, to deal with classification situations, we need to have knowledge on class labels. Even if the class labels are known, still there might be sub populations within each class labels. For instance, let us consider the OGTT (Oral Glucose Tolerance Test) data,[16] which has 21 sample observations and the histogram of the same is shown in figure 1(a). Here, an attempt is made to see whether there are any hidden sub populations with in healthy and diseased of OGTT data by using the EM algorithm.[17] On such exploration, a bi-modal pattern is witnessed in the diseased population, turning out to a total of three classes in the OGTT data shown in figure 1(b). So, before proceeding to build a classification model, it is better to do such an exercise to explore and find out the hidden sub populations in each of the known class labels, if any. One of the points of exploring the sub populations in each known class label is to further classify the individuals for better diagnosis or treatment regime. Now, the two-class problem extended to three-class problem, in general, it can be referred as a multi-class classification problem.

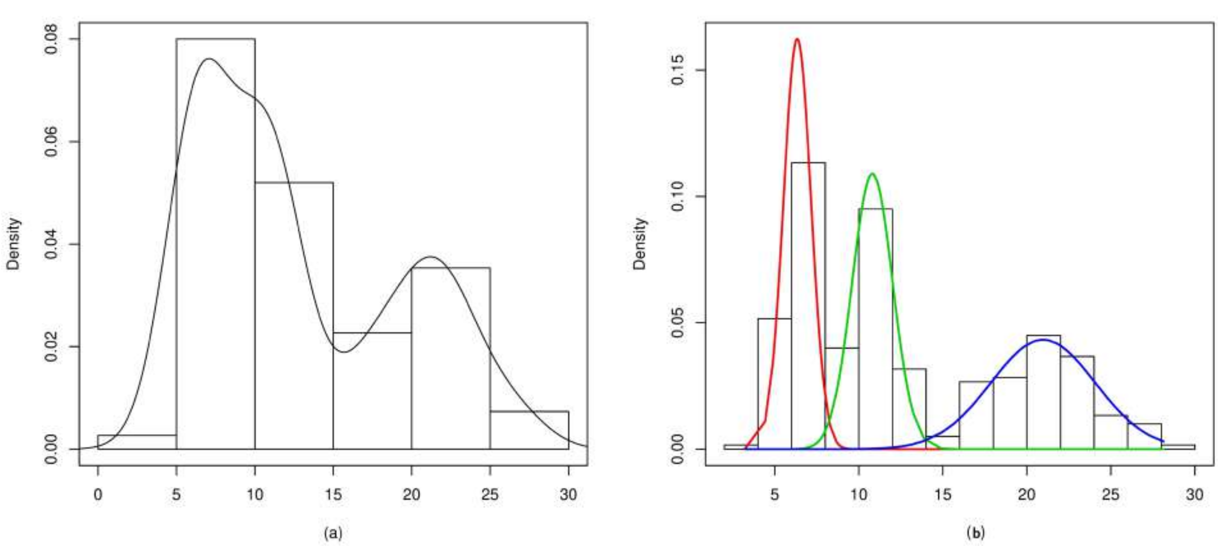

Figure 1. (a) The overall density plot of OGTT, (b) Plot after identifying the components in the OGTT data set.
The plots consists of
(a) Histogram and overlying the density curve of OGTT data.
(b) Identified sub components in OGTT data and resulted in 3- components.

Generally, the Bi-normal ROC curve and Multivariate Receiver Operating Characteristic Curve (MROC) curve will be useful when we deal with a binary class framework. Hence, there is a need to come out with an ROC curve in both univariate and multivariate setup that can accommodate the multi-class scenario. Here, to have the mathematical flexibility and advantages in constructing ROC curve, the multi-class ROC models are proposed in the framework of Finite Mixture Models (FMM).

Finite Mixture Models

These are widely used in many scientific areas, where the data likely to have several sub populations that are to be determined. FMM provides extreme flexibility in model fitting when the data have many modes, skewness and non-distributional characteristics. For more detailed account of major issues, methodologies on FMM and its applications in diversified areas, readers can see.[18, 19]

Generally, a p-component mixture density is given by

$$f(x_j;\Psi) = \sum_{i=1}^{p} \lambda_i \, \phi\left(x_j;\theta_i\right)$$

where the vector Ψ contains all unknown parameters of mixture model i.e., $\Psi = \left(\lambda_1,\lambda_2,\ldots,\lambda_{p-1},\xi'\right)'$, λ_i's are the mixing proportions ($0<\lambda_i<1$); $\sum_{i=1}^{p}\lambda_i = 1$ and ξ is the vector that consists all the distinct parameters in $(\theta_1,\theta_2,\ldots,\theta_p)$.

Let us define the density function as $\phi(x_j;\theta_i)$,

$$\phi\left(x_j;\ \theta_i\right)=\begin{cases}\phi\left(x_j;\ \mu_i,\sigma_i^2\right); \ Normal \\ \phi\left(\boldsymbol{x}_j;\ \boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i\right); \ Multivariate\ Normal\end{cases}$$

here, $\phi\left(x_j;\ \mu_i,\sigma_i^2\right) = \frac{1}{\sigma_i\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{x_j-\mu_i}{\sigma_i}\right)^2\right\}$

is the normal density; μ_i and σ_i^2 are mean and variance of i^{th} population.

$$\phi\left(\boldsymbol{x}_j;\ \boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i\right) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1}(x-\boldsymbol{\mu}_i)\right\}$$

is the multivariate normal density; $\boldsymbol{\mu}_i$ and Σ_i are mean vector and covariance matrix of i^{th} population.

In next sub-sections, a brief introduction about Bi-Normal ROC and MROC models are given.

Bi-Normal ROC model[2]

Let S denote the test scores in Normal/Healthy (H) and Abnormal/Diseased (D) populations, respectively,

$$\text{i.e., } S_H \sim N(\mu_H,\sigma_H^2),\ S_D \sim N(\mu_D,\sigma_D^2)$$

where μ_H, μ_D and σ_H^2, σ_D^2 are the means and variances of H and D populations, respectively. It is assumed that the mean of population D is greater than the mean of population H (*i.e.,* $\mu_D > \mu_H$), but no constraints are placed on the standard deviations.

The false positive rate (FPR) and true positive rate (TPR) are defined as[20]

$$FPR = P(S>c|H) = \Phi\left(\frac{\mu_H - c}{\sigma_H}\right);$$

$$TPR = P(S>c|D) = \Phi\left(\frac{\mu_D - c}{\sigma_D}\right)$$

Using FPR, we can obtain the expression of 'c' as $c=\mu_H-\sigma_H\,\Phi^{-1}[FPR]$, where Φ^{-1} is the inverse cumulative distribution function of Normal distribution. Then on substituting the 'c' in TPR, the expression for the Bi-normal ROC model will be

$$ROC = \Phi\left[\frac{\mu_D-\mu_H}{\sigma_D} + \frac{\sigma_H}{\sigma_D}\Phi^{-1}[FPR]\right]$$

Area under the Curve (AUC) is an important summary measure of an ROC curve and plays a prominent role in assessing the performance of a test, which takes value in the interval [0, 1]. A perfect diagnostic test is one with an area equal to 1 and a test with an area less than or equal to 0.5 is said to be inaccurate. The AUC of an ROC curve can be interpreted as the average sensitivity for all possible values of specificity and vice versa. AUC can be obtained by integrating the ROC expression over the range [0, 1]

The AUC expression for the Bi-Normal ROC is

$$AUC = P(S_D > S_H) = \Phi\left(\frac{\mu_D - \mu_H}{\sqrt{\sigma_D^2 + \sigma_H^2}}\right)$$

Multivariate ROC (MROC) model[15]

Let H and D be two populations assumed to follow multivariate normal distribution with mean vectors $\boldsymbol{\mu}_H, \boldsymbol{\mu}_D$ and covariance matrices Σ_H, Σ_D respectively. The intrinsic measures are follow as:

$$FPR = 1 - \Phi\left(\frac{c - b'\boldsymbol{\mu}_H}{\sqrt{b'\boldsymbol{\Sigma}_H b}}\right); TPR = \Phi\left(\frac{b'\boldsymbol{\mu}_D - c}{\sqrt{b'\boldsymbol{\Sigma}_D b}}\right)$$

where the vector 'b' is obtained using minimax procedure[21] as

$$b = [\, t\,\boldsymbol{\Sigma}_D + (1-t)\boldsymbol{\Sigma}_H]^{-1}(\boldsymbol{\mu}_D - \boldsymbol{\mu}_H) \quad (1)$$

here 't' is constant determined by the trial and error method in the interval (0,1) and the expression of cut-off point 'c' at each 't' can be obtain using minimax procedure by equating TPR and FPR (for detailed note on minimax approach in MROC, refer[15]). Upon simplifying the FPR expression and substituting the quantity c in TPR we obtain the following MROC expression.

$$MROC = \Phi\left[\frac{b'(\boldsymbol{\mu}_D - \boldsymbol{\mu}_H) - \sqrt{b'\boldsymbol{\Sigma}_H\, b}\,\Phi^{-1}(1 - FPR)}{\sqrt{b'\boldsymbol{\Sigma}_D\, b}}\right]$$

The AUC expression for the MROC is

$$AUC = P(S_D > S_H) = \Phi\left[\frac{b'\,(\boldsymbol{\mu}_D - \boldsymbol{\mu}_H)}{\sqrt{b'\,(\boldsymbol{\Sigma}_D + \boldsymbol{\Sigma}_H)\,b}}\right]$$

The linear combination for the MROC model is given by

$$U = b_1 X_1 + b_2 X_2 + \ldots + b_k X_k \Rightarrow U = b'X$$

where $b' = [b_1, b_2, \ldots, b_k]$, X is a $(k \times n)$ data matrix, k is the number of variables and here, b is the vector of coefficients obtained through equation.[1] Here, U contains the scores derived from each profile of the individuals. If $U \leq c$, then the individual will be classified into H populations, otherwise into D population.

Initial works on extending two-class problem to multi-class was by.[22-24] Further, a straight forward approach in generalizing the AUC as Volume Under Surface (VUS) was proposed by[25] and later some works on estimations of VUS was by.[26, 27, 28] Even though, the articles referred above deal with both binary and multiclass problems, but their complexity will be in visualizing the ROC curve with more than 3 classes and also the computation of AUC will be bit complicated.

In this work, we have made an attempt to propose the multi-class ROC models using the flexible approach of finite mixture models. The advantages that we have noticed in working with mixture ROC models are (i) even if there are p-components in the data, still the overall ROC curve can be presented in a 2-dimensional space (ii) the ROC mixtures will provide an easy understanding to a practitioner about classification/allocation of subjects (iii) the interpretation and presentation of summary measures will be with more ease. Hence, we have chosen EM algorithm and FMM as the mathematical treatment to determine the number of classes with in the data and also to model the mixture ROC form. For illustration purpose, the proposed methodology is given for three-class (k=3) population (figure (2)).

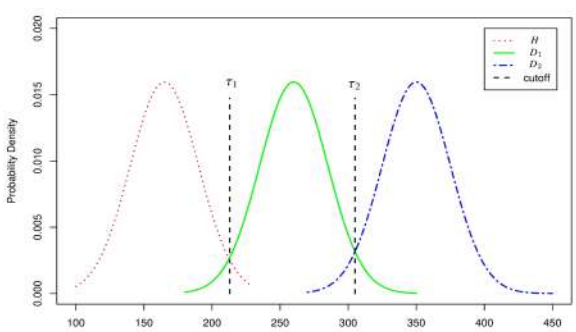

Figure 2. Hypothetically overlapping density curves and the cutoff values
The plot depict the hypothetically overlying densities of three populations and two possible cutoff points.

The subsequent sections will be on the proposed methodology, discussions with necessary illustrations using real and simulated data sets.

Proposed Methodology
Mixture of Univariate Receiver Operating Characteristic (mUROC) curve

Let S denote the test scores of the identified classes as H, D_1 and D_2 populations

i.e., $S_H \sim N(\mu_H, \sigma_H^2); S_{D_1} \sim N(\mu_{D_1}, \sigma_{D1}^2)$ and

$S_{D_2} \sim N(\mu_{D_2}, \sigma_{D_2}^2)$.

Intrinsic measures of mUROC curve
The expression for the FPR in the mixture form is defined as

$$FPR = \lambda_1 \ FPR_1 + \lambda_2 \ FPR_2 \quad (2)$$

here, λ_1 and λ_2 are the mixing proportions and will be obtained using EM algorithm. By definition, FPR_1 and FPR_2 are derived from H and D_1 populations, and will take the following forms

$$FPR_1 = \Phi\left(\frac{\mu_H - c_1}{\sigma_H}\right); FPR_2 = \Phi\left(\frac{\mu_{D_1} - c_2}{\sigma_{D_1}}\right) \quad (3)$$

using equation (3), c_1 and c_2 can be written as

$$c_1 = \mu_H - \sigma_H \Phi^{-1}(FPR_1); \ c_2 = \mu_{D1} - \sigma_{D1} \Phi^{-1}(FPR_2) \quad (4)$$

where Φ^{-1} is the inverse cumulative distribution function of Normal distribution. By definition, the TPR_1 and TPR_2 are obtained from D_1 and D_2 populations, and are as follows

$$TPR_1 = \Phi\left(\frac{\mu_{D_1} - c_1}{\sigma_{D_1}}\right); TPR_2 = \Phi\left(\frac{\mu_{D_2} - c_2}{\sigma_{D_2}}\right) \quad (5)$$

The expression for the TPR in the mixture form is defined as

$$TPR = \lambda_1 \ TPR_1 + \lambda_2 \ TPR_2 \quad (6)$$

by substituting equation (4) in equation (5) and thereafter equation (5) in equation (6), we get the expression for mixture ROC model

$$mUROC = \lambda_1 \ [\Phi(A_1 + B_1 \ \Phi^{-1}(FPR_1))] + \lambda_2 \ [\Phi(A_2 + B_2 \ \Phi^{-1}(FPR_2))] \quad (7)$$

which is called a 2-component mixture of Univariate Normal Receiver Operating Characteristic (mUROC) curve, where

$$A_1 = \frac{\mu_{D_1} - \mu_H}{\sigma_{D_1}}, A_2 = \frac{\mu_{D_2} - \mu_{D_1}}{\sigma_{D_2}}, B_1 = \frac{\sigma_H}{\sigma_{D_1}},$$

$B_2 = \frac{\sigma_{D_1}}{\sigma_{D_2}}$. The AUC expression for mUROC curve will be

$$mUAUC = \lambda_1 \Phi\left(\frac{A_1}{\sqrt{1+B_1^2}}\right) + \lambda_2 \Phi\left(\frac{A_2}{\sqrt{1+B_2^2}}\right) \quad (8)$$

In general, the p-component mixture ROC expression of equation (7) and AUC of equation (8) can be expressed as

$$mUROC_p = \sum_{i=1}^{p-1} \lambda_i \left[\Phi(A_i + B_i \Phi^{-1}(FPR_i)\right]$$

$$mUAUC_p = \sum_{i=1}^{p-1} \lambda_i \Phi\left(\frac{A_i}{\sqrt{1+B_i^2}}\right)$$

where

$\sum_{i=1}^{p-1} \lambda_i = 1 ; A_i = \frac{\mu_{i+1} - \mu_i}{\sigma_{i+1}} ; B_i = \frac{\sigma_i}{\sigma_{i+1}} ; i = 1,2, \dots, (p-1)$

If there are 'p' populations, there will be (p-1) linear combinations. Hence, the summation part λ_i's and ROC expression will range from 1 to (p-1).

Mixture of Multivariate Receiver Operating Characteristic (mMROC) Curve

Let $S_H \sim N_p(\boldsymbol{\mu_H}, \boldsymbol{\Sigma_H})$; $S_{D_1} \sim N_p(\boldsymbol{\mu_{D_1}}, \boldsymbol{\Sigma_{D_1}})$ and $S_{D_2} \sim Np(\boldsymbol{\mu_{D_2}}, \boldsymbol{\Sigma_{D_2}})$, where H, D_1

and D_2 are class labels pertaining to healthy/normal (H) and two abnormal populations, namely D_1 and D_2.

Intrinsic measures of mMROC curve

The expression for FPR in the mixture form is defined as

$$FPR = \lambda_1 FPR_1 + \lambda_2 FPR_2 \quad (9)$$

here

$$FPR_1 = 1 - \Phi\left(\frac{c_1 - b_1'\boldsymbol{\mu}_H}{\sqrt{b_1'\boldsymbol{\Sigma}_H b_1}}\right); FPR_2 = 1 - \Phi\left(\frac{c_2 - b_2'\boldsymbol{\mu}_{D_1}}{\sqrt{b_2'\boldsymbol{\Sigma}_{D_1} b_2}}\right)$$

where b_1, and b_2 ($\neq 0$) are the vector of coefficients of k variables obtained from a minimax procedure and are given as

$b_1 = [t\Sigma_{D1} + (1-t) \Sigma_H]^{-1} (\mu_{D1} - \mu_H); b_2 = [t\Sigma_{D_2} + (1-t) \Sigma_{D_1}]^{-1} (\mu_{D_2} - \mu_{D_1})$

where 't' is a constant lies between 0 and 1 and determined by trial and error method with an increment of 0.1. Here, b_1 corresponds to vector of coefficients of H and D_1 populations and b_2 corresponds to vector of coefficients of D_1 and D_2 populations. From equation (10), c_1 and c_2 can be written as

$$c_1 = b_1'\boldsymbol{\mu}_H + \sqrt{b_1'\boldsymbol{\Sigma}_H b_1}\, \Phi^{-1}(1 - FPR_1);$$

$$c_2 = b_2'\boldsymbol{\mu}_{D_1} + \sqrt{b_2'\boldsymbol{\Sigma}_{D_1} b_2}\, \Phi^{-1}(1 - FPR_2)$$

where Φ^{-1} is the inverse cumulative distribution of multivariate normal. The expression for TPR in the mixture form is defined as

$$TPR = \lambda_1 TPR_1 + \lambda_2 TPR_2 \quad (11)$$

Where;

$$TPR_1 = \Phi\left(\frac{b_1'\boldsymbol{\mu}_{D_1} - c_1}{\sqrt{b_1'\boldsymbol{\Sigma}_{D_1} b_1}}\right); TPR_2 = \Phi\left(\frac{b_2'\boldsymbol{\mu}_{D_2} - c_2}{\sqrt{b_2'\boldsymbol{\Sigma}_{D_2} b_2}}\right)$$

Substituting the expressions of c_1 and c_2 in equation (11), we obtain the ROC expression in the following mixture form as

mMROC $=\lambda_1 \Phi[\alpha_1 + \beta_1 \Phi^{-1}(1\text{-FPR}_1)]$
$+\lambda_2 \Phi[\alpha_2 + \beta_2 \Phi^{-1}(1\text{-FPR}_2)]$ (12)

which is called a 2- component mixture of Multivariate Receiver Operating Characteristic (mMROC) curve where

$$\alpha_1 = \frac{b_1'(\boldsymbol{\mu}_{D_1} - \boldsymbol{\mu}_H)}{\sqrt{b_1'\boldsymbol{\Sigma}_{D_1} b_1}},\ \alpha_2 = \frac{b_2'(\boldsymbol{\mu}_{D_2} - \boldsymbol{\mu}_{D_1})}{\sqrt{b_2'\boldsymbol{\Sigma}_{D_2} b_2}},$$

$$\beta_1 = -\frac{\sqrt{b_1'\boldsymbol{\Sigma}_H b_1}}{\sqrt{b_1'\boldsymbol{\Sigma}_{D_1} b_1}}\ \beta_2 = -\frac{\sqrt{b_2'\boldsymbol{\Sigma}_{D_1} b_2}}{\sqrt{b_2'\boldsymbol{\Sigma}_{D_2} b_2}}.$$

The AUC expression for mMROC curve will be

$$mMAUC = \lambda_1 \Phi\left(\frac{\alpha_1}{\sqrt{1+\beta_1^2}}\right) + \lambda_2 \Phi\left(\frac{\alpha_2}{\sqrt{1+\beta_2^2}}\right) \quad (13)$$

In general, a p-component mixture ROC expression of equation (12) and AUC of equation (13) can be expressed as

$$mMROC_p = \sum_{i=1}^{p-1} \lambda_i\ \Phi[\alpha_i + \beta_i \Phi^{-1}(1 - FPR_i)]$$

$$mMAUC_p = \sum_{i=1}^{p-1} \lambda_i \Phi\left(\frac{\alpha_i}{\sqrt{1+\beta_i^2}}\right)$$

Here

$$\sum_{i=1}^{p-1} \lambda_i = 1,\ \alpha_i = \frac{b_i'(\boldsymbol{\mu}_{i+1} - \boldsymbol{\mu}_i)}{\sqrt{b_i'\boldsymbol{\Sigma}_{i+1} b_i}},$$

$$\beta_i = -\frac{\sqrt{b_i'\boldsymbol{\Sigma}_i b_i}}{\sqrt{b_i'\boldsymbol{\Sigma}_{i+1} b_i}}, i = 1,2,\dots,(p-1)$$

Youden's Index

Usually, the test's capability or the performance of a biomarker depends on the optimal threshold, which provides maximum degree of correct classification. In order to determine the optimal threshold Youden's Index (J) is most preferably used measure and used by many researchers in the context to determine the optimal threshold,[29-32] it is given as

J = Max (TPR + TNR -1)

Here, TNR is the true negative rate (TNR = 1-FPR). Using the expressions of mUROC curve and mMROC curve, equations (2, 6, 9 & 11), Youden's index takes the following forms

$$J_U = Max\left\{\lambda_1\left[\Phi\left(\frac{\mu_{D_1} - c_1}{\sigma_{D_1}}\right) + \Phi\left(\frac{c_1 - \mu_H}{\sigma_H}\right)\right] + \lambda_2\left[\Phi\left(\frac{\mu_{D_2} - c_2}{\sigma_{D_2}}\right) + \Phi\left(\frac{c_2 - \mu_{D_1}}{\sigma_{D_1}}\right)\right]\right\} \quad (14)$$

$$J_M = Max\left\{\lambda_1\left[\Phi\left(\frac{b_1'\mu_{D_1} - c_1}{\sqrt{b_1'\Sigma_{D_1} b_1}}\right) + \Phi\left(\frac{c_1 - b_1'\mu_H}{\sqrt{b_1'\Sigma_H b_1}}\right)\right] + \lambda_2\left[\Phi\left(\frac{b_2'\mu_{D_2} - c_2}{\sqrt{b_2'\Sigma_{D_2} b_2}}\right) + \Phi\left(\frac{c_2 - b_2'\mu_{D_1}}{\sqrt{b_2'\Sigma_{D_1} b_2}}\right)\right]\right\} \quad (15)$$

where J_U & J_M are the Youden's indices of univariate and multivariate mixture ROC curves. Using the equations (14 & 15), the corresponding score related to the max J_U and J_M will be taken as optimal thresholds, say τ_1 and τ_2, for classifying into several populations.

Comparing two AUCs

In order to compare the AUC's obtained from the different models under univariate and multivariate setup, the null and alternative hypotheses along with the test statistic is given for both univariate and multivariate cases respectively.
For univariate case:

$$H_0: AUC = mUAUC \quad Vs \quad H_1: AUC \neq mUAUC$$

$$Z_{AUC} = \frac{\widehat{AUC} - m\widehat{UAUC}}{\sqrt{var(\widehat{AUC}) + var(m\widehat{UAUC})}}$$

For multivariate case:

$$H_0: AUC = mMAUC \quad Vs \quad H_1: AUC \neq mMAUC$$

$$Z_{AUC} = \frac{\widehat{AUC} - m\widehat{MAUC}}{\sqrt{var(\widehat{AUC}) + var(m\widehat{MAUC})}}$$

The test statistic value follows normal distribution with significance level α asymptotically.

Results and Discussions

The proposed methodology is supported with real data sets, namely, OGTT[16] for univariate and Disk Hernia[33] for multivariate ROC models, respectively. Simulations are performed for mUROC, mMROC, Bi-Normal ROC and MROC curves, respectively with different parameter combinations at varying sample sizes. Detailed discussion along with necessary results are presented in different sub sections.

Results of mUROC model
Real data set

The OGTT data (n=21) is visualized using histogram and an overlying density curve in figure (1). Figure 1(a) depicts the tri-model pattern of OGTT data and on applying EM-algorithm, 3-components are witnessed (see figure 1(b)). The proposed mUROC model is then applied to this 3-component data. The estimated summary and intrinsic measures cut-off are reported in table (1). The estimated mixing proportions are $\hat{\lambda}_1$ = 0.4992 and $\hat{\lambda}_2$ = 0.5008.
From the results of mUROC model, it is seen that identification of a hidden population helped out in exhibiting the true accuracy and

Table 1. ROC Curve parameters and AUC measures

	μ_H	μ_D		σ_H	σ_D		c		FPR	TPR	AUC	J	Z_{AUC}	Sig
Bi-Normal	6.7118	15.5550		1.9840	6.4567		9.79		0.0603	0.8140	0.9048	0.7537		
	μ_H	μ_{D_1}	μ_{D_2}	σ_H	σ_{D_1}	σ_{D_2}	τ_1	τ_2	FPR	TPR	mUAUC	J_U	1.99932	0.02279*
mUROC	6.7118	9.8793	21.2127	1.9840	1.6068	2.8680	9.22	12.06	0.1013	0.8273	0.9463	0.7260		

*Significant
The table consists of parameters of conventional Bi-Normal ROC and the proposed mixture univariate ROC model and its AUC measures, respectively.

reliable information in the data and which was at higher value than the Bi-Normal ROC model. This result and claim is based on using the expressions given in the above section of comparing two AUC's. From the results of Z statistic, it is clearly evident that the proposed mixture univariate ROC model provides better accuracy by accounting the hidden components in the model than the Bi-Normal ROC which has only two components. So, even if you have known class labels, always it is better to investigate further for observing hidden sub components.

The optimal cut points τ_1 (9.22) and τ_2 (12.06) are obtained at max J_U = 0.7260. In similar manner, the optimal cut point for the Bi-Normal ROC model is observed at max J = 0.7537 i.e., c = 9.79. If the test score is greater than the optimal cut point τ_2 = 12.06, then the individual is classified to diseased component 2 (D_2) else if the score lies between τ_1= 9.22 and τ_2=12.06, then the individual is classified to diseased component 1 (D_1) else individual will be assigned in healthy population (H). The above explanation can be presented in the following way

The individual is classified as =

$$\begin{cases} Normal & if \quad S \leq \tau_1 \\ Diseased\ Component\ 1 & if \quad \tau_1 \leq S \leq \tau_2 \\ Diseased\ Component\ 2 & if \quad S > \tau_2 \end{cases}$$

The ROC curves for Bi-Normal and mUROC are shown in figure (3). From the graph it can be clearly understood that mUROC curve is slightly superior with FPR (0.1013) and higher TPR (0.8273) values than that of Bi-Normal ROC curve (0.0603, 0.8140).

Simulation studies

In table (2), four sets of means and variances are considered along with the initial values for mixing proportions. Of which, the first two sets (A & B) are with unequal variances and the last two sets (C & D) are with equal variances. In each population, random samples of size n = {25,50,100,250,500} were generated using the parameter values defined in four sets and for each set, 1000 iterations were performed.

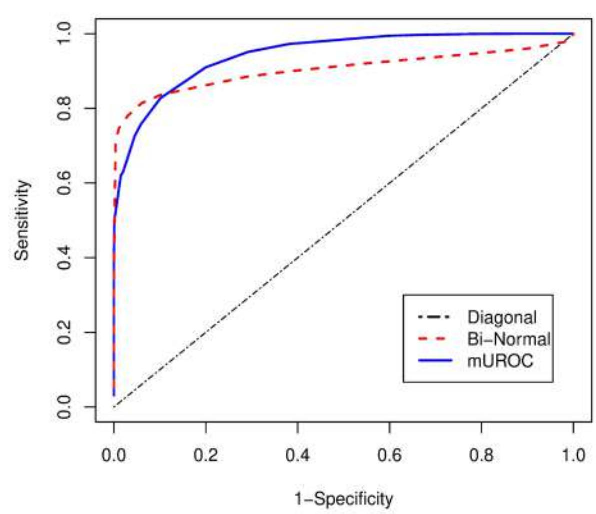

Figure 3. ROC Curves for the OGTT data estimated from Bi-Normal and mUROC

Table 2. Mixing Proportions, Means and Variances are considered for simulation studies

Sets	λ_1	λ_2	μ_H	μ_{D_1}	μ_{D_2}	σ_H	σ_{D_1}	σ_{D_2}
A	0.5	0.5	30.3	33.5	42.2	1.0	1.5	2.0
B	0.5	0.5	30.3	30.3	30.3	1.0	1.5	2.0
C	0.5	0.5	30.3	33.5	42.2	1.5	1.5	1.5
D	0.5	0.5	30.3	30.3	30.3	1.5	1.5	1.5

The table depicts the four sets of means and variances where the first two sets have unequal variances while the other have equal variances.

Table 3. Bi-Normal ROC and mUROC parameters at various sample sizes

Set	Sample size	Bi-Normal ROC parameters				mUROC parameters							
		$\hat{\mu}_H$	$\hat{\mu}_D$	$\hat{\sigma}_H$	$\hat{\sigma}_D$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\mu}_H$	$\hat{\mu}_{D_1}$	$\hat{\mu}_{D_2}$	$\hat{\sigma}_H$	$\hat{\sigma}_{D_1}$	$\hat{\sigma}_{D_2}$
	25	30.28999	37.83501	0.99371	4.80298	0.49897	0.50104	30.28999	33.53889	42.24360	0.99371	1.42354	2.28759
	50	30.29443	37.80698	0.99815	4.80505	0.49487	0.50513	30.29443	33.52135	42.19013	0.99815	1.44216	2.43348
A	100	30.29929	37.84866	0.99618	4.80581	0.49794	0.50207	30.29929	33.50459	42.19767	0.99618	1.47008	2.47394
	250	30.30241	37.85816	0.99984	4.81676	0.49865	0.50135	30.30241	33.50111	42.19369	0.99984	1.48849	2.47896
	500	30.30145	37.83916	1.00027	4.80862	0.50008	0.49992	30.30145	33.50060	42.20033	1.00027	1.49460	2.48484
	25	30.30087	30.27936	0.98916	2.02794	0.50893	0.49107	30.30087	30.30915	30.32666	0.98916	1.46140	2.34095
	50	30.29896	30.29042	0.99306	2.04269	0.50735	0.49265	30.29896	30.29435	30.30553	0.99306	1.47204	2.41244
B	100	30.30428	30.30635	0.99711	2.05678	0.50247	0.49753	30.30428	30.29453	30.28683	0.99711	1.49322	2.45562
	250	30.30187	30.29860	0.99877	2.05799	0.50106	0.49894	30.30187	30.29931	30.29915	0.99877	1.49452	2.48817
	500	30.30071	30.30195	0.99981	2.06128	0.50039	0.49961	30.30071	30.29562	30.29857	0.99981	1.49729	2.49579
	25	30.29027	37.86248	1.47544	4.58276	0.49819	0.50181	30.29027	33.52860	42.19653	1.47544	1.38771	1.40285
	50	30.30231	37.87722	1.48463	4.59170	0.50002	0.49998	30.30231	33.51888	42.20491	1.48463	1.47319	1.44567
C	100	30.29901	37.86276	1.49659	4.59827	0.50170	0.49830	30.29901	33.50236	42.19209	1.49659	1.47012	1.47389
	250	30.30195	37.85228	1.49666	4.59508	0.50047	0.49953	30.30195	33.50322	42.19834	1.49666	1.49416	1.49620
	500	30.29718	37.84655	1.49739	4.60281	0.50041	0.49959	30.29718	33.50430	42.19980	1.49739	1.49373	1.49321
	25	30.29557	30.28307	1.48423	1.49149	0.50246	0.49754	30.29557	30.30278	30.30278	1.48423	1.45850	1.45850
	50	30.30643	30.30014	1.49269	1.48801	0.50013	0.49987	30.30643	30.31424	30.31424	1.49269	1.47716	1.47716
D	100	30.30320	30.30215	1.49595	1.49531	0.50345	0.49655	30.30320	30.30023	30.30023	1.49595	1.49175	1.49175
	250	30.30292	30.30036	1.49844	1.49698	0.50925	0.49075	30.30292	30.30347	30.30347	1.49844	1.49467	1.49467
	500	30.29990	30.29905	1.50123	1.49692	0.50345	0.49655	30.29990	30.30083	30.30083	1.50123	1.49714	1.49714

The table depicts the estimates of mixing proportions, means and variances of the four sets considered for simulation studies at various sample sizes.

Using EM algorithm, mixing proportions, means and variances were estimated (table, 3) and table (4) has the estimates of the intrinsic and AUC measures of Bi-Normal ROC and mUROC curves at different sample sizes.

In continuation to the results obtained through simulation studies presented in tables (3 & 4), table (5) is about the AUC comparisons between Bi-Normal and mUROC models respectively. The p-value at each sample size for respective sets A and C indicate that there is evidence that the AUC obtained through

Table 4. Bi-Normal ROC and mUROC measures at various sample sizes

Set	Sample size	Bi-Normal ROC measures				mUROC measures				
		c	FPR	TPR	J	τ_1	τ_2	FPR	TPR	J_U
A	25	37.66963	0.02022	0.85858	0.83836	30.33720	37.98989	0.00808	0.93907	0.93099
	50	37.35715	0.00684	0.86308	0.85623	30.41233	37.71593	0.00406	0.93922	0.93516
	100	37.53743	0.01767	0.86980	0.85213	30.37432	37.85399	0.00666	0.94144	0.93478
	250	37.33082	0.01682	0.87378	0.85696	30.37836	38.00560	0.00778	0.94628	0.93849
	500	37.33457	0.02269	0.88892	0.86623	30.39347	37.55582	0.00892	0.95067	0.94175
B	25	30.27284	0.10281	0.21574	0.11293	30.34705	30.43388	0.08364	0.24271	0.15908
	50	30.37245	0.08438	0.22962	0.14523	30.35976	30.38076	0.08964	0.24785	0.15821
	100	30.36387	0.09169	0.23242	0.14073	30.39627	30.47292	0.10124	0.25127	0.15002
	250	30.42390	0.09125	0.23347	0.14221	30.37676	30.44896	0.07135	0.25767	0.18632
	500	30.48969	0.08408	0.25025	0.16617	30.39324	30.54108	0.08038	0.26974	0.18936
C	25	37.48985	0.04690	0.82295	0.77605	30.38452	37.84553	0.02949	0.88754	0.85804
	50	37.64905	0.03460	0.82665	0.79206	30.40888	37.85051	0.02231	0.89500	0.87269
	100	37.43438	0.03144	0.86252	0.83108	30.39531	38.02926	0.02631	0.89541	0.86909
	250	37.40727	0.02916	0.86581	0.83665	30.36950	37.65668	0.02459	0.90736	0.88276
	500	37.37727	0.03213	0.86587	0.83374	30.45430	38.01432	0.03082	0.90992	0.87911
D	25	30.25927	0.10035	0.11446	0.01411	30.32118	30.23058	0.17075	0.20607	0.03531
	50	30.39390	0.16276	0.22156	0.05880	30.33703	30.34987	0.18308	0.20898	0.02590
	100	30.27969	0.77375	0.79293	0.01918	30.40987	30.32450	0.23867	0.24340	0.00474
	250	30.32154	0.78966	0.80224	0.01258	30.29218	30.31667	0.82486	0.83294	0.00808
	500	30.30986	0.90768	0.91334	0.00566	30.24286	30.32719	0.88734	0.89423	0.00690

The table depicts the measures of Bi-Normal and mUROC model of the four sets considered for simulation studies at various sample sizes.

Table 5. The estimated AUC's of Bi-Nomral and mUROC along with values at various sample sizes

Set	Sample size	$\widehat{AUC}$	$m\widehat{UAUC}$	Z_{AUC}	Sig.
A	25	0.93479	0.97863	1.65738	0.04872*
	50	0.93544	0.98012	2.07329	0.01907*
	100	0.93714	0.98013	2.88499	0.00196*
	250	0.93736	0.98035	4.47766	0.00000*
	500	0.93741	0.98035	6.43528	0.00000*
B	25	0.49611	0.50264	0.06851	0.47269^{NS}
	50	0.49869	0.49942	0.05286	0.47892^{NS}
	100	0.49856	0.50045	0.28207	0.61105^{NS}
	250	0.49943	0.49995	0.03249	0.48704^{NS}
	500	0.49943	0.50021	0.03203	0.48722^{NS}
C	25	0.93773	0.96783	1.64521	0.04996*
	50	0.93973	0.96694	1.85489	0.03181*
	100	0.94013	0.96693	1.94829	0.02569*
	250	0.94052	0.96718	2.63752	0.00418*
	500	0.94040	0.96721	3.68464	0.00011*
D	25	0.49716	0.50026	0.03908	0.48441^{NS}
	50	0.49883	0.50098	0.02551	0.48982^{NS}
	100	0.49970	0.49996	0.01895	0.49244^{NS}
	250	0.49953	0.50000	0.01998	0.49203^{NS}
	500	0.49983	0.50023	0.01656	0.49340^{NS}

*Significant;
NS, Not significant.

mUROC model is comparatively better than the existing Bi-Normal ROC model. The sets A & C are taken in such a way that they exhibited the better classification and sets B & D relates to explain the behaviour of worst classification scenario. Since the sets B & D are taken to mimic the worst cases situation, the insignificant p-values are witnessed with nearer mean values of Bi-Normal and mUROC models.
The Bi-Normal and mUROC curves for each set at various sample sizes can be seen in figure (4) & (5). It is observed that, in each graph, the conventional Bi-Normal and proposed Mixture ROC curves almost overlap each other and which indicates that the extent of classification is similar at varying sample sizes. Here, the primary point is on focusing the use of mUROC curve, when unseen or hidden populations are extracted. In such cases, the Bi-Normal ROC may not be used and masks the true accuracy, FPR and TPR.

Results of mMROC model
Real data set

In multivariate case, to demonstrate the practical applicability of the proposed mMROC

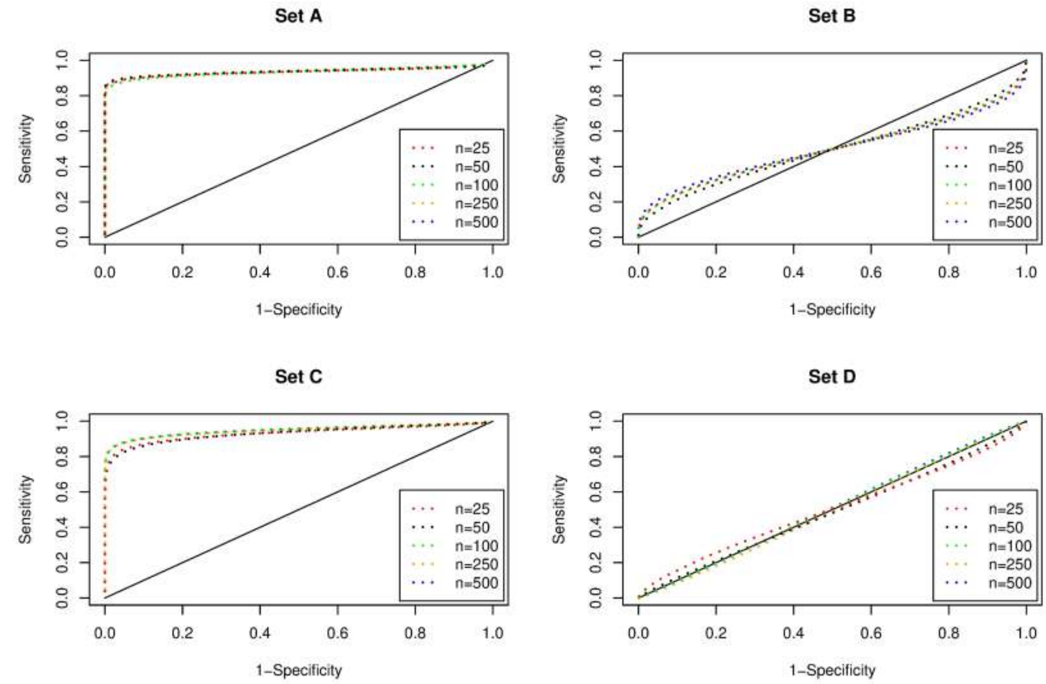

Figure 4. Bi-Normal ROC Curves for the simulated data sets at various sample sizes
The Bi-Normal ROC curves for the four sets are shown here. The first and third sets are the examples of best cases while the second and fourth are the worst cases of classification.

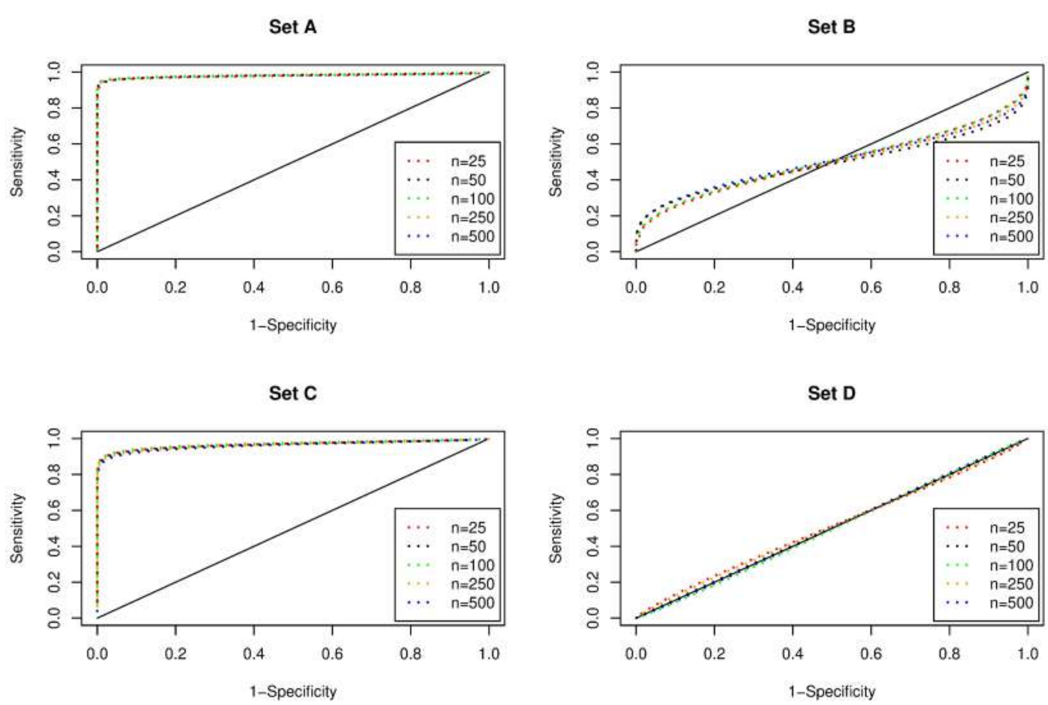

Figure 5. mUROC Curves for the simulated data sets at various sample sizes
The mUROC curves for the four sets are shown here. The first and third sets are the examples of best cases while the second and fourth are the worst cases of classification.

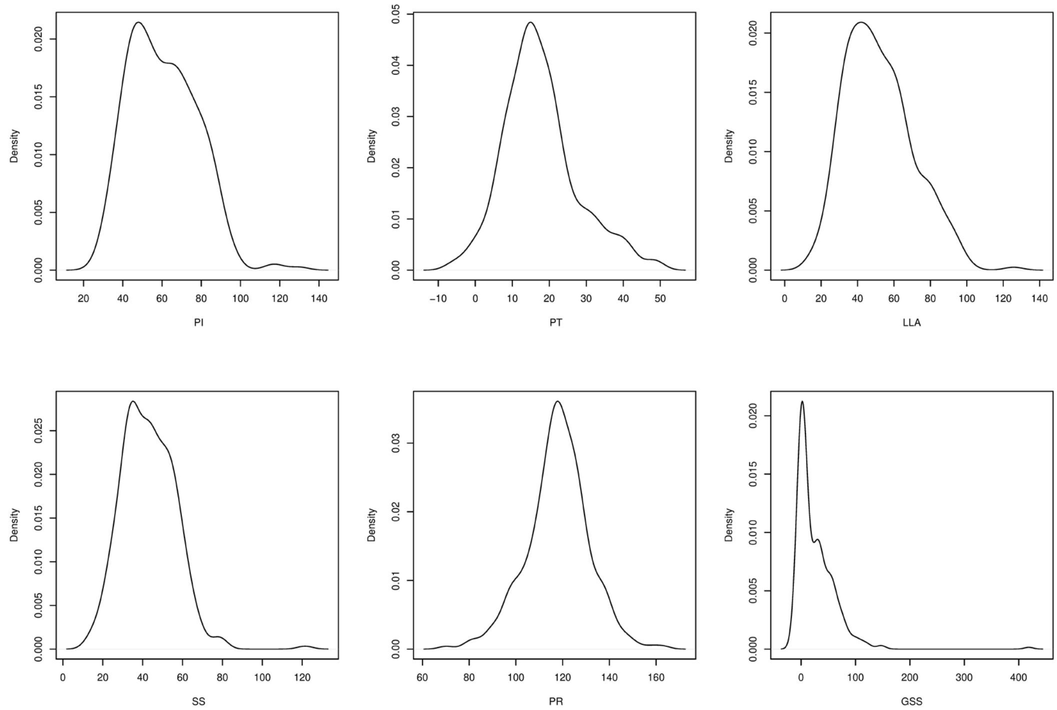

Figure 6. Density plots of each variable in Disk Hernia data set

Density plots of each variable in Disk Hernia data set shown here for observing the density patterns of each variable.

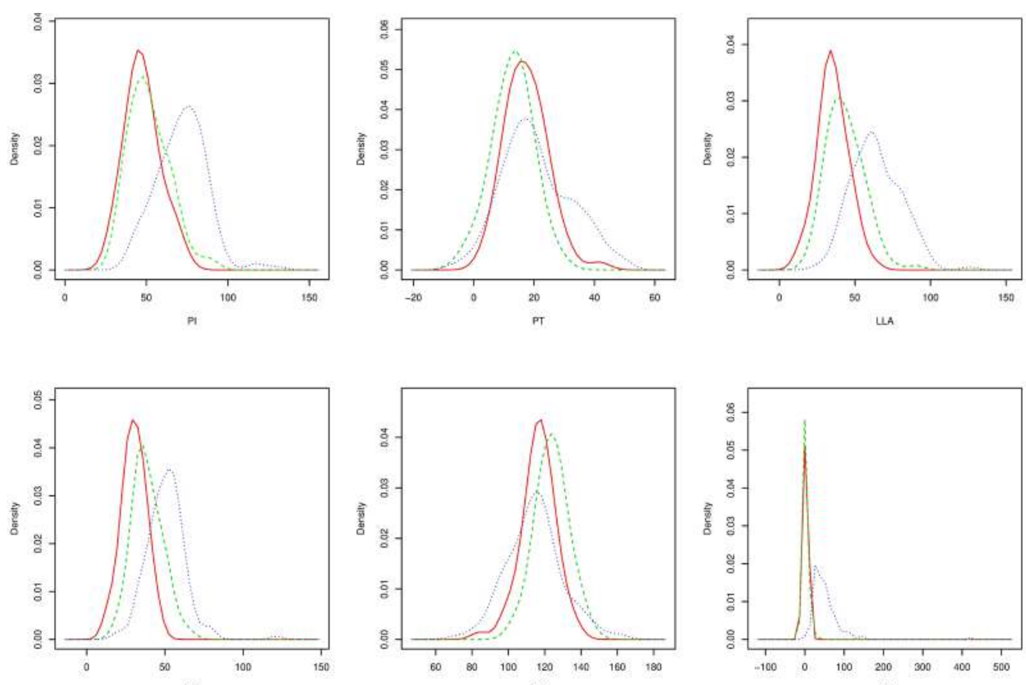

Figure 7. Identified mixture density plots of each variable in Disk Hernia data set

The identified components of each variable in Disk Hernia data set shown here, results shows that there are tri-modal patters exists in the data.

model, Disk Hernia data set is considered. It consists of 310 samples and 6 variables (PI, PT, LLA, SS, PR, GS) the density plots are for each variable is presented in figure (6).

Actually, this data set has three categories: Normal (N), Spondylolisthesis (ST) and Disk Hernia (DH). But, for illustration purpose, these class labels were ignored and on the complete data, EM algorithm has been applied to expose and extract the hidden components. Upon this exercise, three components were identified and depicted in figure (7).

The estimated mean vectors and covariance matrices of MROC and mMROC models are

MROC Model

Using these mean vectors and covariance matrices, the vectors b, b_1 and b_2 values, the intrinsic measures and summary

MROC Model

$$\hat{\mu}_H = \begin{pmatrix} 51.6856 \\ 12.8218 \\ 43.5423 \\ 38.8638 \\ 123.8912 \\ 2.1870 \end{pmatrix}, \quad \hat{\mu}_D = \begin{pmatrix} 64.6921 \\ 19.7911 \\ 55.9252 \\ 44.9015 \\ 115.0774 \\ 37.7776 \end{pmatrix}$$

$$\hat{\Sigma}_H = \begin{pmatrix} 152.9650 & 53.1515 & 107.1944 & 99.8139 & -54.6642 & 15.6189 \\ 53.1515 & 45.9502 & 25.3629 & 7.1993 & -21.1770 & 7.7775 \\ 107.1944 & 25.3629 & 152.8087 & 81.8374 & -27.5706 & 20.3356 \\ 99.8139 & 7.1993 & 81.8374 & 92.6171 & -33.4884 & 7.8483 \\ -54.6642 & -21.1770 & -27.5706 & -33.4884 & 81.2478 & -3.3084 \\ 15.6189 & 7.7775 & 20.3356 & 7.8483 & -3.3084 & 39.7785 \end{pmatrix}$$

$$\hat{\Sigma}_D = \begin{pmatrix} 311.9394 & 105.9087 & 236.1581 & 206.0250 & -20.9393 & 453.9016 \\ 105.9087 & 110.5790 & 78.8221 & -4.6700 & 36.3767 & 137.0192 \\ 236.1581 & 78.8221 & 386.8685 & 157.3297 & 19.0751 & 397.3477 \\ 206.0250 & -4.6700 & 157.3297 & 210.6891 & -57.3146 & 316.8820 \\ -20.9393 & 36.3767 & 19.0751 & -57.3146 & 198.5553 & 83.9594 \\ 453.9016 & 137.0192 & 397.3477 & 316.8820 & 83.9594 & 1656.2245 \end{pmatrix}$$

mMROC model

$$\hat{\lambda}_1 = 0.7171, \hat{\lambda}_2 = 0.2829$$

$$\hat{\mu}_H = \begin{pmatrix} 51.6856 \\ 12.8218 \\ 43.5423 \\ 38.8638 \\ 123.8912 \\ 2.1870 \end{pmatrix}, \quad \hat{\mu}_{D_1} = \begin{pmatrix} 58.1137 \\ 15.3218 \\ 49.0018 \\ 42.7924 \\ 112.3824 \\ 22.4897 \end{pmatrix}, \quad \hat{\mu}_{D_2} = \begin{pmatrix} 81.3650 \\ 31.1183 \\ 73.4723 \\ 50.2471 \\ 121.9078 \\ 76.5243 \end{pmatrix}$$

$$\hat{\Sigma}_H = \begin{pmatrix} 152.9650 & 53.1515 & 107.1944 & 99.8139 & -54.6642 & 15.6189 \\ 53.1515 & 45.9502 & 25.3629 & 7.1993 & -21.1770 & 7.7775 \\ 107.1944 & 25.3629 & 152.8087 & 81.8374 & -27.5706 & 20.3356 \\ 99.8139 & 7.1993 & 81.8374 & 92.6171 & -33.4884 & 7.8483 \\ -54.6642 & -21.1770 & -27.5706 & -33.4884 & 81.2478 & -3.3084 \\ 15.6189 & 7.7775 & 20.3356 & 7.8483 & -3.3084 & 39.7785 \end{pmatrix}$$

$$\hat{\Sigma}_{D_1} = \begin{pmatrix} 214.9352 & 42.7037 & 186.1923 & 172.2225 & -56.5213 & 178.1567 \\ 42.7037 & 48.6219 & 9.7010 & -5.9223 & 8.5664 & -19.1297 \\ 186.1923 & 9.7010 & 264.8324 & 176.4905 & -63.2351 & 218.4417 \\ 172.2225 & -5.9223 & 176.4905 & 178.1399 & -65.0804 & 197.2867 \\ -56.5213 & 8.5664 & -63.2351 & -65.0804 & 171.3358 & -69.6850 \\ 178.1567 & -19.1297 & 218.4417 & 197.2867 & -69.6850 & 371.3049 \end{pmatrix}$$

$$\hat{\Sigma}_{D_2} = \begin{pmatrix} 164.8756 & 0.9426 & 49.1751 & 163.9377 & -89.2222 & 244.2145 \\ 0.9426 & 86.8138 & -24.5043 & -85.8586 & -1.6486 & -81.5979 \\ 49.1751 & -24.5043 & 260.2649 & -24.6890 & 60.2220 & -104.0618 \\ 163.9377 & -85.8586 & -24.6890 & 249.7885 & -87.5861 & 325.8146 \\ -89.2222 & -1.6486 & 60.2220 & -87.5861 & 199.1375 & 102.8736 \\ 244.2145 & -81.5979 & -104.0618 & 325.8146 & 102.8736 & 2791.2693 \end{pmatrix}$$

Table 6. Cutoffs and measures of MROC curve for varying t values

t	c	FPR	TPR	AUC	J
0.1	-15.7252	0.1920	0.8080	0.8521	0.6159
0.2	-14.8812	0.1847	0.8153	0.8708	0.6306
0.3	-13.8502	0.1837	0.8163	0.8790	0.6325
0.4	-12.0972	0.1875	0.8125	0.8822	0.6250
0.5	**-12.9164**	**0.1852**	**0.8148**	**0.8830**	**0.6297**
0.6	-11.3782	0.1902	0.8098	0.8825	0.6196
0.7	-10.7431	0.1929	0.8071	0.8813	0.6141
0.8	-10.1777	0.1956	0.8044	0.8798	0.6088
0.9	-9.6706	0.1982	0.8018	0.8782	0.6036

The table consists of MROC measures at various values of ‘t’, highlighting the optimal cut point identified using Youden’s Index (J)

measures of MROC and mMROC curve were computed and are presented in tables (5) and (6), respectively. The optimal cut points of mMROC are τ_1= -13.0535 and τ_2= 24.0451 are observed at max J_M = 0.7294 with mMAUC = 0.9324 at t = 0.5. At t = 0.5, the linear combinations for MROC model and mMROC models are

MROC model

*U=-6.7887*PI+6.8242*PT+0.0283*LLA+6.7237*SS-0.0993*PR+0.0491*GS*

mMROC model

*U_I=-13.0695*PI+13.1404*PT-0.0291*LLA+12.9842*SS-0.0900*PR+0.1440*GS*

*U_{II}=5.0773*PI-4.7765*PT+0.0784*LLA-5.0192*SS+0.0676*PR+0.0306*GS*

Here, U denote the scores derived from the linear combinations of MROC model. Similarly, U_I and U_{II} are the scores obtained from the linear combinations of 2-component mMROC model.

If the test score is greater than τ_2= 24.0451, then individual is classified as diseased category, namely, Disk Hernia (D_2) else if the score lies between τ_1= -13.0535 and τ_2= 24.0451, the individual is classified into diseased category, Spondylolisthesis (D_1) else individual considered to be in healthy (H) population. From the results, it is observed that the AUC of proposed mMROC model (mMAUC=0.9324) is better than the AUC of MROC curve (AUC=0.8830). This shows that mMROC has the better accuracy of correct classification (Z=2.431513, p-value=0.007518).

In figure (8), the ROC curves for MROC and mMROC are depicted at t = 0.5. This reveals the fact that the performance is superior with mMROC than the MROC curve.

Table 7. Cutoffs and measures of mMROC curve for varying t values

t	τ_1	τ_2	FPR	TPR	mMAUC	J_M
0.1	-17.4021	16.1824	0.1523	0.8477	0.9062	0.6954
0.2	-15.7196	18.0715	0.1414	0.8586	0.9206	0.7172
0.3	-14.5492	19.9254	0.1367	0.8633	0.9280	0.7266
0.4	-13.6936	21.8854	0.1353	0.8647	0.9314	0.7279
0.5	**-13.0535**	**24.0466**	**0.1360**	**0.8640**	**0.9324**	**0.7294**
0.6	-12.5701	26.5148	0.1381	0.8619	0.9316	0.7237
0.7	-12.2059	29.4317	0.1412	0.8588	0.9298	0.7177
0.8	-11.9355	33.0070	0.1448	0.8552	0.9271	0.7104
0.9	-11.7414	37.5768	0.1489	0.8511	0.9240	0.7022

The table consists of mMROC measures at various values of ‘t’, highlighting the optimal cut point identified using Youden’s Index (J)

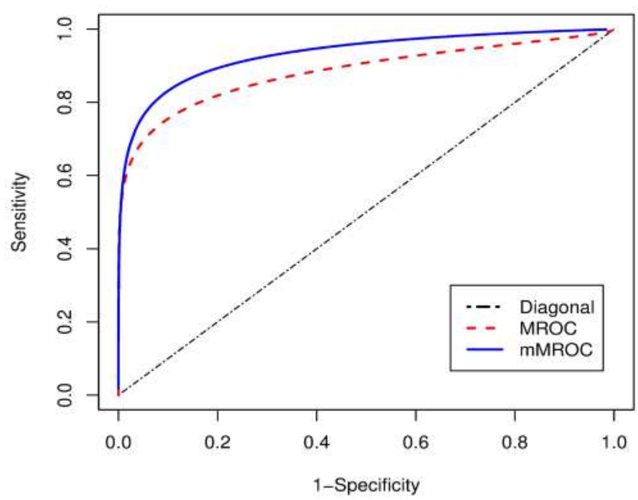

Figure 8. The MROC and mMROC curves for Disk Hernia data set

Simulation studies

Simulation studies are carried out for the proposed mMROC using bivariate normal random variables (X_1 and X_2). In table (7), four sets of mean vectors and covariance matrices were considered along with the initial mixing proportions to depict the behaviour of the proposed mMROC. Of these, the first two sets (A & B) are with unequal covariances matrices and the last two sets (C & D) are with equal covariance matrices. In each population, random samples of size n = {25,50,100,250,500} were generated.

Table (8) has the estimated mean vectors and covariance matrices of MROC model at different sample sizes. The estimated mixing proportions, mean vectors and covariance matrices of the proposed mMROC model are reported in table (9).

Table (10) and (11) depicts the linear combinations, cut-off points and measures of MROC and mMROC curves for various sample sizes at t = 0.5.

Furthermore, the AUC comparisons for simulated studies at different sample sizes

Table 8. Mixing Proportions, Mean vectors and Covariance matrices for simulation studies

Sets	λ_1	λ_2	μ_H	μ_{D_1}	μ_{D_2}	Σ_H	Σ_{D_1}	Σ_{D_2}
1	0.5	0.5	$\begin{pmatrix} 10.1 \\ 6.7 \end{pmatrix}$	$\begin{pmatrix} 13.2 \\ 9.8 \end{pmatrix}$	$\begin{pmatrix} 21.8 \\ 12.1 \end{pmatrix}$	$\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$	$\begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$	$\begin{pmatrix} 2 & 1 \\ 1 & 6 \end{pmatrix}$
2	0.5	0.5	$\begin{pmatrix} 10.1 \\ 6.7 \end{pmatrix}$	$\begin{pmatrix} 10.1 \\ 6.7 \end{pmatrix}$	$\begin{pmatrix} 10.1 \\ 6.7 \end{pmatrix}$	$\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$	$\begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$	$\begin{pmatrix} 2 & 1 \\ 1 & 6 \end{pmatrix}$
3	0.5	0.5	$\begin{pmatrix} 10.1 \\ 6.7 \end{pmatrix}$	$\begin{pmatrix} 13.2 \\ 9.8 \end{pmatrix}$	$\begin{pmatrix} 21.8 \\ 12.1 \end{pmatrix}$	$\begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$	$\begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$	$\begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$
4	0.5	0.5	$\begin{pmatrix} 10.1 \\ 6.7 \end{pmatrix}$	$\begin{pmatrix} 10.1 \\ 6.7 \end{pmatrix}$	$\begin{pmatrix} 10.1 \\ 6.7 \end{pmatrix}$	$\begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$	$\begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$	$\begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$

The table depicts the four sets of means and covariance matrices where the first two sets have unequal covariance matrices while the other have equal covariance matrices.

are carried out and are reported in table (13). Results indicate significant outcomes meaning to the difference between the AUC's of MROC and mMROC models. This relates to that the AUC obtained through a mixture MROC with three components provides better information about the classifier by having more accuracy than the two component MROC model. Under simulations, the sets A & C will speak about the better classification scenario and sets B & D relates to worst class scenario. The AUC comparisons were found to be significant with sets A & C as they have minimum overlapping areas. Further, due to the maximum overlapping areas exhibited by the sets B & D resulted insignificant p-values where the AUC's are close to each other.

The MROC and mMROC curves at each parameter combination for varying sample sizes are shown in figures (9) and (10).

Summary

In this work, the main aim is to address on the issue of constructing an ROC model when there exists multi-model patterns in the known class labels. In medical scenario, most of data sets exhibit such patterns. In such situations, before we proceed for ROC modelling, it is suggested to look for such multi-model patterns that might exist in the data, if any. An illustration of this kind is demonstrated and to model such patterns, mUROC and mMROC models are proposed. The practical applicability and necessary illustrations of these models are given using OGTT, Disk Hernia data sets and simulation studies. Based on the results, it is understood that proposed ROC forms notifies better accuracy when compared with Bi-Normal and MROC models. Basically, the Bi-Normal ROC model and Multivariate ROC models were developed for a two –class classifications, so when we have a multi-class situation, one cannot rely on these models. So, in such cases, the proposed mUROC and mMROC models will be of great help and the graphical depiction of the ROC curves can be presented in a 2-dimensional space rather than with a complex higher dimension. On the whole, it is always suggested that before proceeding to modelling ROC curves, it is good to take a look at the density patterns of the study variable(s), which in turn will help in explaining the true information between the classes and also provides good amount of "true" accuracy about the marker(s).

Table 9. Estimated means and covariance matrices of MROC model at various sample sizes

Set	Sample size	MROC parameters			
		$\hat{\mu}_H$	$\hat{\mu}_D$	$\hat{\Sigma}_H$	$\hat{\Sigma}_D$
A	25	$\begin{pmatrix} 10.10454 \\ 6.69076 \end{pmatrix}$	$\begin{pmatrix} 17.66999 \\ 10.98911 \end{pmatrix}$	$\begin{pmatrix} 2.00686 & 1.01319 \\ 1.01319 & 2.99530 \end{pmatrix}$	$\begin{pmatrix} 21.10063 & 6.07393 \\ 6.07393 & 6.46281 \end{pmatrix}$
	50	$\begin{pmatrix} 10.09904 \\ 6.70402 \end{pmatrix}$	$\begin{pmatrix} 17.49570 \\ 10.96989 \end{pmatrix}$	$\begin{pmatrix} 1.99840 & 0.99545 \\ 0.99545 & 3.01684 \end{pmatrix}$	$\begin{pmatrix} 20.82954 & 6.02989 \\ 6.02989 & 6.34245 \end{pmatrix}$
	100	$\begin{pmatrix} 10.09583 \\ 6.70182 \end{pmatrix}$	$\begin{pmatrix} 17.50580 \\ 10.95309 \end{pmatrix}$	$\begin{pmatrix} 1.99281 & 0.99341 \\ 0.99341 & 2.98802 \end{pmatrix}$	$\begin{pmatrix} 20.62403 & 5.97472 \\ 5.97472 & 6.31814 \end{pmatrix}$
	250	$\begin{pmatrix} 10.09869 \\ 6.70271 \end{pmatrix}$	$\begin{pmatrix} 17.49997 \\ 10.95445 \end{pmatrix}$	$\begin{pmatrix} 1.98469 & 0.98826 \\ 0.98826 & 2.98348 \end{pmatrix}$	$\begin{pmatrix} 20.55593 & 5.95532 \\ 5.95532 & 6.32159 \end{pmatrix}$
	500	$\begin{pmatrix} 10.10112 \\ 6.70264 \end{pmatrix}$	$\begin{pmatrix} 17.50082 \\ 10.95650 \end{pmatrix}$	$\begin{pmatrix} 1.99749 & 0.99903 \\ 0.99903 & 2.99258 \end{pmatrix}$	$\begin{pmatrix} 20.51221 & 5.93948 \\ 5.93948 & 6.31243 \end{pmatrix}$
B	25	$\begin{pmatrix} 10.09200 \\ 6.69816 \end{pmatrix}$	$\begin{pmatrix} 10.10787 \\ 6.68796 \end{pmatrix}$	$\begin{pmatrix} 1.98508 & 1.00712 \\ 1.00712 & 3.02619 \end{pmatrix}$	$\begin{pmatrix} 1.99265 & 0.97685 \\ 0.97685 & 5.10118 \end{pmatrix}$
	50	$\begin{pmatrix} 10.10507 \\ 6.69237 \end{pmatrix}$	$\begin{pmatrix} 10.09890 \\ 6.70950 \end{pmatrix}$	$\begin{pmatrix} 2.00315 & 1.00625 \\ 1.00625 & 2.98878)) \end{pmatrix}$	$\begin{pmatrix} 1.98671 & 0.99043 \\ 0.99043 & 5.00878 \end{pmatrix}$
	100	$\begin{pmatrix} 10.10062 \\ 6.69737 \end{pmatrix}$	$\begin{pmatrix} 10.09922 \\ 6.69334 \end{pmatrix}$	$\begin{pmatrix} 2.00079 & 1.00461 \\ 1.00461 & 3.02001 \end{pmatrix}$	$\begin{pmatrix} 2.00555 & 1.00879 \\ 1.00879 & 5.01907 \end{pmatrix}$
	250	$\begin{pmatrix} 10.09578 \\ 6.69398 \end{pmatrix}$	$\begin{pmatrix} 10.09922 \\ 6.69441 \end{pmatrix}$	$\begin{pmatrix} 2.00310 & 0.99800 \\ 0.99800 & 2.99502 \end{pmatrix}$	$\begin{pmatrix} 1.99577 & 0.99195 \\ 0.99195 & 4.99197 \end{pmatrix}$
	500	$\begin{pmatrix} 10.10159 \\ 6.69711 \end{pmatrix}$	$\begin{pmatrix} 10.10162 \\ 6.69796 \end{pmatrix}$	$\begin{pmatrix} 2.00576 & 0.99948 \\ 0.99948 & 2.99706 \end{pmatrix}$	$\begin{pmatrix} 1.99757 & 0.99724 \\ 0.99724 & 4.98736 \end{pmatrix}$
C	25	$\begin{pmatrix} 10.09550 \\ 6.69838 \end{pmatrix}$	$\begin{pmatrix} 17.65491 \\ 10.98152 \end{pmatrix}$	$\begin{pmatrix} 1.99751 & 1.00236 \\ 1.00236 & 3.99925 \end{pmatrix}$	$\begin{pmatrix} 21.47138 & 6.15339 \\ 6.15339 & 5.35069 \end{pmatrix}$
	50	$\begin{pmatrix} 10.10188 \\ 6.69930 \end{pmatrix}$	$\begin{pmatrix} 17.49161 \\ 10.93902 \end{pmatrix}$	$\begin{pmatrix} 1.98604 & 0.97591 \\ 0.97591 & 4.00349 \end{pmatrix}$	$\begin{pmatrix} 20.95016 & 6.05907 \\ 6.05907 & 5.33796 \end{pmatrix}$
	100	$\begin{pmatrix} 10.09579 \\ 6.69797)) \end{pmatrix}$	$\begin{pmatrix} 17.49404 \\ 10.95740 \end{pmatrix}$	$\begin{pmatrix} 1.99842 & 0.98288 \\ 0.98288 & 4.00100 \end{pmatrix}$	$\begin{pmatrix} 20.68148 & 5.98281 \\ 5.98281 & 5.31919 \end{pmatrix}$
	250	$\begin{pmatrix} 10.09810 \\ 6.70231 \end{pmatrix}$	$\begin{pmatrix} 17.49812 \\ 10.94947 \end{pmatrix}$	$\begin{pmatrix} 2.00667 & 1.01494 \\ 1.01494 & 3.99923 \end{pmatrix}$	$\begin{pmatrix} 20.55439 & 5.95898 \\ 5.95898 & 5.31075 \end{pmatrix}$
	500	$\begin{pmatrix} 10.10036 \\ 6.70202 \end{pmatrix}$	$\begin{pmatrix} 17.50165 \\ 10.94792 \end{pmatrix}$	$\begin{pmatrix} 1.99661 & 0.99025 \\ 0.99025 & 3.98883 \end{pmatrix}$	$\begin{pmatrix} 20.52614 & 5.95700 \\ 5.95700 & 5.34308 \end{pmatrix}$
D	25	$\begin{pmatrix} 10.10083 \\ 6.71296 \end{pmatrix}$	$\begin{pmatrix} 10.10859 \\ 6.72491 \end{pmatrix}$	$\begin{pmatrix} 1.96491 & 1.01037 \\ 1.01037 & 3.99736 \end{pmatrix}$	$\begin{pmatrix} 1.99364 & 1.00184 \\ 1.00184 & 3.99035 \end{pmatrix}$
	50	$\begin{pmatrix} 10.11116 \\ 6.69497 \end{pmatrix}$	$\begin{pmatrix} 10.10245 \\ 6.69943 \end{pmatrix}$	$\begin{pmatrix} 2.00147 & 1.01289 \\ 1.01289 & 3.99149 \end{pmatrix}$	$\begin{pmatrix} 2.00248 & 0.99474 \\ 0.99474 & 3.93676 \end{pmatrix}$
	100	$\begin{pmatrix} 10.10983 \\ 6.70429 \end{pmatrix}$	$\begin{pmatrix} 10.09890 \\ 6.69710 \end{pmatrix}$	$\begin{pmatrix} 1.99516 & 0.98727 \\ 0.98727 & 3.97426 \end{pmatrix}$	$\begin{pmatrix} 2.01451 & 1.00626 \\ 1.00626 & 4.00825 \end{pmatrix}$
	250	$\begin{pmatrix} 10.09920 \\ 6.69115 \end{pmatrix}$	$\begin{pmatrix} 10.09907 \\ 6.70065 \end{pmatrix}$	$\begin{pmatrix} 1.99713 & 1.00714 \\ 1.00714 & 4.01377 \end{pmatrix}$	$\begin{pmatrix} 1.99593 & 0.99290 \\ 0.99290 & 3.98825 \end{pmatrix}$
	500	$\begin{pmatrix} 10.09996 \\ 6.69683 \end{pmatrix}$	$\begin{pmatrix} 10.10075 \\ 6.69958 \end{pmatrix}$	$\begin{pmatrix} 2.00361 & 0.99686 \\ 0.99686 & 3.99104 \end{pmatrix}$	$\begin{pmatrix} 1.99583 & 0.99669 \\ 0.99669 & 3.98328 \end{pmatrix}$

The table depicts the estimates of Mean vectors and Covariance matrices of the four sets considered for simulation studies at various samples sizes.

Table 10. Estimated mixing proportions, means and covariance matrices of proposed mMROC model at various sample sizes

Set	Sample size	mMROC parameters							
		$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\mu}_H$	$\hat{\mu}_{D_1}$	$\hat{\mu}_{D_2}$	$\hat{\Sigma}_H$	$\hat{\Sigma}_{D_1}$	$\hat{\Sigma}_{D_2}$
	25	0.49127	0.50873	$\begin{pmatrix}10.10454\\6.69076\end{pmatrix}$	$\begin{pmatrix}13.21104\\9.80754\end{pmatrix}$	$\begin{pmatrix}21.78593\\12.07980\end{pmatrix}$	$\begin{pmatrix}2.00686 & 1.01319\\1.01319 & 2.99530\end{pmatrix}$	$\begin{pmatrix}1.98524 & 0.98058\\0.98058 & 3.97023\end{pmatrix}$	$\begin{pmatrix}1.96485 & 1.02805\\1.02805 & 6.18519\end{pmatrix}$
	50	0.49927	0.50073	$\begin{pmatrix}10.09904\\6.70402\end{pmatrix}$	$\begin{pmatrix}13.20086\\9.82320\end{pmatrix}$	$\begin{pmatrix}21.79054\\12.11657\end{pmatrix}$	$\begin{pmatrix}1.99840 & 0.99545\\0.99545 & 3.01684\end{pmatrix}$	$\begin{pmatrix}1.99731 & 0.99534\\0.99534 & 3.96877\end{pmatrix}$	$\begin{pmatrix}2.01656 & 1.01002\\1.01002 & 6.02197\end{pmatrix}$
A	100	0.49929	0.50071	$\begin{pmatrix}10.09583\\6.70182\end{pmatrix}$	$\begin{pmatrix}13.21230\\9.80711\end{pmatrix}$	$\begin{pmatrix}21.79931\\12.09907\end{pmatrix}$	$\begin{pmatrix}1.99281 & 0.99341\\0.99341 & 2.98802\end{pmatrix}$	$\begin{pmatrix}1.99454 & 0.99022\\0.99022 & 3.98587\end{pmatrix}$	$\begin{pmatrix}2.01246 & 1.02165\\1.02165 & 5.99822\end{pmatrix}$
	250	0.49965	0.50035	$\begin{pmatrix}10.09869\\6.70271\end{pmatrix}$	$\begin{pmatrix}13.20145\\9.80519\end{pmatrix}$	$\begin{pmatrix}21.79850\\12.10371\end{pmatrix}$	$\begin{pmatrix}1.98469 & 0.98826\\0.98826 & 2.98348\end{pmatrix}$	$\begin{pmatrix}2.00966 & 0.99165\\0.99165 & 3.99214\end{pmatrix}$	$\begin{pmatrix}1.99881 & 0.99844\\0.99844 & 6.00218\end{pmatrix}$
	500	0.49998	0.50002	$\begin{pmatrix}10.10112\\6.70264\end{pmatrix}$	$\begin{pmatrix}13.20264\\9.80916\end{pmatrix}$	$\begin{pmatrix}21.79900\\12.10383\end{pmatrix}$	$\begin{pmatrix}1.99749 & 0.99903\\0.99903 & 2.99258\end{pmatrix}$	$\begin{pmatrix}1.99987 & 0.98793\\0.98793 & 3.99468\end{pmatrix}$	$\begin{pmatrix}2.00201 & 1.00886\\1.00886 & 5.99283\end{pmatrix}$
	25	0.50093	0.49907	$\begin{pmatrix}10.09200\\6.69816\end{pmatrix}$	$\begin{pmatrix}10.11433\\6.70119\end{pmatrix}$	$\begin{pmatrix}10.10190\\6.67575\end{pmatrix}$	$\begin{pmatrix}1.98508 & 1.00712\\1.00712 & 3.02619\end{pmatrix}$	$\begin{pmatrix}2.00313 & 1.00074\\1.00074 & 3.99882\end{pmatrix}$	$\begin{pmatrix}1.97246 & 0.95437\\0.95437 & 6.11482\end{pmatrix}$
	50	0.50105	0.49895	$\begin{pmatrix}10.10507\\6.69237\end{pmatrix}$	$\begin{pmatrix}10.10010\\6.72143\end{pmatrix}$	$\begin{pmatrix}10.09769\\6.69758\end{pmatrix}$	$\begin{pmatrix}2.00315 & 1.00625\\1.00625 & 2.98878\end{pmatrix}$	$\begin{pmatrix}1.98096 & 0.99777\\0.99777 & 4.05185\end{pmatrix}$	$\begin{pmatrix}1.99221 & 0.98156\\0.98156 & 5.96817\end{pmatrix}$
B	100	0.50843	0.49157	$\begin{pmatrix}10.10062\\6.69737\end{pmatrix}$	$\begin{pmatrix}10.09895\\6.69230\end{pmatrix}$	$\begin{pmatrix}10.09949\\6.69438\end{pmatrix}$	$\begin{pmatrix}2.00079 & 1.00461\\1.00461 & 3.02001\end{pmatrix}$	$\begin{pmatrix}1.99755 & 1.00190\\1.00190 & 4.01536\end{pmatrix}$	$\begin{pmatrix}2.01510 & 1.01556\\1.01556 & 6.02106\end{pmatrix}$
	250	0.50501	0.49499	$\begin{pmatrix}10.09578\\6.69398\end{pmatrix}$	$\begin{pmatrix}10.09937\\6.69374\end{pmatrix}$	$\begin{pmatrix}10.09906\\6.69509\end{pmatrix}$	$\begin{pmatrix}2.00310 & 0.99800\\0.99800 & 2.99502\end{pmatrix}$	$\begin{pmatrix}2.00751 & 0.99489\\0.99489 & 3.98802\end{pmatrix}$	$\begin{pmatrix}1.98479 & 0.98985\\0.98985 & 5.99721\end{pmatrix}$
	500	0.50860	0.49140	$\begin{pmatrix}10.10159\\6.69711\end{pmatrix}$	$\begin{pmatrix}10.10154\\6.69905\end{pmatrix}$	$\begin{pmatrix}10.10169\\6.69688\end{pmatrix}$	$\begin{pmatrix}2.00576 & 0.99948\\0.99948 & 2.99706\end{pmatrix}$	$\begin{pmatrix}1.99830 & 1.00449\\1.00449 & 3.98450\end{pmatrix}$	$\begin{pmatrix}1.99688 & 0.99017\\0.99017 & 5.98912\end{pmatrix}$
	25	0.49092	0.50908	$\begin{pmatrix}10.09550\\6.69838\end{pmatrix}$	$\begin{pmatrix}13.15651\\9.79545\end{pmatrix}$	$\begin{pmatrix}21.80729\\12.07636\end{pmatrix}$	$\begin{pmatrix}1.99751 & 1.00236\\1.00236 & 3.99925\end{pmatrix}$	$\begin{pmatrix}1.97665 & 0.96691\\0.96691 & 3.97046\end{pmatrix}$	$\begin{pmatrix}2.04887 & 1.07399\\1.07399 & 4.03201\end{pmatrix}$
	50	0.49802	0.50198	$\begin{pmatrix}10.10188\\6.69930\end{pmatrix}$	$\begin{pmatrix}13.18281\\9.78740\end{pmatrix}$	$\begin{pmatrix}21.80040\\12.09065\end{pmatrix}$	$\begin{pmatrix}1.98604 & 0.97591\\0.97591 & 4.00349\end{pmatrix}$	$\begin{pmatrix}2.00042 & 1.00927\\1.00927 & 4.05739\end{pmatrix}$	$\begin{pmatrix}2.00497 & 0.97327\\0.97327 & 3.90720\end{pmatrix}$
C	100	0.49907	0.50093	$\begin{pmatrix}10.09579\\6.69797\end{pmatrix}$	$\begin{pmatrix}13.19387\\9.80767\end{pmatrix}$	$\begin{pmatrix}21.79421\\12.10713\end{pmatrix}$	$\begin{pmatrix}1.99842 & 0.98288\\0.98288 & 4.00100\end{pmatrix}$	$\begin{pmatrix}2.00003 & 0.98627\\0.98627 & 3.97636\end{pmatrix}$	$\begin{pmatrix}2.00741 & 0.98873\\0.98873 & 3.98610\end{pmatrix}$
	250	0.49923	0.50077	$\begin{pmatrix}10.09810\\6.70231\end{pmatrix}$	$\begin{pmatrix}13.19969\\9.79946\end{pmatrix}$	$\begin{pmatrix}21.79656\\12.09948\end{pmatrix}$	$\begin{pmatrix}2.00667 & 1.01494\\1.01494 & 3.99923\end{pmatrix}$	$\begin{pmatrix}2.01018 & 0.99708\\0.99708 & 3.98460\end{pmatrix}$	$\begin{pmatrix}1.99708 & 0.99388\\0.99388 & 3.98274\end{pmatrix}$
	500	0.50112	0.49888	$\begin{pmatrix}10.10036\\6.70202\end{pmatrix}$	$\begin{pmatrix}13.20202\\9.79707\end{pmatrix}$	$\begin{pmatrix}21.80128\\12.09877\end{pmatrix}$	$\begin{pmatrix}1.99661 & 0.99025\\0.99025 & 3.98883\end{pmatrix}$	$\begin{pmatrix}2.00753 & 0.99785\\0.99785 & 4.01736\end{pmatrix}$	$\begin{pmatrix}1.99722 & 1.00055\\1.00055 & 4.01531\end{pmatrix}$
	25	0.50114	0.49886	$\begin{pmatrix}10.10083\\6.71296\end{pmatrix}$	$\begin{pmatrix}10.09875\\6.73157\end{pmatrix}$	$\begin{pmatrix}10.11768\\6.71876\end{pmatrix}$	$\begin{pmatrix}1.96491 & 1.01037\\1.01037 & 3.99736\end{pmatrix}$	$\begin{pmatrix}2.00809 & 1.01407\\1.01407 & 3.98593\end{pmatrix}$	$\begin{pmatrix}1.98168 & 0.98925\\0.98925 & 3.98779\end{pmatrix}$
	50	0.50121	0.49879	$\begin{pmatrix}10.11116\\6.69497\end{pmatrix}$	$\begin{pmatrix}10.11027\\6.70019\end{pmatrix}$	$\begin{pmatrix}10.09463\\6.69867\end{pmatrix}$	$\begin{pmatrix}2.00147 & 1.01289\\1.01289 & 3.99149\end{pmatrix}$	$\begin{pmatrix}2.01250 & 1.00685\\1.00685 & 3.92122\end{pmatrix}$	$\begin{pmatrix}1.99582 & 0.98459\\0.98459 & 3.95430\end{pmatrix}$
D	100	0.50130	0.49870	$\begin{pmatrix}10.10983\\6.70429\end{pmatrix}$	$\begin{pmatrix}10.09784\\6.70444\end{pmatrix}$	$\begin{pmatrix}10.09996\\6.68976\end{pmatrix}$	$\begin{pmatrix}1.99516 & 0.98727\\0.98727 & 3.97426\end{pmatrix}$	$\begin{pmatrix}2.01704 & 0.99471\\0.99471 & 3.99336\end{pmatrix}$	$\begin{pmatrix}2.01485 & 1.02091\\1.02091 & 4.02971\end{pmatrix}$
	250	0.50239	0.49761	$\begin{pmatrix}10.09920\\6.69115\end{pmatrix}$	$\begin{pmatrix}10.10184\\6.69570\end{pmatrix}$	$\begin{pmatrix}10.09630\\6.70561\end{pmatrix}$	$\begin{pmatrix}1.99713 & 1.00714\\1.00714 & 4.01377\end{pmatrix}$	$\begin{pmatrix}1.98638 & 0.98227\\0.98227 & 3.99369\end{pmatrix}$	$\begin{pmatrix}2.00607 & 1.00420\\1.00420 & 3.98474\end{pmatrix}$
	500	0.50233	0.49768	$\begin{pmatrix}10.09996\\6.69683\end{pmatrix}$	$\begin{pmatrix}10.10272\\6.70330\end{pmatrix}$	$\begin{pmatrix}10.09878\\6.69587\end{pmatrix}$	$\begin{pmatrix}2.00361 & 0.99686\\0.99686 & 3.99104\end{pmatrix}$	$\begin{pmatrix}1.99838 & 0.99824\\0.99824 & 3.97943\end{pmatrix}$	$\begin{pmatrix}1.99366 & 0.99477\\0.99477 & 3.98684\end{pmatrix}$

The table depicts the estimates of mixing proportions, Mean vectors and Covariance matrices of the four sets considered for simulation studies at various samples sizes.

Table 11. The linear combinations, cutoff points and measures of MROC for various sample sizes at

Set	Sample size	MROC measures				
		U	c	FPR	TPR	J
A	25	$0.49503^{*}x_1 + 0.59497^{*}x_2$	10.92129	0.09030	0.90970	0.81940
	50	$0.48277^{*}x_1 + 0.57865^{*}x_2$	10.61223	0.09459	0.90541	0.81081
	100	$0.48848^{*}x_1 + 0.56082^{*}x_2$	10.52059	0.09454	0.90546	0.81092
	250	$0.48841^{*}x_1 + 0.55482^{*}x_2$	10.46879	0.09482	0.90518	0.81036
	500	$0.48824^{*}x_1 + 0.55245^{*}x_2$	10.45343	0.09503	0.90497	0.80994
B	25	$0.01161^{*}x_1 - 0.00363^{*}x_2$	0.09753	0.42827	0.57173	0.14346
	50	$-0.0074^{*}x_1 + 0.007456^{*}x_2$	-0.02238	0.44859	0.55141	0.10281
	100	$-0.00012^{*}x_1 - 0.00076^{*}x_2$	-0.00498	0.46483	0.53517	0.07034
	250	$0.00184^{*}x_1 - 0.00013^{*}x_2$	0.01831	0.47736	0.52264	0.04527
	500	$-0.00007^{*}x_1 + 0.00015^{*}x_2$	-0.00402	0.48397	0.51603	0.03205
C	25	$0.47911^{*}x_1 + 0.60050^{*}x_2$	10.88749	0.09416	0.90584	0.81169
	50	$0.480569^{*}x_1 + 0.5695^{*}x_2$	10.61016	0.09813	0.90187	0.80374
	100	$0.48418^{*}x_1 + 0.56287^{*}x_2$	10.60277	0.09764	0.90236	0.80472
	250	$0.48759^{*}x_1 + 0.55150^{*}x_2$	10.55049	0.09807	0.90193	0.80386
	500	$0.48937^{*}x_1 + 0.54786^{*}x_2$	10.53425	0.09786	0.90214	0.80427
D	25	$0.00609^{*}x_1 + 0.00190^{*}x_2$	0.10743	0.42781	0.57219	0.14438
	50	$-0.00535^{*}x_1 + 0.00227^{*}x_2$	-0.03975	0.45048	0.54952	0.09904
	100	$-0.00498^{*}x_1 - 0.00032^{*}x_2$	-0.05225	0.46457	0.53543	0.07085
	250	$-0.00113^{*}x_1 + 0.00254^{*}x_2$	0.00521	0.47755	0.52245	0.04490
	500	$0.00008^{*}x_1 + 0.00065^{*}x_2$	0.00520	0.48425	0.51575	0.03150

The table contains the best linear combinations and AUC measures of the MROC model for simulation studies at various samples sizes.

Table 12. The linear combinations, cutoff points and measures of mMROC for various sample sizes at

Set	Sample size	mMROC measures						
		U_1	U_2	τ_1	τ_2	FPR	TPR	J_M
A	25	$1.49517^*x_1 + 0.59813^*x_2$	$5.23432^*x_1 - 0.53711^*x_2$	22.54886	85.80980	0.05549	0.94451	0.88901
	50	$1.36389^*x_1 + 0.56187^*x_2$	$4.77574^*x_1 - 0.48233^*x_2$	20.57286	78.23666	0.05726	0.94274	0.88548
	100	$1.34323^*x_1 + 0.53634^*x_2$	$4.64862^*x_1 - 0.46568^*x_2$	20.08756	76.18369	0.05770	0.94230	0.88460
	250	$1.30705^*x_1 + 0.52759^*x_2$	$4.55944^*x_1 - 0.44248^*x_2$	19.55593	74.92880	0.05866	0.94134	0.88269
	500	$1.29827^*x_1 + 0.52402^*x_2$	$4.54416^*x_1 - 0.44578^*x_2$	19.42758	74.60272	0.05892	0.94108	0.88216
B	25	$0.01557^*x_1 + 0.00120^*x_2$	$-0.00148^*x_1 - 0.00398^*x_2$	0.22413	-0.03913	0.40087	0.59913	0.19826
	50	$-0.01024^*x_1 + 0.01111^*x_2$	$0.00141^*x_1 - 0.00310^*x_2$	-0.00584	-0.00044	0.43259	0.56741	0.13483
	100	$0.00121^*x_1 - 0.00179^*x_2$	$-0.00124^*x_1 + 0.00090^*x_2$	0.01106	-0.00308	0.45294	0.54706	0.09412
	250	$0.00202^*x_1 - 0.00018^*x_2$	$0.00102^*x_1 + 0.00013^*x_2$	0.02381	0.00169	0.47001	0.52999	0.05998
	500	$-0.00018^*x_1 + 0.00063^*x_2$	$0.00036^*x_1 - 0.00064^*x_2$	-0.00526	0.00611	0.47868	0.52132	0.04265
C	25	$1.51072^*x_1 + 0.51681^*x_2$	$5.27942^*x_1 - 0.71801^*x_2$	22.02856	84.23608	0.05761	0.94239	0.88479
	50	$1.40755^*x_1 + 0.46361^*x_2$	$4.87680^*x_1 - 0.60704^*x_2$	20.27628	78.66009	0.05941	0.94059	0.88117
	100	$1.36271^*x_1 + 0.46260^*x_2$	$4.70298^*x_1 - 0.57506^*x_2$	19.71705	75.95861	0.05947	0.94053	0.88105
	250	$1.33538^*x_1 + 0.44649^*x_2$	$4.62757^*x_1 - 0.57342^*x_2$	19.25432	74.73332	0.06041	0.93959	0.87919
	500	$1.33518^*x_1 + 0.44507^*x_2$	$4.60400^*x_1 - 0.56903^*x_2$	19.23025	74.37660	0.06033	0.93967	0.87934
D	25	$-0.00264^*x_1 + 0.00735^*x_2$	$0.01219^*x_1 - 0.01394^*x_2$	0.23539	-0.21874	0.40334	0.59666	0.19332
	50	$0.00039^*x_1 + 0.00192^*x_2$	$-0.00699^*x_1 - 0.00046^*x_2$	0.03735	-0.07294	0.43399	0.56601	0.13202
	100	$-0.00653^*x_1 + 0.00137^*x_2$	$0.00322^*x_1 - 0.00331^*x_2$	-0.04720	0.01146	0.45404	0.54596	0.09192
	250	$0.00112^*x_1 + 0.00082^*x_2$	$-0.00423^*x_1 + 0.00363^*x_2$	0.02053	-0.01882	0.47070	0.52930	0.05860
	500	$0.00069^*x_1 + 0.00152^*x_2$	$-0.00151^*x_1 - 0.00157^*x_2$	0.01925	-0.02596	0.47928	0.52072	0.04144

The table contains the best linear combinations and AUC measures of the mMROC model for simulation studies at various samples sizes.

Table 13.The estimated AUC's of MROC and mMROC curve along with values at various sample sizes

Set	Sample size	$\widehat{AUC}$	$m\widehat{UAUC}$	Z_{AUC}	$S_{ig.}$
A	25	0.95945	0.97649	1.66738	0.04878*
	50	0.95752	0.97684	2.05329	0.02002*
	100	0.95779	0.97710	2.86499	0.00209*
	250	0.95776	0.97682	4.45766	0.00000*
	500	0.95778	0.97676	6.41528	0.00000*
B	25	0.59953	0.63591	0.04851	0.48065NS
	50	0.57187	0.59391	0.03286	0.48689NS
	100	0.54937	0.56600	0.26207	0.39663NS
	250	0.53184	0.54222	0.01249	0.4950NS
	500	0.52256	0.53005	0.01203	0.49520NS
C	25	0.95859	0.97517	1.65011	0.04946*
	50	0.95667	0.97528	1.83489	0.03326*
	100	0.95757	0.97587	1.92829	0.02691*
	250	0.95750	0.97557	2.61752	0.00443*
	500	0.95759	0.97576	3.66464	0.00012*
D	25	0.60044	0.63270	0.01908	0.49239NS
	50	0.56952	0.59206	0.00551	0.49780NS
	100	0.54992	0.56457	0.00344	0.50137NS
	250	0.53170	0.54133	0.00105	0.50042NS
	500	0.52226	0.52926	0.00002	0.50001NS

*Significant;
NS, Not significant

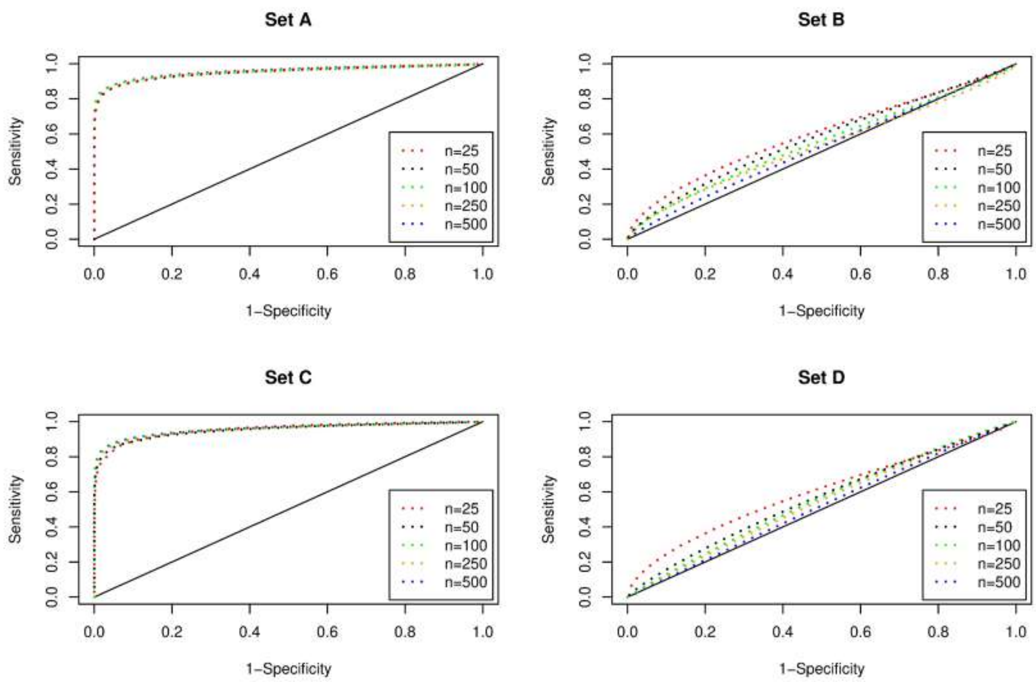

Figure 9. MROC Curves for the simulated data sets at various sample sizes
The MROC curves for the four sets are shown here. The first and third sets are the examples of best cases while the second and fourth are the worst cases of classification.

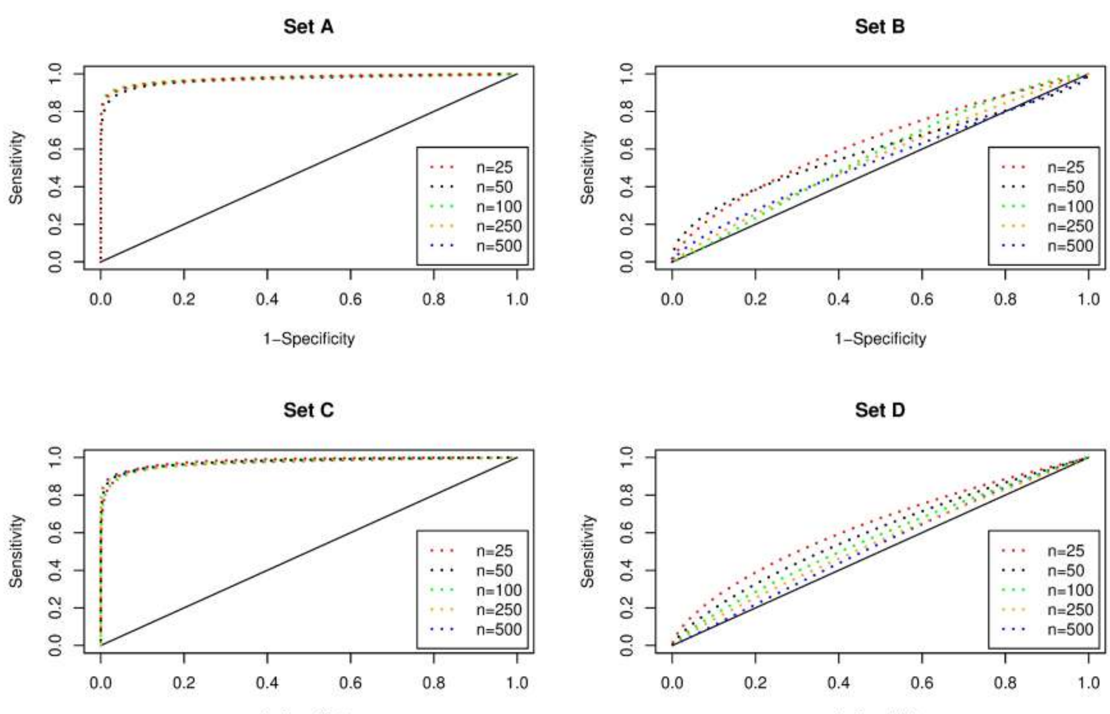

Figure 10. mMROC Curves for the simulated data sets at various sample sizes
The mMROC curves for the four sets are shown here. The first and third sets are the examples of best cases while the second and fourth are the worst cases of classification.

References

1. Green DM, Swets JA. Signal detection theory and psychophysics. New York: Wiley. 1966.

2. Egan JP. Signal Detection Theory and ROC Analysis "Academic Press. New York. 1975.

3. England WL. An exponential model used for optimal threshold selection on ROC Curves. Med Deci Mak, 1988;8:120-131.

4. Campbell G, Ratnaparkhi MV. An application of Lomax distributions in receiver operating characteristic (ROC) curve analysis. Comm in Stat -Theory and Methods. 1993;22:1681-1687.

5. Hussain E. The bi-gamma ROC curve in a straightforward manner. J Bas & App Sci. 2012;8:309-314.

6. Mossman D, Peng H. Using dual beta distributions to create "proper" ROC curves based on rating category data. Med Deci Mak. 2016;36:349-365.

7. Balaswamy S, Vardhan RV, Sarma, KVS. The Hybrid ROC Curve and its Divergence Measures for Binary Classification. Int J Stat in Med Res. 2015a;4(1):94–102.

8. Balaswamy S, Vardhan RV. Estimation of Confidence Intervals of a GHROC Curve in the presence of Scale and Shape parameters. Res J Math and Stat Sci. 2015b;3(10):4-11.

9. Balaswamy S, Vardhan RV. ROC curve Estimation using the combination of Generalized Half Normal and Weibull distributions. J Indian Society for Prob and Stat. 2016a;17(1):11-23.

10. Balaswamy S, Vishnu Vardhan R. An Anthology of Parametric ROC Models. Research & Reviews: J Sta. 2016b;5(2):32-46.

11. Su JQ, Liu JS. Linear combinations of multiple diagnostic markers. J Ame Stat Assoc. 1993;88:1350-1355.

12. Schisterman EF, Faraggi D, Reiser B. Adjusting the generalized ROC curve for covariates. Stat in Med. 2004;23:3319-3331.

13. Yuan Z, Ghosh D. Combining multiple biomarker models in logistic regression. Biometrics. 2008;64:431-439.

14. Yin J, Tian L. Optimal linear combinations of multiple diagnostic biomarkers based on Youden index. Stat in Med. 2014;33:1426-1440.

15. Sameera G, Vardhan, RV, Sarma KVS. Binary classification using multivariate receiver operating characteristic curve for continuous data. J Biopha Stat. 2016;26:421-431.

16. Lasko TA, Bhagwat JG, Zou KH, Ohno-Machado L. The use of receiver operating characteristic curves in biomedical informatics. J Biomed Inf. 2005;38(5):404-415.

17. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J Royal Stat Soc: Ser B (Meth). 1977;39:1-22.

18. Titterington DM, Smith AFM, Makov UE. Statistical analysis of finite mixture distributions (Vol. 198). John Wiley & Sons Inc. 1985.

19. McLachlan G, Peel D. Finite Mixture Models. John Wiley and Sons, Inc., New York. 2000.

20. Krzanowski WJ, Hand DJ. ROC curves for continuous data. Crc Press. 2009.

21. Anderson TW, Bahadur RR. Classification into two multivariate normal distributions with different covariance matrices. The Ann Math Sta. 1962;33:420-431.

22. Mossman D. Three-way rocs. Med Deci Mak. 1999;19:78-89.

23. Srinivasan A. Note on the location of optimal classifiers in n-dimensional ROC space. 1999.

24. Hand DJ, Till RJ. A simple generalisation of the area under the ROC curve for multiple class classification problems. Mach lear, 2001;45:171-186.

25. Ferri C, Hernandez-Orallo J, Salido MA. Volume under the ROC surface for multi-class problems. In Eur conf mach lear, Spri, Berlin, Heidelberg, 2003:108-120.

26. He X, Frey EC. The meaning and use of the volume under a three-class ROC surface (VUS). IEEE Trans Med Imag. 2008;27(5):577-588.

27. Kang L, Tian L. Estimation of the volume under the ROC surface with three ordinal diagnostic categories. Comp Stat & Dat Ana. 2013;62:39-51.

28. Liu S, Zhu H, Yi K, Sun X, Xu W, Wang C. Fast and Unbiased Estimation of Volume Under Ordered Three-Class ROC Surface (VUS) With Continuous or Discrete Measurements. IEEE Access. 2020;8:136206-136222.

29. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3:32-35.

30. Perkins NJ, Schisterman EF. The Youden Index and the optimal cut-point corrected for measurement error. J Math Meth in Biosci. 2005;47:428-441.

31. Perkins NJ, Schisterman EF. The inconsistency of “optimal” cutpoints obtained using two criteria based on the receiver operating characteristic curve. Ame J Epid. 2006;163:670-675.

32. Leal12 J, Oliveira M, Sanches12 JM. Analysis of cut-off criteria in ROC curve for endarterectomy decision making. 2011.

33. da Rocha Neto AR, de Alencar Barreto G. On the application of ensembles of classifiers to the diagnosis of pathologies of the vertebral column: A comparative analysis. IEEE Lat Am Trans. 2009;7:487-496.