Statistics and Applications {ISSN 2454-7395 (online)}
Volume 24, No. 1, 2026 (New Series), pp 237–248
http://www.ssca.org.in/journal

ROC Curve for Binary Classification using X Lindley Distribution

Sandhya Singh[1] **and Saebugari Balaswamy**[2]

[1] *Department of Community Medicine, Bundelkhand Medical College, Madhya Pradesh*
[2] *Department of Statistics, Indira Gandhi National Tribal University, Madhya Pradesh*

Received: 15 May 2025; Revised: 29 September 2025; Accepted: 10 October 2025

Abstract

A lot of ROC models have been derived in the area of classification to classify the subjects using many distributional assumptions based upon the nature of the data. Though, there are many models in the literature, still such models are of need current situation to address the different needs of the nature of data due to the skewness or non normal behavior of the data in reality. Therefore, this paper addresses such an ROC model which consists of the X Lindley (XL) distribution and this distribution has more flexibility than other one-parameter distributions. An attempt has been made to develop an ROC model in classification when the healthy population is at the higher side than the abnormal/diseased population. Further, simulations as well as real data set have been used to find out area under the curve. The simulations are done with various parameter values of the distribution, which will represent the better, moderate and worst case of classification in ROC analysis. The X Lindley distribution can be used quite effectively in analyzing the data in classification and which is easy to make computations even for a non statistician. Further, the properties of the XL ROC curve are verified mathematically and the likelihood ratio test is also proposed and the relation is established with the slope of the ROC curve.

Key words: ROC curve; X Lindley distribution; AUC; Likelihood ratio test; Slope of the ROC curve; Optimal threshold.

AMS Subject Classifications: 62P10

1. Introduction

Various statistical methods have been developed to identify class labels, which is a primary goal in classification tasks. A typical challenge in classification involves assigning an individual to one of several predefined groups, such as healthy (normal) or diseased (abnormal). To address these challenges, tools like the Receiver Operating Characteristic (ROC) curve are commonly used. The ROC curve originated in the 1940s when electrical and radar engineers developed it for detecting enemy objects during World War II. Since then, ROC analysis has found widespread application across diverse fields, including medicine, radiol-

Corresponding Author: Saebugari Balaswamy
Email: balaswamy.stats@gmail.com

ogy, biometrics, natural hazard forecasting, and meteorology, and it has become increasingly important in machine learning and data mining. An ROC curve plots sensitivity against 1-specificity for a diagnostic test. Sensitivity refers to the probability of correctly identifying individuals with a disease, while specificity measures the probability of correctly identifying healthy individuals. Additionally, the area under the ROC curve (AUC) is used as a summary measure, reflecting the binary classifier's ability to differentiate between classes.

If the outcome S of a medical test is measured on a continuous scale, a specific threshold t can be set to classify individuals. For any diagnostic test, a person with a score $s \leq t$ can be classified as healthy (H), while those with a score above t are classified as diseased (D). Based on this classification, a 2×2 contingency table, also known as a confusion matrix, can be constructed. This matrix includes four possible outcomes: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

The probabilistic definitions are given below,

1. The probability that an individual from D is correctly classified.
   True positive rate, $TPR = P(S > t|D)$ Sensitivity

2. The probability that an individual from H is misclassified.
   False positive Rate, $FPR = P(S > t|H)$

3. The probability that an individual from is correctly classified.
   True negative rate, $TNR = P(S \leq t|H)$ Specificity

4. The probability that an individual from D is misclassified.
   False Negative Rate, $FNR = P(S \leq t|D)$ 1-Sensitivity.

Over time, a wide variety of ROC (Receiver Operating Characteristic) models have been introduced in the literature, particularly those relying on bi-distributional assumptions. The following are notable examples: Bilognormal (Dorfman and Alf, 1968); Binormal (Egan, 1975); Bibeta (Zou *et al.*, 1997); Bigamma (Hussain, 2012); Biexponential (Tang and Balakrishnan, 2011); Hybrid ROC Curve (Balaswamy *et al.*, 2015). Additionally, methods for estimating confidence intervals for ROC curves have been enhanced by applying distributions such as: Generalized Half Normal and Weibull (Balaswamy *et al.*, 2015). Further, innovative applications and extensions are developed in the recent past, which includes the analysis of the relationship between the ratio β and the area under the ROC curve (AUC) using combined Half Normal and Rayleigh distributions (Balaswamy and Vardhan, 2015); AUC estimation for non-normal data (Balaswamy and Vardhan, 2022); Estimation of AUC with the Bi-Generalised Exponential ROC curve (Dashina and Vishnu Vardhan, 2022); ROC curve area estimation within gamma mixture models (Arunima and Vishnu Vardhan, 2022); Multiclass ROC statistics and non-normal data applications (Arunima and Vishnu Vardhan, 2023); Comparative AUC estimation using generalized exponential distributions (Balaswamy and Vardhan, 2023). Despite the variety of ROC models based on different distributions, new models are required to address complex, real-world data scenarios. Motivated by the limitations in existing approaches and the need for increased flexibility, the present work introduces an ROC model founded on the X Lindley distribution. This choice is justified by ease of application and greater flexibility compared to other single-parameter distributions.

This new model aims to better accommodate diverse data characteristics encountered in practical classification problems where standard assumptions may not hold.

Therefore, the next section deals with the methodology, where the ROC model is developed using the X Lindley distribution.

2. Methodology

In this section, a mixture of two known distributions (Exponential and Lindley) used to give new distribution called X Lindley distribution. Let X be a random variable with the mixture Distribution viz, X Lindley distribution. The probability density function and the distribution functions of X Lindley distribution are as follows.

$$f_{XL}(x,\theta) = \frac{\theta^2(2+\theta+x)}{(1+\theta)^2}e^{-x\theta} \qquad ,x,\theta > 0 \tag{1}$$

$$F_{XL}(x,\theta) = 1 - \left[1 + \frac{x\theta}{(1+\theta)^2}\right]e^{-x\theta} \qquad ,x,\theta > 0 \tag{2}$$

Let the test scores (S) of normal (0) and abnormal (1) populations follows a X Lindley distribution. As the ROC curve is the tradeoff between Sensitivity and 1-Specificity, the FPR (1-Specificity) is defined as

$$FPR = x(t) = P(S > t|0) = \left[1 + \frac{t\theta_0}{(1+\theta_0)^2}\right]e^{-t\theta_0} \tag{3}$$

Where, t denotes threshold of a diagnostic test and θ_0 is the parameter of XL distribution from healthy (normal) population.

Applying log on both sides in (3), we get

$$\log(x(t)) = \log\left[1 + \frac{t\theta_0}{(1+\theta_0)^2}\right] - t(\theta_0)$$

$$\log(x(t)) = \frac{t\theta_0}{(1+\theta_0)^2} - t\theta_0$$

Further, the threshold can be obtained from the above equation as

$$t = \frac{(1+\theta_0)^2}{[\theta_0 - (1+\theta_0)^2\theta_0]}\log(x(t)) \tag{4}$$

Another intrinsic measure Sensitivity is defined as

$$TPR = Y(t) = P(S > t|0) = \left[1 + \frac{t\theta_1}{(1+\theta_1)^2}\right]e^{-t\theta_1} \tag{5}$$

where θ_1 is the parameter of XL distribution from diseased (abnormal) population and it is assumed to be lesser than the value of θ_0. The ROC curve can be obtained by substituting the above threshold value t in the y(t), we get

$$ROC(t) = \left[1 + \frac{\theta_1}{(1+\theta_1)^2}\frac{(1+\theta_0)^2}{[\theta_0 - (1+\theta_0)^2\theta_0]}\log(x(t))\right]e^{\frac{-\theta_1(1+\theta_0)^2}{\theta_0-(1+\theta_0)^2\theta_0}\log(x(t))} \tag{6}$$

On further simplification, the final expression for TPR is given by

$$ROC(t) = \left[1 + \frac{\theta_1}{\theta_0}\left(\frac{1}{\left(\frac{1+\theta_1}{1+\theta_0}\right)^2 - (1+\theta_1)^2}\log(x(t))\right)\right] e^{-\frac{\theta_1}{\theta_0}\left[\frac{1}{\frac{1}{(1+\theta_0)^2}-1}\right]\log(x(t))} \tag{7}$$

The above expression is defined as the ROC curve for X Lindley distribution and is named as XL ROC curve in binary classification. Along with the above two intrinsic measures namely Sensitivity and Specificity; Area under the Curve (AUC) is also an important measure of ROC curve and plays a prominent role in assessing the performance of a test. AUC takes values between 0 and 1. A perfect diagnostic test is one with an area equal to 1 and a test with an area 0 is perfectly inaccurate. The AUC of an ROC curve can be interpreted as the average of sensitivity for all possible values of specificity and vice versa. AUC of an ROC curve can be obtained by integrating the ROC expression over the range [0, 1] *i.e.*,

$$AUC = \int_0^1 ROC(t)dt$$

Further the area under the curve for XL ROC curve is given by

$$AUC = \int_0^1 \left[1 + \frac{\theta_1}{\theta_0}\left(\frac{1}{\left(\frac{1+\theta_1}{1+\theta_0}\right)^2 - (1+\theta_1)^2}\right)\log(x(t))\right] e^{-\frac{\theta_1}{\theta_0}\left[\frac{1}{\frac{1}{(1+\theta_0)^2}-1}\right]\log(x(t))} dx(t) \tag{8}$$

For the purpose of simplification, let us consider the above AUC in (8) as,

$$AUC = \int_0^1 \left\{1 + a\log(x(t))\right\} e^{-b\log(x(t))}dx(t) \tag{9}$$

Where a and b are defined as follows

$$a = \frac{\theta_1}{\theta_0}\left(\frac{1}{\left(\frac{1+\theta_1}{1+\theta_0}\right)^2 - (1+\theta_1)^2}\right)$$

$$b = \frac{\theta_1}{\theta_0}\left[\frac{1}{\frac{1}{(1+\theta_0)^2} - 1}\right]$$

Further, the AUC in (9) can be simplified as

$$AUC = \lim_{\epsilon\to 0}\int_\epsilon^1 \{1 + a\log(x(t))\}e^{-b\log(x(t))}dx(t)$$

$$AUC = \lim_{\epsilon\to 0}\int_\epsilon^1 \{1 + a\log(x(t))\}\frac{1}{x^b}dx(t)$$

On further simplification, we get

$$AUC = \lim_{\epsilon\to 0}\left\{\left(\frac{x^{1-b}}{1-b}\right)_\epsilon^1 + a\log(x)\int_\epsilon^1 \frac{1}{x^b}dx(t) - a\int_\epsilon^1 \frac{1}{x}\frac{x^{1-b}}{1-b}dx(t)\right\}$$

$$AUC = \lim_{\epsilon \to 0} \left(\frac{1-\epsilon^{(1-b)}}{1-b} - \frac{\epsilon^{(1-b)} a \log(\epsilon)}{1-b} - \frac{a}{(1-b)^2} \left(1-\epsilon^{1-b}\right) \right)$$

$$AUC = \frac{1}{1-b} - \frac{a}{(1-b)^2}$$

$$AUC = \frac{1-b-a}{(1-b)^2} \tag{10}$$

The above expression (10) is the final Area under the XL ROC curve, which provides the accuracy of the test in classification. The next section deals with the optimal threshold and the Youden's Index, which are crucial in identifying the optimal threshold to classify the subjects with greater accuracy.

2.1. Optimal threshold and Youden's index

Once, the ROC Curve and AUC are estimated, another important aspect in classification is to obtain the optimal threshold, which is unique and provides a better percentage of correct classification. This criterion can be met in two ways, first one depends on the condition that the pdf's of two populations are equal and second way is to make use of the widely accepted measure, namely Youden's Index (J). To obtain the optimal threshold for the ROC Curve, let us assume that both population densities are equal,

$$f_x(x, \theta_0) = f_x(x, \theta_1)$$

On further simplification, we have

$$\frac{\theta_0^2(2+\theta_0+t)e^{-t\theta_0}}{(1+\theta_0)^2} = \frac{\theta_1^2(2+\theta_1+t)e^{-t\theta_1}}{(1+\theta_1)^2}$$

The final expression for the optimal threshold is given by,

$$opt\ t = \frac{2}{(\theta_1-\theta_0)} \log\left(\frac{\theta_1}{\theta_0}\right) + \frac{2}{(\theta_1-\theta_0)} \log\left(\frac{1+\theta_0}{1+\theta_1}\right) + 1 \tag{11}$$

This is the optimal threshold for obtaining the best threshold or cutoff which will classify the populations with higher accuracy and lesser misclassification rate. Here, description on the use of Youden's Index in validating the obtained optimal threshold is given. In ROC context, the Youden's Index is defined as

$$J = \max\{Sensitivity(t) + Specificity(t) - 1\} \tag{12}$$

over all cut points t, $-\infty < t < \infty$. The index (J) ranges between 0 and 1 with a value of 1 indicating perfect diagnostic effectiveness and 0 indicating an ineffective test. With respect to the ROC curve, J is the maximum vertical distance between the curve and the diagonal (chance line) and acts as a global measure of the optimum diagnostic ability. Further, the Youden's index is obtained as

$$J = \max\left\{\left[1+\frac{t\theta_1}{(1+\theta_1)^2}\right]e^{-t\theta_1} - \left[1+\frac{t\theta_0}{(1+\theta_0)^2}\right]e^{-t\theta_0}\right\} \tag{13}$$

Further, ROC curve should possess the properties of monotonic transformation, invariance with respect to some increasing transformation. That is, the ROC curve is invariant to any monotonic (*e.g.*, linear, logarithmic) transformation of the test results (Krzanowski and Hand, 2009). Here, a brief outline about these basic properties of an ROC curve is discussed.

Properties of the ROC

i. $Y = h(x)$ is the mathematical model of the ROC curve, where y denotes the true positive rate and x denotes the false positive rate. The curve is a monotonic increasing function in the positive quadrant, lying between $y = 0 \ \ at \ \ x = 0$ and $y = 1 \ \ at \ \ x = 1$. Let us consider two false positives values P_1 and P_2 such that $P_1 < P_2$

$$\begin{aligned}
-b\log P_1 &< -b\log P_2 \\
e^{-b\log P_1} &< e^{-b\log P_2} \\
a\log(P_1)e^{-b\log P_1} &\leq a\log(P_2)e^{-b\log P_2} \\
e^{-b*\log P_1} + [a*\log P_1]e^{-b\log P_1} &\leq e^{-b*\log P_2}[a*\log P_2]e^{-b\log P_2} \\
e^{-b*\log P_1} + [1+a*\log P_1] &\leq e^{-b*\log P_2}[1+a*\log P_2] \\
XLROC(P_1) &\leq XLROC(P_2)
\end{aligned}$$

Hence the XL ROC curve is monotonically increasing.

ii. The ROC curve is unaltered if the classification scores undergo a strictly increasing transformation.

Let S denote the set of scores with $S \subset R$ and $h(\cdot)$ is strictly increasing function. Let $a, \ b \in S$ and $a < b$, then by using the strictly increasing function, we can write $h(a) < h(b)$.

The transformed random variables U and V from the respective healthy and diseased classes are

$$P(U \leq t) = P[h(U) \leq h(t)]$$
$$P(V \leq t) = P[h(V) \leq h(t)]$$

Let us consider the points $(x^*(t), y^*(t))$ on the ROC Curve for the transformed scores

$$x^*(t) = P\{h(U) > h(t)|H\} = 1 - P[h(U) \leq h(t)] = 1 - P(U \leq t) = x(t)$$

$$y^*(t) = P\{h(V) > h(t)|D\} = 1 - P[h(V) \leq h(t)] = 1 - P(V \leq t) = y(t)$$

Thus, the XLROC Curve is invariant to transformation.

iii. The slope of the XL ROC curve at threshold value t is given by

$$\frac{dy}{dx} = \frac{P(S > t|1)}{P(S > t|0)}$$

The derivative of the roc curve at a given pair of coordinates equals the likelihood ratio. Let us parameterize x and y in terms of t and derivative can be written as

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}} = \frac{g(t)}{f(t)}$$

Therefore,

$$\frac{dy}{dx} = \left\{\frac{\theta_1^2(2+\theta_1+t)e^{-t\theta_1}}{(1+\theta_1)^2}\right\}\left\{\frac{(1+\theta_0)^2}{\theta_0^2(2+\theta_0+t)e^{-t\theta_0}}\right\}$$

$$\frac{dy}{dx} = \left(\frac{\theta_1}{1+\theta_1}\right)^2 \left(\frac{1+\theta_0}{\theta_0}\right)^2 \left(\frac{2+\theta_1+t}{2+\theta_0+t}\right) e^{(\theta_0-\theta_1)t} \tag{14}$$

Here $\theta_0,\ \theta_1$ and $t \geq 0$. Therefore, $\frac{dy}{dx} \geq 0$. The behavior of the curve and the typical forms of ROC curve are totally dependent on the slope given in (14). This is the ratio of distribution of diseased scores to healthy scores of the two probability densities at the value of t. This is referred to as likelihood ratio of XL ROC curve, which is provided in the next section. In the next section, the well known Likelihood Ratio test is described and integrated in the context of ROC curve analysis. It is also a measure, which helps in understanding the shape of the ROC curve.

2.2. Likelihood ratio test

In order to establish an interesting relationship between the likelihood ratio test and the slope of the ROC curve, we have considered this concept of likelihood ratio test in classification as follows. In general, the likelihood ratio test has the hypothesis as

$$H_0 : \theta = \theta_0$$

Versus

$$H_1 : \theta = \theta_1$$

To test the hypothesis Likelihood Ratio Test is given by

$$\lambda = \frac{supL(x_1, x_2, \ldots, x_n; \theta_0)}{supL(x_1, x_2, \ldots, x_n; \theta_1)}$$

The likelihood function of XL distribution is

$$L = \left(\frac{\theta}{1+\theta}\right)^{2n} e^{-\theta \sum_{i=1}^{n} t_i} \prod_{i=1}^{n} (2+\theta+t_i)$$

Therefore, the likelihood ratio test for testing the hypothesis is

$$\lambda = \frac{\left(\frac{\theta_0}{1+\theta_0}\right)^{2n} e^{-\theta_0 \sum_{i=1}^{n} t_i} \prod_{i=1}^{n}(2+\theta_0+t_i)}{\left(\frac{\theta_1}{1+\theta_1}\right)^{2n} e^{-\theta_1 \sum_{i=1}^{n} t_i} \prod_{i=1}^{n}(2+\theta_1+t_i)}$$

$$\lambda = \left(\frac{\theta_0}{1+\theta_0}\right)^{2n} \left(\frac{1+\theta_1}{\theta_1}\right)^{2n} e^{(-\theta_0+\theta_1)\sum_{i=1}^{n} t_i} \prod_{i=1}^{n} \frac{(2+\theta_0+t_i)}{(2+\theta_1+t_i)}$$

This is the likelihood ratio test for testing the defined hypothesis, where $\theta_1 > \theta_0$. But, in case of XL ROC curve, we have $\theta_0 > \theta_1$. Therefore, λ can be rewritten as

$$\lambda = \left(\frac{\theta_1}{1+\theta_1}\right)^{2n} \left(\frac{1+\theta_0}{\theta_0}\right)^{2n} e^{-(\theta_1+\theta_0)\sum_{i=1}^{n} t_i} \prod_{i=1}^{n} \frac{(2+\theta_1+t_i)}{(2+\theta_0+t_i)} \tag{15}$$

For $n = 1$, the slope of the XL ROC curve (14) is equals the likelihood ratio (15) in the context of classification. Therefore, the functional relationship between slope of the ROC curve and the likelihood ratio is established to study the behavior of the ROC curve and its parameters.

The next section deals with the results and discussion to explain the proposed ROC curve in classification.

3. Results and discussion

This section presents the results obtained from the following experiments. Several tests were performed to demonstrate the characteristics and behavior of the XL ROC curve. Three different simulations were carried out using varying distribution parameters to illustrate better, moderate and worst cases classification scenarios in ROC analysis.

Table 1: Results of XLROC curve for simulated data at particular values of parameters

case	θ_0	θ_1	b	a	AUC	J	Opt t
Better	3.5	0.2	-0.06011	-0.04174	0.980441	0.32582	1.933603
Better	3	0.5	-0.17778	-0.07901	0.906016	0.673449	1.648744
Better	2.5	0.8	-0.34844	-0.10754	0.800741	0.125075	1.558186
Better	2	1.1	-0.61875	-0.14031	0.671305	0.253085	1.535916
Moderate	1.5	0.5	-0.39683	-0.17637	0.806302	0.482189	2.175573
Moderate	2.25	0.9	-0.44183	-0.12239	0.752437	0.162233	1.562207
Moderate	3.5	1.5	-0.45083	-0.07213	0.723527	0.263832	1.259511
Moderate	1.001	1	-1.33156	-0.33289	0.490134	0.000232	1.999251
Worst	0.2	3.5	-57.2727	-2.82828	0.017994	-0.41979	1.933603
Worst	0.5	3	-10.8	-0.675	0.089594	-0.05391	1.648744
Worst	0.8	2.5	-4.52009	-0.36899	0.193266	-0.1352	1.558186
Worst	1.1	2	-2.35137	-0.26126	0.321646	-0.05129	1.535916

In Table 1 (better case), as the scale parameter of healthy (normal) population decreases with opposite phenomenon in the diseased (abnormal) population, then it is found that area under the curve is decreasing and value of optimum threshold is also decreasing with the case better curve (Figure 1 - Better case). This means that as the parameters of the distributions moves away from the center to produce the largest discrimination between the two populations, the area under the curve increases with better classification rate.

Further, the moderate case of ROC curve is considered with different parameter values of the considered distribution and the results with the corresponding XL ROC curves are plotted (Figure 1 - Moderate case). Here, the area under the curve and its corresponding Youden's index are found to be good with moderate classification rate. Also, the when the AUC is about 50% (parameter values are equal), then the ROC curve is at the chance line.

Finally, the worst case of ROC curve is considered, where the normal and abnormal population parameter values are inverted and the corresponding AUC and the ROC curve are plotted (Figure 1 - Worst case). This is the case, where an abnormal population value tends to be higher than the normal population parameter values. This particular mathematical model of ROC curve is considered with the condition that the parameter values of normal population are to be higher side that the abnormal population ($\theta_0 \geq \theta_1$). So far, many

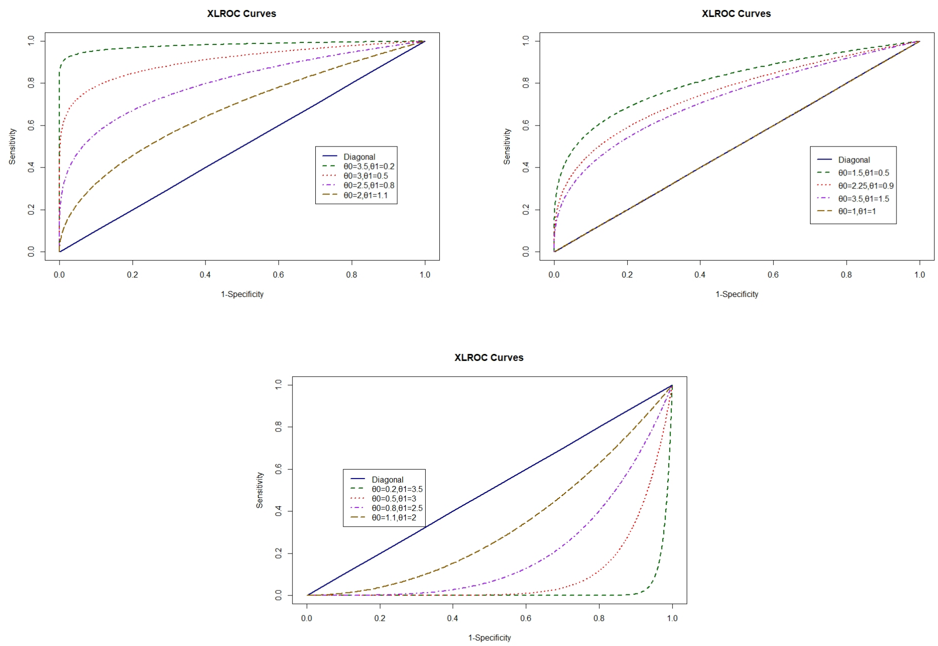

Figure 1: XL ROC curve using different simulations

models in ROC literature considered the other way, where the normal population values are at lower side than the abnormal population for example bi-normal, bi-exponential ROC curve.

3.1. Real dataset - APACHE IV

A real dataset APACHE IV (Balaswamy and Vardhan, 2023) is used to check the performance of the proposed ROC curve and the other existing ROC models like Lindley ROC (LROC) and Exponential ROC (EROC) curves (Balaswamy and Vardhan, 2021) and their corresponding AUC along with the optimal threshold values are given in Table 2.

Table 2: Comparison of different ROC curves using APACHE IV dataset

Method	θ_0	θ_1	AUC	J	Opt t (FPR, TPR)	KS test (H)	KS test (D)
XL ROC	0.2178	0.0912	0.7200	0.4400	12 (0.2022, 0.6422)	D = 0.1641 p-value = 0.0670	D = 0.1653 p-value = 0.1705
LROC	0.2464	0.1152	0.6945	0.3891	10 (0.2531, 0.6422)	D = 0.1682 p-value = 0.0477	D = 0.1413 p-value = 0.3301
EROC	0.1367	0.0607	0.6449	0.2899	10 (0.2547, 0.5446)	D = 0.0786 p-value = 0.8088	D = 0.0794 p-value = 0.9390

From Table 2, it is observed that the optimal threshold value is calculated and is reported as 12 (FPR = 0.2022 & TPR = 0.6422) and its corresponding Area under the curve for the APACHE IV dataset is 0.7200 with the youden's index value as 0.4400. This means that if the APACHE IV score of an individual who admitted to ICU is more than the optimal threshold 12, then the condition of the individual is considered as severe and belongs to the abnormal (diseased) group. The Kolmogorov-Smirnov test values are also given in the table 2 to check the validation of the data with the considered X Lindley distribution and the result shows that the data follows XL distribution. The corresponding XL ROC curve for the considered dataset is also plotted and shown in Figure 2. Further, the proposed ROC

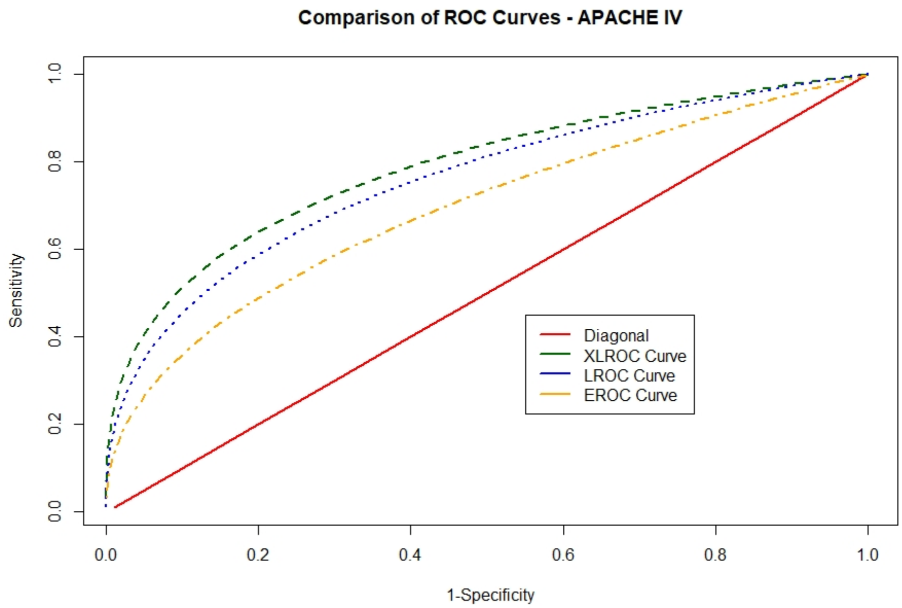

Figure 2: Comparison of ROC curves for APACHE IV dataset

curve methodology, XL ROC (AUC = 0.7200) is compared with the existing ROC models LROC (AUC = 0.6945) and EROC (AUC = 0.6449), which means that the XL ROC curve is performing better than the other models when the data follows the distribution assumed. The comparative plot is also drawn in the Figure 2, which explains that the proposed method is better than the other two models considered here.

Further, on performing the non-parametric ROC curve analysis, it is observed that the XL ROC curve methodology is found to be better with higher AUC values than the AUC of nonparametric ROC curve analysis both in simulations and real dataset. It is observed that the simulation results are also giving a higher accuracy than the non-parametric method ($At\ \theta_0 = 3.5$ & $\theta_1 = 0.5;\ AUC\ of\ XL\ ROC = 0.9804$ & $AUC\ of\ Non$-Parametric method $= 0.7435;\ At\ \theta_0 = 3$ & $\theta_1 = 0.5;\ AUC\ of\ XL\ ROC = 0.9060$ & $AUC\ of\ Non$-Parametric method $= 0.7193$).

On comparing the proposed methodology with non-parametric ROC curve, the XL ROC curve is found to be better than the non-parametric ROC method (AUC of XL ROC = 0.7200 with optimal threshold 12 and its corresponding FPR as 0.2022 & TPR as 0.6422 & AUC of Non-Parametric method = 0.6787 with optimal threshold 10 and its corresponding FPR as 0.2424, TPR as 0.6000). Therefore, the XL ROC methodology is recommended as compared to the Non-Parametric ROC method when the data follows an assumed distribution.

4. Conclusion

A Receiver Operating Characteristic (ROC) model based on the X Lindley (XL) distribution has been developed for classification tasks due to its greater flexibility compared to other single parameter distributions. The XL distribution is particularly effective for data analysis in classification contexts, offering computational simplicity that is accessible even to non-statisticians. The mathematical properties of the XL-based ROC curve have been rigorously verified, and a likelihood ratio test has been introduced for classification purposes. Extensive simulation studies demonstrate the model's performance under various scenarios, revealing better, moderate, and poorer ROC curve shapes as the scale parameters of the healthy and diseased populations vary. This model is especially relevant when parameters of the normal (healthy) group are higher than those of the abnormal (diseased) group in classification problems.

Acknowledgements

The authors are grateful to the Editor and reviewers for their guidance, valuable comments and suggestions to improve the paper.

Conflict of interest

The authors do not have any conflict of interest to declare for the research work included in this article.

References

Arunima, S. K. and Vardhan, V. R. (2023). Estimation of area under the multiclass ROC for non-normal data, *Statistics and Applications*, **21**, 113-121.

Arunima, S. K. and Vardhan, V. R. (2022). Estimation of area under the ROC curve in the framework of gamma mixtures. *Communications in Statistics Case Studies, Data Analysis and Applications*, **8**, 714-727.

Balaswamy, S. and Vardhan, V. R. (2023). AUC estimation and ROC model comparison in the perspective of generalized exponential distribution. *International Journal of Statistics and Reliability Engineering*, **10**, 346-351.

Balaswamy, S. and Vardhan, V. R. (2015). Confidence interval estimation of an ROC curve - an application of generalised half normal and weibull distribution. *Journal of Probability and Statistics*, **ID 934362**, 1-8.

Balaswamy, S. and Vardhan, V. R. (2021). Estimation of the area under the ROC curve in the framework of lindley centered distributions. *Statistics and Applications*, **19**, 231-240.

Balaswamy, S. and Vardhan, V. R. (2022). Estimation of the area under the ROC curve for non normal data. *Communication in Statistics - Case Studies, Data Analysis and Applications*, **8**, 393-406.

Balaswamy, S. and Vardhan, V. R. (2015). Interface between the ratio β with area under the ROC curve and kullback-leibler divergence under the combination of half normal and rayleigh distributions. *American Journal of Biostatistics*, **5**, 69-77.

Balaswamy, S., Vardhan, V. R., and Sarma, K. V. S. (2015). The Hybrid ROC (HROC) curve and its divergence measures for binary classification. *International Journal of Statistics in Medical Research*, **4**, 94-102.

Dashina, P. and Vardhan, V. R. (2022). Estimation of AUC of bi-generalised exponential ROC curve and its asymptotic results. *Advances and Applications in Statistics*, **79**, 105-119.

Dorfman, D. D. and Alf, Jr. E. (1968). Maximum likelihood estimation of parameters of signal detection theory - a direct solution. *Psychometrika*, **33**, 117-124.

Egan, J. P. (1975). *Signal Detection Theory and ROC Analysis.* Academic Press, New York.

Hussain, E. (2012). The bi-gamma ROC curve in a straightforward manner. *Journal of Basic and Applied Sciences*, **8**, 309-314.

Krzanowski, W. J. and Hand, D. J. (2009). *ROC Curves for Continuous data*, Monographs on Statistics and Applied Probability. CRC Press, New York.

Tang, L. L. and Balakrishnan, N. (2011). A random-sum wilcoxon statistic and its application to analysis of ROC and LROC data. *Journal of Statistical Planning and Inference*, **141**, 335-344.

Zou, K. H., Hall W. J., and Shapiro, D. E. (1997). Smooth nonparametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Statistics in Medicine*, **16**, 2143-2156.