Statistics and Applications {ISSN 2454-7395 (online)}
Volume 23, No. 1, 2025 (New Series), pp 303–320
https://www.ssca.org.in/journal

Determining Optimal Threshold and Some Inferential Procedures for a Skewed ROC Model in the Binary Classification Framework

Sandhya Singh[1], Saebugari Balaswamy[1] and R.Vishnu Vardhan[2]
[1]*Department of Statistics*
Indira Gandhi National Tribal University, Amarkantak, Madhya Pradesh, India
[2]*Department of Statistics, Pondicherry University, Puducherry*

Received: 31 May 2024; Revised: 20 June 2024; Accepted: 25 June 2024

Abstract

ROC curve is a useful tool in the assessment of the performance of a diagnostic test over the range of possible values of a predictor variable and the sensitivity, specificity, optimal threshold and Area under the curve (AUC) are its intrinsic measures to know the accuracy of the diagnostic test. The area under the curve is a measure of accuracy which provides the extent of correct classification of the test and also it is a measure of discrimination to compare the performance of two or more diagnostic tests. Further, the optimal threshold is a cut-off, which discriminates the populations into one of the two groups with a maximum of accurate accuracy. The Youden's Index method is the usual approach to identify the optimal threshold. The alternate approaches to compute the optimal threshold have been provided in this paper when the data is of skewed nature in the ROC context. For this purpose, ROC model is considered to show how the discriminatory ability of a test changes on changing the location and scale parameters by using a generalized half normal distribution. Further, the simulation studies are conducted to study the proposed methodology and also compared with the existing ROC models using both simulations and real datasets.

Key words: ROC Curve; Sensitivity; Specificity; AUC; Index of union; Concordance of probability.

AMS Subject Classifications: 62P10.

1. Introduction

The ROC curve was first developed by radar engineers during World War II for truly detecting enemy objects in battle fields starting in 1941 which led to its name "Receiver Operating Characteristic" (ROC) curve. Now-a-days, this technique is being extensively used in diverse areas of research such as banking, Finance, Engineering, Machine learning and Medical Sciences, *etc.* ROC curve was introduced in medicine for analysis of radiographic images (Lusted, 1971). This is an important tool applied in classification problems mostly

Corresponding Author: R. Vishnu Vardhan
Email: vrstatsguru@gmail.com

associated with evaluating the performance of the diagnostic test(s) by means of the accuracy or sensitivity measures, also to provide accuracy of the classifier/diagnostic test and helps in determining the optimal cut-off of a diagnostic test or classifier. To define ROC curve, there is a need of two intrinsic measures, such as, Sensitivity (True Positive Rate, TPR), which is the probability of a positive test result conditioned on the individual truly being positive and Specificity (True Negative Rate, TNR), which is the probability of a negative test result conditioned on the individual truly being negative. Graphically, the ROC curve can be achieved by using 1-TNR on x-axis and TPR on y-axis, resulting a smooth curve. This smooth curve is embedded with various threshold points; we tend to choose such threshold that attains minimum distance from the chance line. Though this approach is heuristic, there are other established indices that helps in determining the optimal threshold, one such index is the Youden's Index. The portion under the ROC curve is termed as the area under the curve (AUC), theoretically lies between 0 and 1. In a practical point of view, it is interpreted as that higher value of AUC indicates that the performance of marker/diagnostic test is better. Further, a test's AUC should not lie below or close to 0.5, this result in random classification and test is not considered for classification. Even though this technique's role is to classify the subjects into one of the predefined groups, it also allows allocating the new subjects into one of those groups with a proper status label. Further, much theoretical work has been done in the ROC context using different distributional assumptions and the formal statistical definition of ROC curve in terms of cumulative distribution function (CDF) is

$$ROC = 1 - G\left(F^{-1}(1-t)\right), 0 < t < 1$$

Here, F and G are the CDFs of two independent populations and the ROC model so generated is referred to as bi-distributional ROC model. The test score derived from a marker or diagnostic test do have some pattern and follows a particular distribution, then the ROC curve be developed based on that particular distribution, by which one can gets the proper fit of the data, and appropriate results with interpretation. In ROC literature, many models have been proposed based upon bi-distributional assumption such as Bi-lognormal (Dorfman and Alf, 1968, 1969), Bi-normal (Egan, 1975), Bi-gamma (Hussain, 2012), Bi-beta (Zou *et al.*, 1997), Bi-exponential (Tang and Balakrishnan, 2011), Hybrid ROC models (Balaswamy *et al.*, 2015) and many more. In the recent past, estimation of area under the ROC, for non-normal data (Balaswamy and Vardhan, 2022), Bi-Generalised Exponential ROC curve (Balaswamy and Vardhan, 2023), area under the ROC Curve in the framework of gamma mixtures (Arunima and Vishnu Vardhan, 2022), area under the multi-class ROC statistics and applications for non-normal data (Arunima and Vishnu Vardhan, 2023) are few to cite in the ROC framework.

This paper focuses on different procedure to obtain an optimal threshold other than the method of Youden's index. This provides the better and easiest way of calculating the optimal threshold. In order to demonstrate this methodology, a new ROC model is considered based upon a skewed distribution. To illustrate this skewed nature, let us consider a practical illustration. In assessing the subject's life status (alive or dead), a marker by name *Acute Physiology and Chronic Health Evaluation (APACHE II)* will be used. Mostly, the APACHE II score do not satisfy the normality assumption and possesses a skewed pattern. In such case, the conventional bi-normal ROC model may not suitable to assess the performance and threshold of APACHE II. So, there is a need to find a suitable statistical distribution that can meet the requirements of ROC model. Another marker that has similar kind of

non-normality is the *Simplified Acute Physiology Score (SAPS III)*. Hence, there is a need to look into the influence of measures of location, scale and shape to model a newer version of ROC. The present work addresses the above practical situations using the data of APACHE II and SAPS III. Along with these, simulations are also carried out to demonstrate the worst, moderate and better classification scenarios from the proposed ROC model. It is understood that these two datasets follow Generalized Half Normal distribution and for comparison purpose, two other distributions namely, the Normal and the Half-Normal are also considered.

2. Methodology

Let $(x_1, x_2) \in S$ be the test scores which are observed in healthy(0) and diseased (1) populations respectively. It is assumed that '0' and '1' population follow Generalized Half Normal Distribution with $\alpha > 0, \sigma > 0$ as shape and scale parameters, respectively. The probability density function and cumulative distribution function of Generalized Half Normal Distribution are given as follows:

$$g(x, \alpha, \sigma) = \sqrt{\frac{2}{\pi}} \left(\frac{\alpha}{x}\right) \left(\frac{x}{\sigma}\right)^{\alpha} exp\left(-\frac{1}{2}\left(\frac{x}{\sigma}\right)^{2\alpha}\right) \; ; x \geq 0$$

$$G(x, \alpha, \sigma) = 1 - 2\Phi\left[-\left(\frac{x}{\sigma}\right)^{\alpha}\right]$$

where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution. As the ROC curve is a trade-off between False Positive Rate (FPR) and True Positive Rate (TPR). Therefore, the FPR is derived by using probabilistic definition as follows

$$FPR = x(t) = P(S > t|0) = 2\left[1 - \Phi\left(\frac{t}{\sigma_0}\right)^{\alpha_0}\right] \tag{1}$$

on further simplification, the expression for t can be obtained as

$$t = \sigma_0 \left[\Phi^{-1}\left(1 - \frac{x(t)}{2}\right)\right]^{\frac{1}{\alpha_0}} \tag{2}$$

where $\Phi^{-1}(\cdot)$ is the inverse cumulative standard normal distribution function. Similarly, TPR expression is derived by using its probabilistic definition as follows

$$TPR = y(t) = P(S > t|1) = 2\left[1 - \Phi\left(\frac{t}{\sigma_1}\right)^{\alpha_1}\right] \tag{3}$$

substituting (2) in (3),

$$y(t) = 2\left[1 - \Phi\left(\left(\frac{\sigma_0}{\sigma_1}\right)^{\alpha_1} \left[\Phi^{-1}\left(1 - \frac{x(t)}{2}\right)\right]^{\frac{\alpha_1}{\alpha_0}}\right)\right]$$

Let, $\Phi^{-1}\left(1 - \frac{x(t)}{2}\right) = Z_x$ and on further simplification,

$$y(t) = 2\left[1 - \Phi\left(\left(\frac{\sigma_0}{\sigma_1}\right)^{\alpha_1} [Z_x]^{\frac{\alpha_1}{\alpha_0}}\right)\right]$$

Let, $\beta = \frac{\sigma_0}{\sigma_1}$ and $\alpha = \frac{\alpha_1}{\alpha_0}$. Then

$$y(t) = 2\left[1 - \Phi\left(\beta^{\alpha_1}\left[Z_x\right]^{\frac{\alpha_1}{\alpha_0}}\right)\right] \tag{4}$$

on further simplification, the expression for ROC curve is

$$y(t) = 1 - erf\left(\frac{\beta^{\alpha_1}\left[Z_x\right]^{\alpha}}{\sqrt{2}}\right) \tag{5}$$

The expression in (5) can be referred to as Generalized Half Normal ROC curve. In ROC methodology AUC measures the entire two dimensional area underneath the ROC curve.

$$AUC = \int_0^1 y(t)\mathrm{d}t$$

$$AUC = \int_0^1 1 - erf\left(\frac{\beta^{\alpha_1}\left[\Phi^{-1}\left(1 - \frac{x(t)}{2}\right)\right]^{\alpha}}{\sqrt{2}}\right)\mathrm{d}x(t) \tag{6}$$

The above expression has no closed form solution; therefore it needs to be evaluated numerically. The numerical evaluations have been carried out using Simpson's method in the results section. Let $\alpha = 1, i.e. \alpha_1 = \alpha_0 = 1$ in equation (5) and on further simplification,

$$AUC = 2 - 2\left[\Phi\left(\left(\frac{\sigma_0}{\sigma_1}\right)\left[\Phi^{-1}\left(1 - \frac{x(t)}{2}\right)\right]\right)\right] \tag{7}$$

The equation (7) is known as ROC curve for Half Normal Distribution (HN ROC curve) and the AUC for the HN ROC is given by

$$AUC = 1 - \frac{2}{\pi}\left(\frac{\sigma_0}{\sigma_1}\right) \tag{8}$$

3. Optimal threshold

The optimal threshold is very important in classification to obtain the good accuracy and to minimize the misclassification rate. Therefore, the four different methods to determine the optimal threshold that are in this paper are as follows.

Youden's index (J): This Index is a single statistic that captures the performance of a dichotomous diagnostic test. J is a function of sensitivity and specificity, such that

$$J(c) = \{Sensitivity(c) + Specificity(c) - 1\}$$

Over all cut point c; "optimal t" denotes the cut-point corresponding to J. When the value of J is maximum, optimal t is the optimum cut point value.

The closest to (0,1) criterion (ER): In this criteria, the optimal cut point is defined as the point closest to the point (0, 1) on the ROC curve.

$$ER(c) = \sqrt{(1 - TPR(c))^2 + (FPR(c))^2}$$

Mathematically, the point C_{ER} minimising the ER(c) function is called the optimal cut point value.

Concordance probability method (CZ): The concordance probability method defines the optimal cut point as the point maximizing the product of sensitivity and specificity.

$$CZ(c) = TPR(c) \times TNR(c)$$

The product gets value between 0 and 1. The concordance probability of dichotomized measure at cut point c can be expressed as the area of a rectangle associated with the ROC curve. Cut point $\hat{c_z}$ maximizing CZ(c) actually maximizes the area of the rectangle.

Index of union (IU): The optimal cut point should be chosen as the point which classifies most of individuals correctly and thus least of them incorrectly. From this point of view, Ilker (2017) proposed the index of union (IU) method to obtain the optimal threshold. This method provides an "optimal" cut point which has maximum sensitivity and specificity values at the same time. In order to find the highest sensitivity and specificity values at the same time, the AUC value is taken as the starting value of them. The above criteria correspond to the following equation,

$$IU(c) = (|TPR(c) - AUC| + |TNR(c) - AUC|)$$

The cut-point optimal t which minimizes the IU(C) function and the $|TPR(c) - TNR(c)|$ difference will be "optimal" cut point value.

Among these four methods of optimal threshold identification, choosing a one optimal threshold with good accuracy is a question. In order to answer this, Ilker (2017) compared these four methods with the mathematical optimal threshold (equating both density curves of healthy/ normal and diseased/abnormal populations and solve for the threshold). But this is not possible in all the cases of distributions, just like the case of proposed GHN ROC curve, here the closed form solution for the threshold is not possible. Therefore, keeping this in mind, we have used TPR value and their corresponding specificity values are considered to be higher. Wherever, these values are higher, that particular threshold will be of good choice with greater accuracy. Further, these four methods are tested at various sample sizes and different classification scenarios. In the next subsection, the inferential aspects of proposed ROC curve are discussed. For which, the variance of AUC is estimated through bootstrapping method as follows.

3.1. Bootstrap estimate of AUC

Since there is no closed form for AUC, its variance can be obtained using bootstrap technique. Let 'B' be the number of bootstraps obtained from the data with the sample sizes n_0 and n_1 respectively from normal and abnormal populations. Then the bootstrapped AUC estimate and its variance are given as

$$\widehat{AUC_B} = \frac{1}{B}\sum_{b=1}^{B} AUC_b \tag{9}$$

$$Var\left(\widehat{AUC_B}\right) = \frac{1}{B-1}\sum_{b=1}^{B}\left(AUC_b - \widehat{AUC}\right)^2 \tag{10}$$

3.2. Confidence intervals for AUC

Let $\widehat{AUC}$ denote the sample AUC value. For large samples, the distribution of AUC is approximately normal. Hence, a $100(1-\alpha)\%$ confidence interval for AUC may be computed using the standard normal distribution as follows

$$\widehat{AUC_B} \pm Z_{\frac{\alpha}{2}} \sqrt{Var\left(\widehat{AUC_B}\right)} \tag{11}$$

where $Z_{\frac{\alpha}{2}}$ is the $\frac{\alpha}{2}$ standard normal percentile.

3.3. Test statistic

A test with $AUC_0 = 0.5$ is considered useless as it classifies only 50% of individuals correctly. For this test, the ROC curve coincides with the chance line and $TPR = FPR$. Hence, the null and alternative hypothesis are defined as $H_0 : AUC = AUC_0 and H_1 : AUC > AUC_0$. Then the test statistic is defined as

$$Z = \frac{\widehat{AUC_B} - AUC_0}{\sqrt{Var\left(\widehat{AUC_B}\right)}} \tag{12}$$

The next subsection deals with the construction of confidence intervals for the proposed ROC Curve to explain the variability of the curve at each and every threshold value.

3.4. Confidence intervals for FPR and TPR

The $100(1-\alpha)\%$ confidence intervals for FPR and TPR, which in turn help in producing the confidence interval for GHN ROC curve. Therefore, the $100(1-\alpha)\%$ confidence intervals for FPR and TPR are as follows,

$$\widehat{FPR} \pm Z_{\frac{\alpha}{2}} \sqrt{Var(\widehat{FPR})}; \qquad \widehat{TPR} \pm Z_{\frac{\alpha}{2}} \sqrt{Var(\widehat{TPR})}$$

where variance of false positive rate and true positive rate are estimated through Delta method. The expression for $Var(\widehat{FPR}$ and $Var(\widehat{TPR}$ are

$$Var(\widehat{FPR}) = \left(\frac{\partial FPR}{\partial \sigma_0}\right)^2 Var(\hat{\sigma}_0) + \left(\frac{\partial FPR}{\partial \alpha_0}\right)^2 Var(\hat{\alpha}_0) \tag{13a}$$

$$Var(\widehat{TPR}) = \left(\frac{\partial TPR}{\partial \sigma_1}\right)^2 Var(\hat{\sigma}_1) + \left(\frac{\partial TPR}{\partial \alpha_1}\right)^2 Var(\hat{\alpha}_1) \tag{13b}$$

Further, the partial differentiations of FPR and TPR with respect to their parameters are as follows,

$$\begin{aligned}
\frac{\partial FPR}{\partial \sigma_0} &= \frac{\partial}{\partial \sigma_0}\left\{2\left[1-\Phi\left[\left(\frac{t}{\sigma_0}\right)^{\alpha_0}\right]\right]\right\}\\
\frac{\partial FPR}{\partial \sigma_0} &= \frac{2\alpha_0 t^{\alpha_0}}{\sigma_0^{\alpha_0+1}}\phi\left(\frac{t}{\sigma_0}\right)^{\alpha_0}\\
\frac{\partial FPR}{\partial \alpha_0} &= \frac{\partial}{\partial \alpha_0}\left\{2\left[1-\Phi\left[\left(\frac{t}{\sigma_0}\right)^{\alpha_0}\right]\right]\right\}\\
\frac{\partial FPR}{\partial \alpha_0} &= -2\phi\left(\frac{t}{\sigma_0}\right)^{\alpha_0}\left(\frac{t}{\sigma_0}\right)^{\alpha_0}\log\left(\frac{t}{\sigma_0}\right)\\
\frac{\partial TPR}{\partial \sigma_1} &= \frac{\partial}{\partial \sigma_1}\left\{2\left[1-\Phi\left[\left(\frac{t}{\sigma_1}\right)^{\alpha_1}\right]\right]\right\}\\
\frac{\partial TPR}{\partial \sigma_1} &= \frac{2\alpha_1 t^{\alpha_1}}{\sigma_1^{\alpha_1+1}}\phi\left(\frac{t}{\sigma_1}\right)^{\alpha_1}\\
\frac{\partial TPR}{\partial \alpha_1} &= \frac{\partial}{\partial \alpha_1}\left\{2\left[1-\Phi\left[\left(\frac{t}{\sigma_1}\right)^{\alpha_1}\right]\right]\right\}\\
\frac{\partial TPR}{\partial \alpha_1} &= -2\phi\left(\frac{t}{\sigma_1}\right)^{\alpha_1}\left(\frac{t}{\sigma_1}\right)^{\alpha_1}\log\left(\frac{t}{\sigma_1}\right)
\end{aligned}$$

Now, by substituting the above expressions in equations (13a) and (13b), we obtain the variances of FPR and TPR as,

$$\begin{aligned}
Var\left(\widehat{FPR}\right) &= \left[\frac{2\alpha_0 t^{\alpha_0}}{\sigma_0^{\alpha_0+1}}\phi\left(\frac{t}{\sigma_0}\right)^{\alpha_0}\right]^2 Var(\hat{\sigma}_0)\\
&\quad + \left[-2\phi\left(\frac{t}{\sigma_0}\right)^{\alpha_0}\left(\frac{t}{\sigma_0}\right)^{\alpha_0}\log\left(\frac{t}{\sigma_0}\right)\right]^2 Var(\hat{\alpha}_0) \qquad (14a)\\
Var\left(\widehat{TPR}\right) &= \left[\frac{2\alpha_1 t^{\alpha_1}}{\sigma_1^{\alpha_1+1}}\phi\left(\frac{t}{\sigma_1}\right)^{\alpha_1}\right]^2 Var(\hat{\sigma}_1)\\
&\quad + \left[-2\phi\left(\frac{t}{\sigma_1}\right)^{\alpha_1}\left(\frac{t}{\sigma_1}\right)^{\alpha_1}\log\left(\frac{t}{\sigma_1}\right)\right]^2 Var(\hat{\alpha}_1) \qquad (14b)
\end{aligned}$$

The variances of the parameters can be estimated through their asymptotic distributions, but in the present context the maximum likelihood estimators of the Generalized Half Normal distribution do not have closed form expressions. Therefore, the maximum likelihood parameters of these distributions can be obtained by direct maximization of log-likelihood function using the Newton-Raphson method in R. The asymptotic variances of the parameters are estimated using the Bootstrap method. Hence, the bootstrapped estimates of σ_0&α_0

and their variance are

$$\hat{\sigma}_0 = \frac{1}{B}\sum_{b=1}^{B}\sigma_{0b}$$
$$Var\left(\hat{\sigma}_0\right) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\sigma_{0b} - \hat{\sigma}_0\right)^2$$
$$\hat{\alpha}_0 = \frac{1}{B}\sum_{b=1}^{B}\alpha_{0b}$$
$$Var\left(\hat{\alpha}_0\right) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\alpha_{0b} - \hat{\alpha}_0\right)^2$$

In a similar manner, we can obtain the bootstrap estimate of $\sigma_1 \& \alpha_1$ as follows,

$$\hat{\sigma}_1 = \frac{1}{B}\sum_{b=1}^{B}\sigma_{1b}$$
$$Var\left(\hat{\sigma}_1\right) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\sigma_{1b} - \hat{\sigma}_1\right)^2$$
$$\hat{\alpha}_1 = \frac{1}{B}\sum_{b=1}^{B}\alpha_{1b}$$
$$Var\left(\hat{\alpha}_1\right) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\alpha_{1b} - \hat{\alpha}_1\right)^2$$

Now, using the above variances for the parameters of Generalized Half Normal distribution along with equations (14a) and (14b), the confidence intervals for FPR and TPR are obtained. By using these confidence intervals, the confidence interval lines can be plotted along with the GHN ROC curve to show the variability of the proposed ROC Curve at each and every point on the ROC space.

In the next section, the results are carried out using simulation studies and real datasets to explain the proposed methodology and the confidence intervals are also evaluated for the summary measure AUC and the proposed ROC Curve.

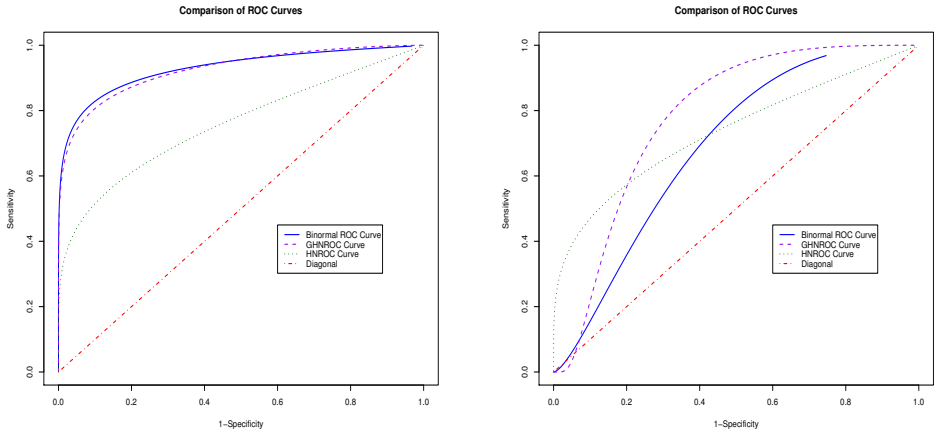
4. Results and discussions

Different simulation studies have been carried out to study the behaviour of the proposed ROC curve and also compared with the existing ROC models in literature. In this results and discussions sections, there are different subsections which will explain the necessity and importance of the proposed ROC curve in detail. The results reported in the tables are given in the appendix.

4.1. Comparison of ROC Curves - simulated datasets

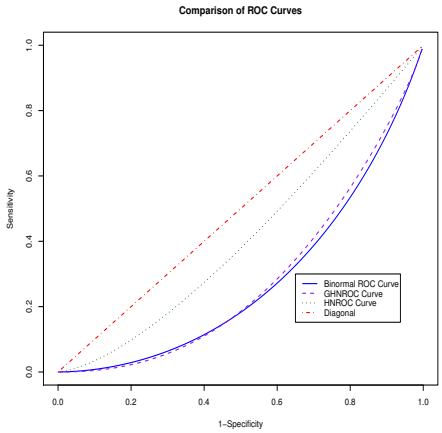
In this section different situations (Better, Moderate and Worst cases) of simulation studies in classification are considered and the results are given in Table 1, which consists of

optimal threshold, AUC, J and One sample KS test for testing the reliability of the simulated data (from GHN distribution) with GHN, Half Normal and Normal distributions. The GHN ROC model is compared with the existing ROC models like HN ROC and Binormal ROC model in all the three different situations of classification.



(a) Better case of Classification

(b) Moderate case of Classification



(c) Worst case of Classification

Figure 1: Comparison of ROC curves for different cases of classification using simulated datasets

Table 1 shows the differences and the importance of proposed GHN ROC model as compared to the existing ROC models with different simulation studies. The first case is better case of classification and the accuracy measure AUC is found to be 90% for the GHN ROC Curve when the data of both populations follows a generalized half normal distribution (KS test statistic values are given in the table). Whereas, when the shape parameter is suppressed, the proposed ROC model reduces to the half normal ROC model and this case has an accuracy of 74% and the data follows half normal distribution. The interesting fact observed is that even though the data of healthy ($D = 0.5631, p - value < 2.2e^{-16}$) and diseased ($D = 0.8461, p - value < 2.2e^{-16}$) populations do not follow the normal distribution, the accuracy is found to be 92%. This means that the Binormal ROC model is over estimating

the accuracy when the data does not satisfy the distributional properties. This is the reason that one must check for the distributional assumptions when you have the data in hand first (This type of situation can be seen in the next section with APACHE II score). Further, the corresponding ROC Curves are drawn in Figure 1a with better accuracy of classification where the curves are nearer to the top right corner of the ROC plot.

The moderate case of classification is considered (Table 1) and the GHN ROC curve (78% of accuracy) is clearly superior than the other two models half normal (72%) and Binormal ROC models (68%). Here also, the distributional properties are verified with the help of KS test statistic and found that when the data follows generalized half normal distribution, the accuracy is higher than the other two models when the data do not follow normality. Similar kind of phenomenon can be seen in Figure 1b, where the curves explain the moderate case of classification.

Finally, the worst case of classification is also considered where the parameters have the higher values in healthy population than the diseased population and the results are placed in Table 1 and Figure 1c. In this experiment also, it is found that the proposed ROC model is better than the Binormal ROC model when the data deviates from normality ($Healthy:\ D = 0.7160,\ p-value < 2.2e^{-16}$ & $Diseased:\ D = 0.6176,\ p-value < 2.2e^{-16}$).

Further, the optimal threshold, Youden's index,false positive rate and true positive rate at the corresponding optimal threshold are also computed and depicted in the Table 1. The optimal threshold is the value or score which divides the data into one of the two possible cases with a good amount of accuracy with lesser misclassification rate. These are computed for all the cases of classification along with the FPR and TPR at that particular optimal threshold.

4.2. Comparison of ROC Curves - real dataset

In this section, two real datasets are used to illustrate the proposed methodology and comparison is made with the existing ROC models and the results are as follows. The APACHE II (Acute Physiology and chronic Health Evaluation II) and SAPS III (Simplified Acute Physiology Score) datasets are considered to explain the proposed methodology and its significance over other ROC models like Binormal and HNROC models. The Tables 2 &

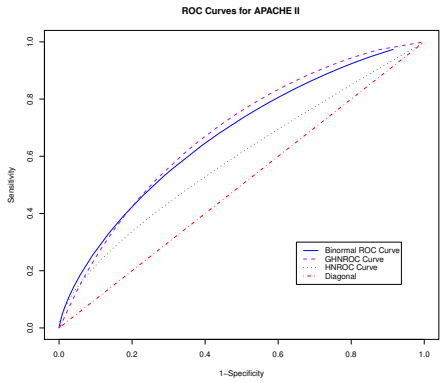


Figure 2: Comparison of ROC curves - APACHE II dataset

3 consists of optimal threshold, FPR, TPR, J and AUC along with the KS test statistics and

their significance values. The real data set is about the ICU scoring system namely APACHE II (Balaswamy and Vardhan, 2015) which is used to predict the status of the patient *i.e.* dead or alive. This is commonly used score which is derived from 11 physiological variables, the Glasgow coma (scores) and the patient's age and chronic health status. A total of 111 patients of which 66(59.46%) are alive and 45(40.54%) dead are present in this study. Further, the GHN ROC curve is plotted and the computations are done with respect to the proposed ROC model and compared with the existing ROC models like Binormal and HN ROC curves. When this data of both alive (D = 0.11785, p-value = 0.3185) and dead (D = 0.089239, p-value = 0.8661) populations follows Generalized Half Normal distribution, the accuracy of the test is 68.3% with the optimal threshold of 26, which classifies the data as abnormal as abnormal about 65% (TPR). Further, it is noticed that the accuracy is lesser in other models like Binormal (67.2%) and HN ROC curve (58.9%) than the GHN ROC model, which means the proposed GHN ROC model is performing better than the existing ROC models when the data follows generalized half normal distribution other than the normal and half normal distributions. Finally, the ROC curves are plotted to show the discrimination ability of the proposed ROC curve with the existing ROC models and is depicted in the Figure 2. The real data set is about the ICU scoring system SAPS III (Balaswamy and Vardhan,

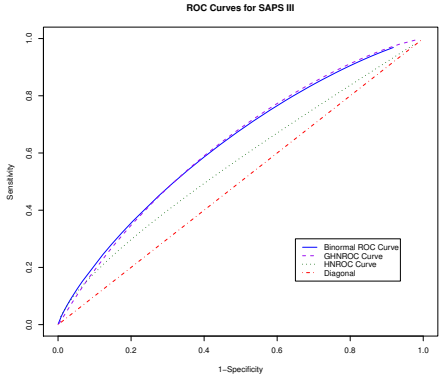


Figure 3: Comparison of ROC curves - SAPS III dataset

2022) III and issued to predict the life status of a subject who is admitted to ICU. The data consists of a total of 111 respondents of which 66(59.46%) are alive and 45(40.54%) dead. In above Table 3, comparison of three ROC curve have been done, when the data follows GHN distribution for both the populations ($normal\ population:\ D = 0.11431,\ p-value = 0.3544$ & $abnormal\ population:\ D = 0.13555,\ p-value = 0.38$) and it is also seen that the data do not follow the normal distribution ($normal\ population:\ D = 0.99997,\ p-value < 2.2e^{-16}$ & $abnormal\ population:\ D = 0.97778,\ p-value < 2.2e^{-16}$). Using the proposed methodology we have used this scoring variable to predict the mortality of patients in ICU. From the obtained result, it is observed that discriminatory ability of generalized half normal distribution (63.07%), and Binormal distribution (63.03%) is almost same *i.e.* 63% whereas when data follow half normal distribution discriminatory ability of the diagnostic test is less *i.e.* 56%. The interesting fact observed is that even though the data doesn't follow the normal distribution, the Binormal ROC curve is providing the similar accuracy with the proposed GHN ROC curve, this means that the Binormal ROC curve is over estimating the accuracy with the optimal threshold of 30, which provides only of 58.5% of true positive rates whereas the GHN ROC curve provides the optimal threshold of 26 with the higher

true positive rates of 63.6%, *i.e.*, the GHN ROC curve is more accurately classifying the data than the existing models when the data follows that particular generalized half normal distribution.

Figure 3, depicts the three ROC models for SAPS III dataset and the GHN ROC curve is slightly higher than the Binormal Roc curve and better than the HN ROC curve with the accuracy of 63%.

4.3. Optimal thresholds and confidence intervals for the ROC curve

In this section, the optimal thresholds are estimated by using different methods and are explained for the ease of medical practitioner. Further, the confidence intervals are also constructed for the proposed GHN ROC curve along with the Z test statistic for the area under the curve (AUC). Here, the effect of sample size on the proposed ROC curve is also be discussed. Three different classification situations (better, moderate and worst) are considered over different sample sizes. The entire simulations and the results are carried out using R programming and a bootstrap methodology is also used for the proposed ROC methodology. The results are as follows. Table4 (Better case) consists of optimal thresholds; FPR and TPR at that particular optimal threshold along with the confidence intervals of AUC and its Z statistic for testing the hypothesis. These optimal thresholds are evaluated using four different methods (J, ER, CZ and IU) and the results are also evaluated at various sample sizes. From these four methods of obtaining an optimal threshold at each and every sample size, the Youden's index method and IU methods are almost same with respect to the better classification scenario with AUC of more than 90%. The optimal threshold can be considered either from method J or IU, since their corresponding true positive rate (sensitivity) is higher than compared to the other methods, which means misclassification rate can be reduced using these methods with higher accuracy. Further, the Z statistic is found to be higher ($Z > 1.96$), *i.e.*, the curves obtained at this combination are significant enough to explain the accuracy of a test.

The confidence intervals are constructed for the considered combination of parameters (Better case) at various sample sizes and are depicted in Figure 4. The optimal threshold identified by method J and IU are also highlighted in the diagram with its corresponding FPR and TPR. Here, one can see the effect of sample sizes clearly, *i.e.*, as the sample size increase, the confidence intervals become closer to each other. Therefore, the accurate results may be obtained with the higher sample sizes.

The moderate case of classification scenario is also considered with $\sigma_0 = 1.5, \sigma_1 = 2.4, \alpha_0 = 0.9 and \alpha_1 = 2.5$ and the results with respect to the optimal threshold and AUC with its confidence intervals are also reported (Table 5). In this situation also, the optimal threshold can be identified by the methods of J and IU, since their sensitivity is higher than the other methods at each and every sample size. The AUC is observed to be more than 70% and the Z value is found be rejected ($Z > 1.96$), this means that the ROC curves are good enough to explain the extent of correct classification with the corresponding optimal thresholds. Further, it is to note that the optimal threshold can be obtained from method J or IU in both the cases of better and moderate case of classification scenarios. The confidence intervals are constructed for the considered moderate case at various sample sizes and are depicted in Figure 5. The optimal threshold identified by method J and IU are

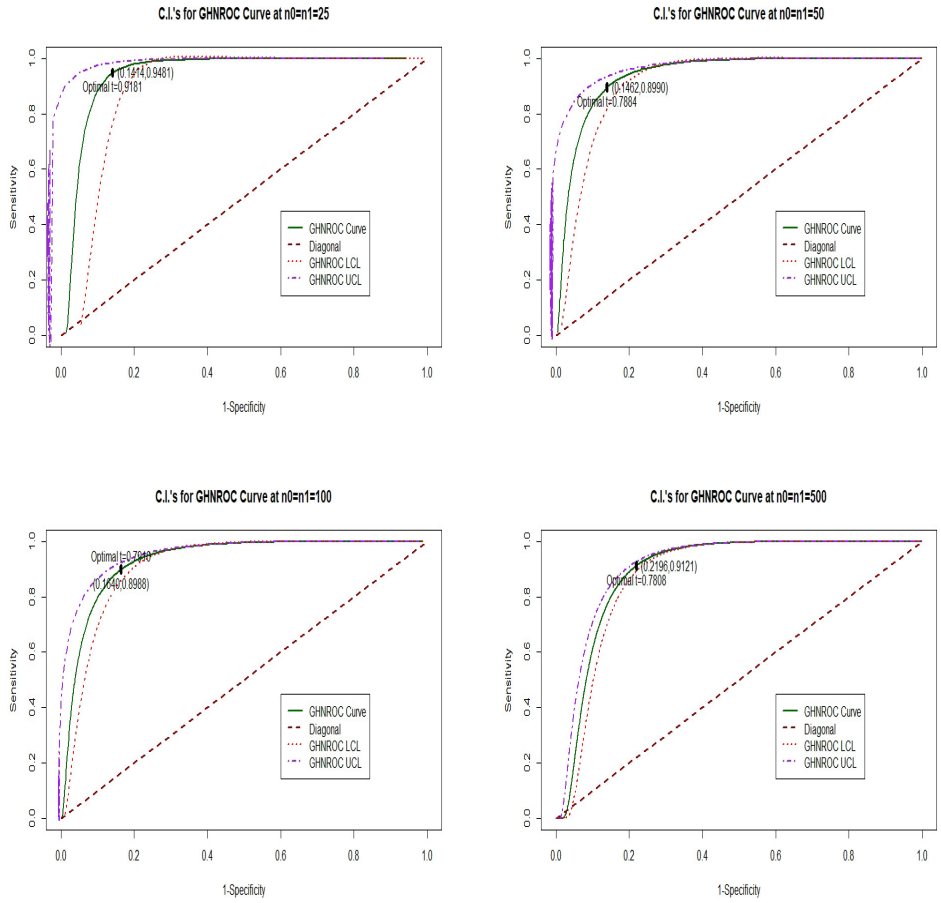


Figure 4: Confidence intervals for GHN ROC curve with its optimal threshold - better case

also highlighted in the diagram with its corresponding FPR and TPR. Here, as the sample size increase, the confidence intervals become closer to each other.Finally, the worst case classification scenario is considered (Table 6) to obtain the optimal threshold and thereby its accuracy. Though this scenario is of not at all useful in reality; the results are carried out to check the methods of obtaining optimal thresholds at various sample sizes. The very interesting factor observed here is that the ER method is found to be better with moderate amount of TPR and reasonably FPR as compared to the other methods. Even though, the sensitivity of method J is higher, but the corresponding FPR is also higher, where it should be minimum. Further, the accuracy is below 50% with the hypothesis is found to be insignificant stating that the curve obtained at this combination is not useful for future classification.

The confidence intervals are constructed for the considered worst case at various sample sizes and are depicted in Figure 6. The optimal threshold identified by method ER is highlighted in the diagram with its corresponding FPR and TPR.

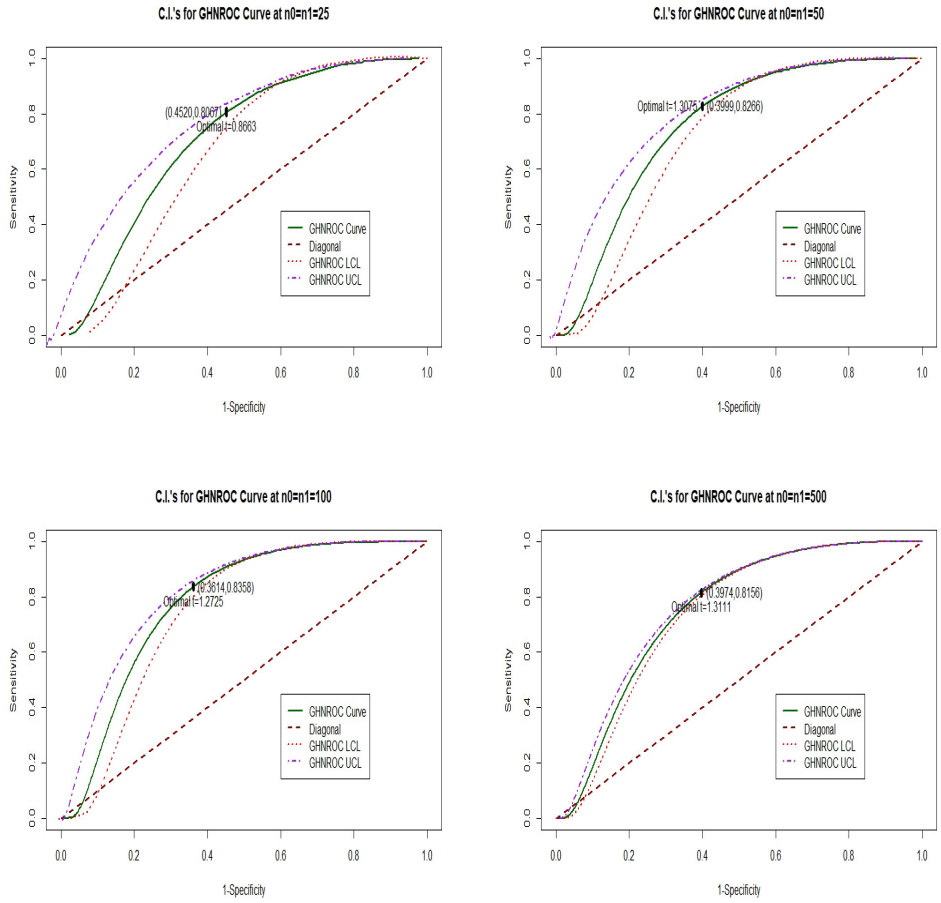


Figure 5: Confidence intervals for GHN ROC curve with its optimal threshold - moderate case

5. Conclusion

The Receiver Operating Characteristic (ROC) curves are useful in detecting the optimal threshold of medical diagnostic test with good extent of correct classification and accuracy. Therefore, on working with real datasets, the knowledge on distributional based ROC curves will be quite useful. Keeping this in mind, the ROC curve for generalized half normal distribution is proposed and the properties are verified. Further, extensive simulation studies are done with respect to the proposed ROC model and this model is also compared with the existing ROC models like Binormal and HN ROC models to show the proposed ROC model is better with skewed data of generalized half normal distribution. The real datasets (APACHE II and SAPS III) are used to demonstrate the behaviour of the proposed ROC curve in the results section. The accuracy measure for the proposed method using SAPS III dataset is higher (63%) than the AUC of SAPS III dataset (56%) proposed by Dashina and Vishnu Vardhan (2023) and that ROC model has incorrect mathematical expressions, which misleads the results. Therefore, this model is more useful than any other when the data is of generalized half normal distribution.

As, the optimal threshold identification is most important in classification, therefore

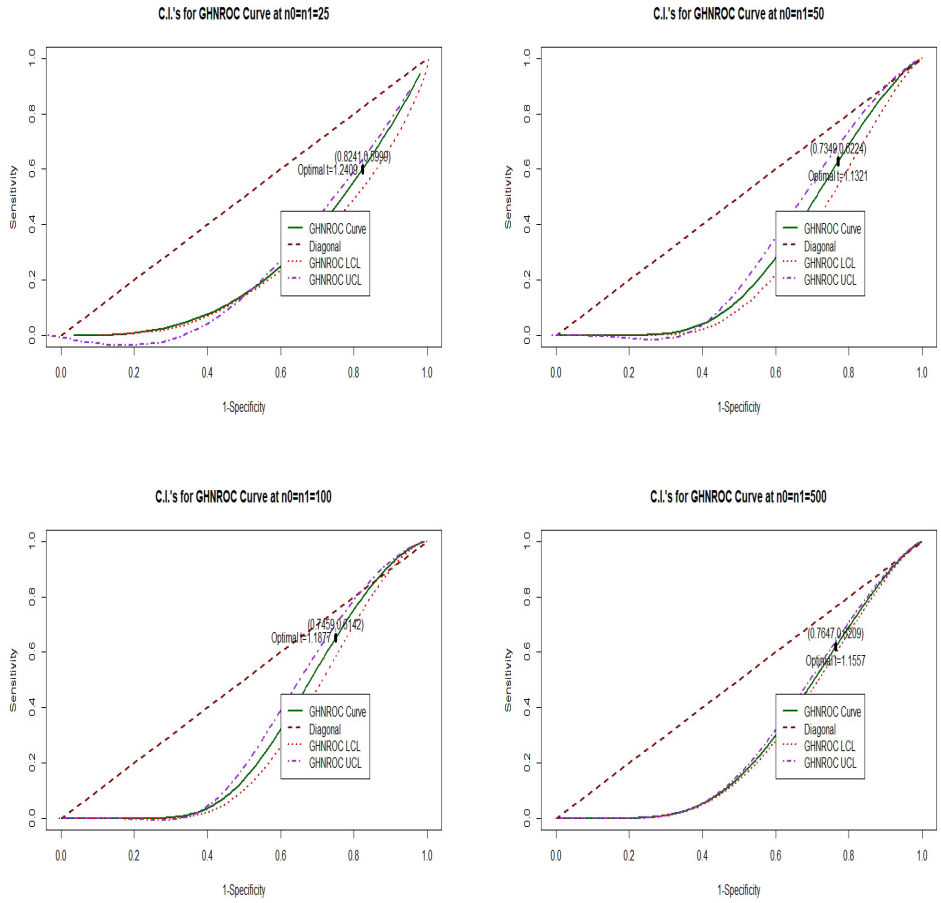


Figure 6: Confidence intervals for GHN ROC curve with its optimal threshold - worst case

four different methods are used to identify the optimal threshold with better accuracy. The interesting results observed is that the methods J and IU are found to be similar though their mathematical formulae are different at various sample sizes (Better and Moderate cases of classification scenario). But, the ER method is found to be good in case of worst classification situation, though this case is not at all considerable. Further, the proposed GHN ROC curve is found to be better with respect to the existing ROC curves when the data is of skewed in nature and follow the generalized half normal distribution.Further, the confidence intervals are also constructed for the ROC curve at various sample sizes and the AUC is also tested with the chance line 50%. Also, it is suggested that among the four methods of optimal threshold, one can consider J or IU methods with equal importance. In order to obtain the best optimal threshold, the usual method of equating densities is always not possible as in this case (no closed form solution). Therefore, we have suggested considering the sensitivity value and their corresponding specificity values to be higher. Wherever, these values are higher, that particular threshold will be of good choice with greater accuracy.

Acknowledgements

The first author would like to thank the Department of Community Medicine, Govt. Bundelkhand Medical College, Sagar, Madhya Pradesh for supporting her to carry out the research work. Also, 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 financial or non-financial conflict of interest to declare for the research work included in this article.

References

Balaswamy S. and Vishnu Vardhan 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 Vishnu Vardhan 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 Vishnu Vardhan 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., Vishnu Vardhan 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 Vishnu Vardhan R. (2023). Estimation of AUC of bi-Generalized half-Normal ROC curve. *Statistics and Applications*, **21**, 59-71.

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

Dorfman, D. D. and Alf, E. (1969). Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals - Ratingmethod data. *Journal of Mathematical Psychology*, **6**, 487-496.

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.

Ilker Unal. (2017). Defining an optimal cut-point value in ROC analysis: an alternative approach. *Computational and Mathematical Methods in Medicine*, Art.Id:3762651, doi:10.1155/2017/3762651.

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

Kannan Arunmima S. and Vishnu Vardhan 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.

Lusted L.B. (1971). Signal detectability & medical decision making. *Science*, **171**, 1217-1219.

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.

Appendix

Table 1: Comparison of ROC curves for different cases of classification using simulated datasets

Experiment	ROC Curve	σ_0	σ_1	α_0	α_1	μ_0	μ_1	Optimal Threshold	FPR	TPR	J	AUC	KS Test (Healthy)	KS Test (Diseased)
Better Case	GHN ROC Curve	1.4463	3.5406	1.4888	2.5065	-	-	2.0457	0.0938	0.8004	0.7066	0.9053	D = 0.0364, p=0.5183	D = 0.0302, p=0.7517
	HNROC Curve	1.4848	3.7447	-	-	-	-	2.2148	0.1339	0.5524	0.4186	0.7481	D = 0.0403, p=0.3887	D = 0.0465, p=0.2287
	Binormal ROC Curve	0.6242	1.0603	-	-	1.1661	2.9710	2.0183	0.0861	0.8155	0.7295	0.9288	$D = 0.5631, p < 2.2e^{-16}$	$D = 0.8461, p < 2.2e^{-16}$
Moderate Case	GHN ROC Curve	0.7587	1.8173	0.4911	1.5118	-	-	0.6257	0.3630	0.8419	0.4789	0.7816	D = 0.0356, p=0.5485	D = 0.0215, p=0.9745
	HNROC Curve	0.7931	1.8117	-	-	-	-	1.1346	0.1531	0.5300	0.3769	0.7283	D = 0.0245, p=0.9234	D = 0.0260, p=0.8854
	Binormal ROC Curve	1.1596	0.7860	-	-	0.7691	1.4589	0.7693	0.4999	0.8099	0.3099	0.6888	$D = 0.5, p < 2.2e^{-16}$	$D = 0.5836, p < 2.2e^{-16}$
Worst Case	GHN ROC Curve	1.9976	1.4890	2.4662	2.0943	-	-	3.2434	0.0010	0.0000	-0.0010	0.2864	D = 0.0225, p=0.9607	D = 0.0326, p=0.6619
	HNROC Curve	1.9358	1.4847	-	-	-	-	6.4408	0.0010	0.0000	-0.0009	0.4162	D = 0.0364,p=0.5187	D = 0.0369, p=0.501
	Binormal ROC Curve	0.5952	0.5047	-	-	1.6822	1.2250	3.2434	0.0044	0.0000	-0.0043	0.2790	$D = 0.7160, p < 2.2e^{-16}$	$D = 0.6176, p < 2.2e^{-16}$

Table 2: Results of ROC curves for the APACHE II dataset

ROC Curve	σ_0	σ_1	α_0	α_1	μ_0	μ_1	Optimal Threshold	FPR	TPR	J	AUC	KS Test (Healthy)	KS Test (Diseased)
GHN ROC Curve	29.411	42.9109	1.0555	1.5772	-	-	26	0.3799	0.6500	0.2700	0.6836	D = 0.11785, p = 0.3185	D = 0.089239, p = 0.8661
HNROC Curve	28.8184	38.4924	-	-	-	-	33	0.2521	0.3912	0.1391	0.5896	D = 0.13462, p = 0.1827	D = 0.20792, p = 0.0408
Binormal ROC Curve	17.0215	17.6888	-	-	23.3486	34.2889	28	0.3923	0.6389	0.2465	0.6720	D = 0.98485, $p < 2.2e^{-16}$	D = 1, $p < 2.2e^{-16}$

Table 3: Results of ROC curves for the SAPS III dataset

ROC Curve	σ_0	σ_1	α_0	α_1	μ_0	μ_1	Optimal Threshold	FPR	TPR	J	AUC	KS Test (Healthy)	KS Test (Diseased)
GHN ROC Curve	32.6943	41.9543	1.1832	1.5636	-	-	26.0000	0.4457	0.6360	0.1903	0.6307	D = 0.11431, p = 0.3544	D = 0.13555, p = 0.38
HNROC Curve	30.9450	38.0462	-	-	-	-	34.0000	0.2719	0.3715	0.0996	0.5646	D = 0.10829, p = 0.4213	D = 0.2031, p = 0.04883
Binormal ROC Curve	17.6210	17.6201	-	-	25.5303	33.8222	30.0000	0.3999	0.5859	0.1860	0.6303	D = 0.99997, $p < 2.2e^{-16}$	D = 0.97778, $p < 2.2e^{-16}$

Table 4: Intrinsic measures of GHN ROC curve using the methods of optimal threshold at different sample sizes - better case ($\sigma_0 = 0.5, \sigma_1 = 1.8, \alpha_0 = 0.5$ *and* $\alpha_1 = 2.5$)

Method	Sample Size	Status	Optimal Threshold	FPR	TPR	Value of method	AUC (LCL, UCL)	Z Statistic for AUC
J	25	1	0.9182	0.1414	0.9481	0.8067	0.9443	19.397
ER	25	1	0.9828	0.1286	0.9343	0.1444		
CZ	25	0	0.9493	0.1351	0.9418	0.8145	(0.8994, 0.9892)	
IU	25	1	0.9182	0.1414	0.9481	0.0919		
J	50	0	0.7885	0.1463	0.8990	0.7528	0.9339	24.2761
ER	50	1	0.8231	0.1365	0.8879	0.1766		
CZ	50	0	0.7885	0.1463	0.8990	0.7675	(0.8989, 0.9690)	
IU	50	0	0.7885	0.1463	0.8990	0.1182		
J	100	1	0.7911	0.1640	0.8988	0.7348	0.9249	30.8472
ER	100	0	0.8641	0.1428	0.8744	0.1902		
CZ	100	0	0.8180	0.1559	0.8902	0.7514	(0.8979, 0.9519)	
IU	100	1	0.7911	0.1640	0.8988	0.1238		
J	500	1	0.7808	0.2196	0.9122	0.6926	0.900	37.7962
ER	500	1	0.8920	0.1899	0.8763	0.2266		
CZ	500	0	0.8185	0.2090	0.9009	0.7126	(0.8793, 0.9208)	
IU	500	0	0.8185	0.2090	0.9009	0.1100		

Table 5: Intrinsic measures of GHN ROC curve using the methods of optimal threshold at different sample sizes - moderate case ($\sigma_0 = 1.5, \sigma_1 = 2.4, \alpha_0 = 0.9$ *and* $\alpha_1 = 2.5$)

Method	Sample Size	Status	Optimal Threshold	FPR	TPR	Value of method	AUC (LCL, UCL)	Z Statistic for AUC
J	25	1	0.8663	0.4521	0.8067	0.3547	0.7153	2.948
ER	25	1	1.1190	0.3592	0.7003	0.4678		
CZ	25	0	0.9501	0.4192	0.7733	0.4491	(0.5721, 0.8585)	
IU	25	1	1.1190	0.3592	0.7003	0.0818		
J	50	0	1.3075	0.4000	0.8266	0.4267	0.7539	4.4455
ER	50	1	1.5305	0.3263	0.7418	0.4161		
CZ	50	1	1.4413	0.3546	0.7779	0.5021	(0.6419, 0.8659)	
IU	50	1	1.5305	0.3263	0.7418	0.0951		
J	100	1	1.2725	0.3614	0.8359	0.4745	0.7766	7.2527
ER	100	1	1.4658	0.3064	0.7695	0.3834		
CZ	100	1	1.3689	0.3331	0.8043	0.5364	(0.7019, 0.8514)	
IU	100	1	1.4658	0.3064	0.7695	0.0956		
J	500	1	1.3112	0.3975	0.8157	0.4182	0.7497	14.3671
ER	500	0	1.5022	0.3380	0.7470	0.4222		
CZ	500	1	1.4373	0.3575	0.7716	0.4958	(0.7156, 0.7837)	
IU	500	1	1.4968	0.3396	0.7491	0.0900		

Table 6: Intrinsic measures of GHN ROC curve using the methods of optimal threshold at different sample sizes - worst case ($\sigma_0 = 2, \sigma_1 = 1.5, \alpha_0 = 2.3$ *and* $\alpha_1 = 3$)

Method	Sample Size	Status	Optimal Threshold	FPR	TPR	Value of method	AUC (LCL, UCL)	Z Statistic for AUC
J	25	0	2.5139	0.0348	0.0000	-0.0347	0.2693	-3.8058
ER	25	0	1.2409	0.8241	0.5999	0.9161		
CZ	25	1	1.4317	0.7258	0.4258	0.1168	(0.1505, 0.3881)	
IU	25	0	1.5985	0.6182	0.2719	0.1162		
J	50	1	0.4999	0.9500	0.9646	0.0146	0.2997	-3.7384
ER	50	1	1.1321	0.7349	0.6224	0.8263		
CZ	50	0	1.1972	0.7040	0.5615	0.1662	(0.2246, 0.4140)	
IU	50	0	1.4125	0.5931	0.3449	0.0856		
J	100	1	0.5197	0.9601	0.9688	0.0087	0.3126	-4.7832
ER	100	1	1.1877	0.7459	0.6142	0.8397		
CZ	100	1	1.2497	0.7161	0.5551	0.1576	(0.2358, 0.3894)	
IU	100	0	1.4805	0.5938	0.3190	0.0872		
J	500	1	0.3895	0.9786	0.9836	0.0050	0.3081	-11.498
ER	500	1	1.1557	0.7647	0.6210	0.8535		
CZ	500	1	1.2344	0.7291	0.5488	0.1487	(0.2753, 0.3408)	
IU	500	0	1.4827	0.6032	0.3056	0.0912		