

Estimation of AUC for Assessing Its Significance in Classification Models

Okeh Uchechukwu Marius¹ Mbegbu Julian Ibezimako²

1. Department of Mathematics and Computer Science, Ebonyi State University Abakaliki, Nigeria.

2. Department of Statistics, Nnamdi Azikiwe University Awka, Nigeria

ABSTRACT

The assessment of the performance of a diagnostic test when test results are measured on continuous scale can be evaluated using the measures of sensitivity and specificity over the range of possible cut-off points for the predictor variable. This is achieved by the use of a receiver operating characteristic (ROC) curve which is a graph of sensitivity against 1-specificity across all possible decision cut-offs values from a diagnostic test result. This curve evaluates the diagnostic ability of tests to discriminate the true state of subjects especially in classification models. These tasks of assessing the predictive accuracy of classification models is always better achieved using a summary measure of accuracy across all possible ranges of cut-off values called the area under the receiver operating characteristic curve (AUC). In this paper, we propose a simple nonparametric method of calculating AUC from predicted probability of positive response involving multiple prediction rules. This method is based on the knowledge of non-parametric Mann-Whitney U statistic. Based on the predicted outcomes and observed outcomes, the performance of diagnostic tests is assessed for the classification models through the AUC calculated from these outcomes. The proposed method when applied on real data, the significance of AUC for the classification models is assessed. The method offers reliable statistical inferences and circumvents the difficulties of deriving the statistical moments of complex summary statistics seen in the parametric method. The proposed method as a non-parametric estimation is recommended for calculating the AUC as it compares favorably with the existing parametric and non-parametric methods.

Keywords: Cut-off value, ROC, Predicted probability, parametric, non-parametric

DOI: 10.7176/JNSR/9-9-03

Publication date: May 31st 2019

INTRODUCTION

In medical sciences, the use of diagnostic procedures is based on clinical investigations or laboratory experiments or trials purposely to classify subject into diseased or non-diseased. These procedures makes for vital decision making aided with advanced machines/tools to detect any given condition. For decades now, receiver operating characteristic curve (ROC) analysis has been used as a popular technique of evaluating the performance or ability of a test to discriminate between alternative health status. The ROC curve represents a graph of sensitivity against 1-specificity across various cut-off values of diagnostic test. It assesses the effectiveness of continuous diagnostic test results to differentiate between groups of healthy and diseased individuals (Greiner et al., 2000; Zhou et al., 2002; Pepe, 2004). It is also a common tool for assessing the performance of various classification tools such as biological test results, diagnostic tests, and statistical distributional models and in assessing accuracy quantitatively or to compare accuracy between tests or predictive models. The ROC curve was originated in the theory of signal detection in the years 1950-1960 (Green and Swets, 1966; Egan, 1975) to discriminate between signal and noise. It has been used in so many areas such as radiology (Metz, 1989), psychiatry (Hsiao et al, 1989), epidemiology (Aoki et al, 1997), biomedical informatics (Lasko et al, 2005). It can provide a direct and visual comparison of two or more diagnostic tests on a single set of scales. It is possible to compare different tests at all decision cut-offs by constructing the ROC curves. For statistical analysis, a recommended numerical index of accuracy associated with an ROC curve are often better used to summarize the information provided for the ROC curve into a single global value or index (Swets and Picket, 1982). This index is called area under the ROC curve. AUC takes values between 0.5 (which corresponds to the diagonal ROC curve that passes through the points (0,0) and (1,1)) and 1 (representing perfect test where all cases are correctly classified). AUC represents the diagnostic accuracy of the test Y, so that the larger the area the better the diagnostic accuracy of Y. This means that values closer to 1 indicate that Y optimally discriminates between healthy and diseased subjects, while values near 0.5 indicate that the test is not informative (Zhou et al, 2002). According to Mann-Whitney (1947), AUC is the probability that the observed

test result of a randomly selected subject from the diseased population (Y_1) is larger than the observed test result of a randomly selected subject from the non-diseased population (Y_0). For comparing two diagnostic processes, the difference between AUCs is often used. In diagnostic imaging it is generally known that the changes due to subjects represent a major component of the overall changes of the AUC. To better control for the sources of changes when comparing diagnostic tests, a paired study design is often advised because it usually induces positive

correlation between the tests results of the same subjects.

This paper is devoted to reviewing some existing methods of calculating AUC. It attempts to calculate AUC based on a simple new method and evaluate its significance in assessing classification models.

EXISTING METHODS OF CALCULATING AUC

There are several methods of calculating the AUC. All these methods differ in the way the distribution functions of both populations are estimated based on their sample values. In Lopez-Raton et al (2012a), a review of all these methods to estimate ROC curve and AUCs is performed. The two basic methods of estimating AUC are the parametric (bi-normal) and non-parametric (empirical).

Parametric (Binormal ROC curve) method

This estimation is based on the assumption that the test results in the diseased and non-diseased populations or some unknown monotonic transformation of the test results follows a bi-normal distribution. Dorfman and Alf, Jr (1968, 1969) and McClish (1989) proposed maximum-likelihood estimates (MLEs) for the parameters of a bi-normal ROC curve and provided parametric methods for estimating and comparing the partial AUC respectively.

According to Dorfman and Alf (1968), McClish (1989), Metz (1978), suppose that continuous diagnostic test results are normally distributed in the healthy and diseased populations, to be able to properly define ROC curve, let Y denote a random variable representing a continuous diagnostic test result. The diagnosis according to any cutoff value c is positive (diseased) if $Y \geq c$ and negative (non-diseased) if $Y < c$. Let D_0 and D_1 denote the non-diseased and diseased populations, respectively. The true and false positive rates at the cut-off value c , true positive rate, $TPR(c)$, and false positive rate, $FPR(c)$ are

$$\left. \begin{aligned} TPR(c) &= P(Y \geq c | D_1) \\ FPR(c) &= P(Y \geq c | D_0) \end{aligned} \right\} \quad 1.11$$

Let

$$Y_1 \sim N(\mu_1, \sigma_1^2), Y_0 \sim N(\mu_0, \sigma_0^2) \quad 1.12$$

Where (μ_1, σ_1^2) and (μ_0, σ_0^2) are the population mean and variance of the disease and non-diseased group. For any cut-off value c , we have

$$\left. \begin{aligned} TPR(c) &= P(Y > c) = \Phi\left(\frac{\mu_1 - c}{\sigma_1}\right) \\ FPR(c) &= P(Y > c) = \Phi\left(\frac{\mu_0 - c}{\sigma_0}\right) \end{aligned} \right\} \quad 1.13$$

where Φ denotes the standard normal cumulative distribution function. Note also that t is the all possible FPRs according to the varying c values in $(-\infty, \infty)$. Simplifying, we have

$$\begin{aligned} \Phi\left(\frac{\mu_0 - c}{\sigma_0}\right) &= t \\ \Phi^{-1}(t) &= \frac{\mu_0 - c}{\sigma_0} \\ \sigma_0 \Phi^{-1}(t) &= \mu_0 - c \\ c &= \mu_0 - \sigma_0 \Phi^{-1}(t). \end{aligned} \quad 1.14$$

Equation 1.14 is the corresponding cut-off value.

Hence

The ROC curve is obtained by substituting for c in equation 1.13 to have the function

$$ROC(t) = \Phi\left(\frac{\mu_1 - c}{\sigma_1}\right) = \Phi\left(\frac{\mu_1 - \mu_0}{\sigma_1} + \frac{\sigma_0}{\sigma_1} \Phi^{-1}(t)\right). \quad 1.15$$

If according to McClish(1989) and Metz(1978), we define $a_1 = \frac{\mu_1 - \mu_0}{\sigma_1}$ and $a_0 = \frac{\sigma_0}{\sigma_1}$, then the ROC curve

$$ROC(t) = \Phi \left(a_1 + a_0 \Phi^{-1}(t) \right) \tag{1.16}$$

In order to show that AUC measure has an analytic form, we recall and show that $AUC = P(Y_1 > Y_0)$ under the binormal or bi-lognormal ROC curve,

Let Y_0 and Y_1 be two random vectors that are independently drawn from the healthy and diseased population and Y_0 follows lognormal distribution with mean, μ_0 and variance, σ_0^2 and Y_1 follows lognormal distribution with mean, μ_1 and variance, σ_1^2 . We here use log transformation to the random vectors Y_0 and Y_1 . Since Y_0 and Y_1 follows log-normal distribution, then logarithm of Y_0 and Y_1 follows normal distribution. Let $\ln Y_1 = y_1$ and $\ln Y_0 = y_0$ be for diseased and non-diseased group respectively, since it is easier to work with normal distribution than the lognormal distribution,

Then $AUC = P(Y_1 > Y_0) = P(\ln Y_1 > \ln Y_0) = P(y_1 > y_0)$

This follows that $y_1 - y_0 \sim N(\mu_1 - \mu_0, \sigma_1^2 + \sigma_0^2)$ implying that $P(y_1 - y_0 > 0) = P\left(Z > \frac{\mu_0 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_0^2}} \right)$

$$AUC = \Phi \left(\frac{\mu_1 - \mu_0}{\sqrt{\sigma_1^2 + \sigma_0^2}} \right) \tag{1.17}$$

Where μ_1 , and μ_0 are the means and σ_1^2, σ_0^2 are the variances of the result y_1 and y_0 respectively. This is the form AUC takes if binormality is assumed. The AUC can be obtained numerically by substituting the estimated values of the parameters given as $\hat{\mu}_1, \hat{\mu}_0, \hat{\sigma}_1^2, \hat{\sigma}_0^2$ from the sample data.

Alternative method of calculating AUC.

Based on the bi-normal assumption, we know that

$$Z_1 = \frac{y_1 - \mu_1}{\sigma_1} \sim N(0,1) \text{ and } Z_0 = \frac{y_0 - \mu_0}{\sigma_0} \sim N(0,1)$$

$$AUC = P(Y_1 > Y_0) = P(\ln Y_1 > \ln Y_0) = P(y_1 > y_0) = P(Z_1 \sigma_1 + \mu_1 > Z_0 \sigma_0 + \mu_0)$$

$$= P(Z_1 \sigma_1 - Z_0 \sigma_0 > \mu_0 - \mu_1) = P\left(\frac{(Z_1 \sigma_1 - Z_0 \sigma_0) - 0}{\sqrt{\sigma_0^2 + \sigma_1^2}} > \frac{(\mu_0 - \mu_1) - 0}{\sqrt{\sigma_0^2 + \sigma_1^2}} \right)$$

$$= P\left(Z > \frac{\mu_0 - \mu_1}{\sqrt{\sigma_0^2 + \sigma_1^2}} \right) = 1 - P\left(Z \leq \frac{\mu_0 - \mu_1}{\sqrt{\sigma_0^2 + \sigma_1^2}} \right) = 1 - \Phi\left(\frac{\mu_0 - \mu_1}{\sqrt{\sigma_0^2 + \sigma_1^2}} \right) = 1 - \Phi\left(-\left[\frac{\mu_1 - \mu_0}{\sqrt{\sigma_0^2 + \sigma_1^2}} \right] \right) = \Phi\left[\frac{\mu_1 - \mu_0}{\sqrt{\sigma_0^2 + \sigma_1^2}} \right]$$

Therefore

$$AUC = P(Y_1 > Y_0) = \Phi\left[\frac{\mu_1 - \mu_0}{\sqrt{\sigma_0^2 + \sigma_1^2}} \right] = \Phi\left(\frac{a_1}{\sqrt{1 + a_0^2}} \right) \tag{1.18}$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution function (Pepe,2003;Metz,1986;Zou et al,1997;Hanley and McNeil,1982).

Estimating Variance of AUC

The respective variances of \hat{a}_1 and a_0 are

$$V(\hat{a}_1) = \frac{n_1(\hat{a}_1^2 + 2) + 2n_0a_0^2}{2n_0n_1}; V(a_0) = \frac{(n_1 + n_0)\hat{a}_0^2}{2n_0n_1} \quad 1.19$$

where n_1 and n_0 are the numbers of diseased and non-diseased study subjects, respectively. The variance of $A\hat{U}C$ can be estimated by substituting estimators for the parameters a_1 and a_0 . To estimate the variance of AUC, recall that

$$AUC = \Phi\left(\frac{\mu_1 - \mu_0}{\sqrt{\sigma_1^2 + \sigma_0^2}}\right) = \Phi(\delta) \quad 1.20$$

The maximum likelihood estimate of δ is obtained substituting the MLE's of the means and variances. The MLE of AUC is found by substituting the MLE's of the means and variances into equation 1.20 and using numerical integration. The AUC now reduce to

$$A\hat{U}C = \Phi\left(\frac{\hat{a}_1}{\sqrt{1 + \hat{a}_0^2}}\right) = \Phi(\hat{\delta}) \quad 1.21$$

Since $\Phi(\hat{\delta})$ is a monotonic increasing function of $\hat{\delta}$, it is enough to find the variance and standard error of $\hat{\delta}$. Since δ is a function of the parameters, μ_1, μ_0, σ_1^2 and σ_0^2 , we will adopt the Delta method (Zhou et al, 2002) for finding the approximate variance and standard error for $\hat{\delta}$.
 By definition,

$$\begin{aligned} V(\hat{\delta}) &= \left(\frac{\partial \delta}{\partial \mu_1}\right)^2 V(\mu_1) + \left(\frac{\partial \delta}{\partial \mu_0}\right)^2 V(\mu_0) + \left(\frac{\partial \delta}{\partial \sigma_1^2}\right)^2 V(\sigma_1^2) + \left(\frac{\partial \delta}{\partial \sigma_0^2}\right)^2 V(\sigma_0^2) \\ &\quad + 2\left(\frac{\partial \delta}{\partial \mu_1}\right)\left(\frac{\partial \delta}{\partial \mu_0}\right)Cov(\mu_1, \mu_0) + 2\left(\frac{\partial \delta}{\partial \sigma_1^2}\right)\left(\frac{\partial \delta}{\partial \sigma_0^2}\right)Cov(\sigma_1^2, \sigma_0^2) \\ &= \left(\frac{1}{\sqrt{\sigma_1^2 + \sigma_0^2}}\right)^2 \frac{\sigma_0^2}{n_0} + \left(-\frac{1}{\sqrt{\sigma_1^2 + \sigma_0^2}}\right)^2 \frac{\sigma_1^2}{n_1} + \left(\frac{-(\mu_1 - \mu_0)}{(\sigma_1^2 + \sigma_0^2)^{\frac{3}{2}}}\right)^2 \frac{2\hat{\sigma}_0^4}{n_0 - 1} + \left(\frac{-(\mu_1 - \mu_0)}{2(\sigma_1^2 + \sigma_0^2)^{\frac{3}{2}}}\right)^2 \frac{2\hat{\sigma}_1^4}{n_1 - 1} \end{aligned}$$

Hence the variance expression for $\hat{\delta}$ can be obtained using the following expression

$$V(\hat{\delta}) = \frac{1}{\sigma_1^2 + \sigma_0^2} \left\{ \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0}\right) + \frac{(\mu_1 - \mu_0)^2}{2(\sigma_1^2 + \sigma_0^2)^3} \left[\frac{\sigma_1^4}{(n_1 - 1)} + \frac{\sigma_0^4}{(n_0 - 1)}\right] \right\} \quad 1.22$$

The variance of $A\hat{U}C$ or $\hat{\delta}$ can be obtained by substituting the estimated values of μ_1, μ_0, σ_1^2 and σ_0^2 in the above expression. The standard error $(SE(\hat{\delta}))$ of $\hat{\delta}$ is just the square root of the above expression for the variance of estimated AUC. To determine the asymptotic $100(1-\alpha)$ % confidence interval for AUC, we have

$$C.I(AUC) = \Phi\left(\hat{\delta} \pm Z_{\alpha/2} \sqrt{V(\hat{\delta})}\right) \quad 1.23$$

Where α is the level of significance and $Z_{\alpha/2}$ is the critical value of Z for a two tailed test at level of significance $\alpha/2$.

NON-PARAMETRIC (EMPIRICAL ROC CURVE) METHOD OF CALCULATING AUC

This method is employed if the normal assumption is violated or tenuous. There are distribution-free methods.

Some methods using the empirical ROC curve are the trapezoidal rule to approximate AUC by integration, the Mann-Whitney statistic and the empirical method by DeLong et al. (1988).

Nonparametric (conventional) Delong et al method

The empirical (nonparametric) method by DeLong et al. (1988) is a popular and best known method to compare two correlated AUCs by using the theory of generalized U statistics. They used the structural components method provided by Sen (1960) to generate consistent variance estimates of the elements of the variance-covariance matrix

of a vector of U statistics, and the resulting test statistic has asymptotically a χ^2 distribution. This method is important as it helps to study the behavior of the type I error and the statistical power of the conventional nonparametric test for comparing two AUCs over a wide range of relevant parameters and against various alternatives. According to DeLong et al. (1988), let the variance of Y_{1i} being the component of the i th subject from

the diseased population, $V(Y_{1i})$ be defined as

$$V(Y_{1i}) = \frac{1}{n_0} \sum_{j=1}^{n_0} \Psi(Y_{1i}, Y_{0j}) \quad 1.24$$

and the variance of Y_{0j} being the component of the j th subject from the healthy population, $V(Y_{0j})$ be defined as

$$V(Y_{0j}) = \frac{1}{n_1} \sum_{i=1}^{n_1} \Psi(Y_{1i}, Y_{0j}) \quad 1.25$$

where n_1 and n_0 are the number of diseased and non-diseased subjects respectively. While Y_{1i} and Y_{0j} are the diagnostic test result of the i th subject with disease and the diagnostic test result of the j th subject without disease respectively and Ψ is a function comparing Y_{1i} and Y_{0j} such that

$$\Psi(Y_{1i}, Y_{0j}) = \begin{cases} 1 & \text{if } Y_{1i} > Y_{0j}, \\ 0.5 & \text{if } Y_{1i} = Y_{0j}, \\ 0 & \text{if } Y_{1i} < Y_{0j}, \end{cases}$$

The empirical AUC is estimated as

$$AUC_{Empirical} = \sum_{i=1}^{n_1} V(Y_{1i})/n_1 = \sum_{j=1}^{n_0} V(Y_{0j})/n_0 \quad 1.26$$

The variance of the estimated AUC is estimated as

$$V(AUC_{Empirical}) = \frac{1}{n_1} S_{Y_1}^2 + \frac{1}{n_0} S_{Y_0}^2 \quad 1.27$$

Where $S_{Y_1}^2$ and $S_{Y_0}^2$, the variances of the diseased and non-diseased subjects is defined as

$$S_{Y_i}^2 = \frac{1}{n_i - 1} \sum_{i=1}^{n_i} [V(Y_{1i}) - AUC_{Empirical}]^2, i = 0, 1.$$

The variance of $S_{Y_0}^2$ is similarly defined.

THE TRAPEZOIDAL RULE AND MANN-WHITNEY U STATISTIC METHODS

To calculate the AUC using the Trapezoidal rule, first the ROC curve is separated into many segments, the area of each segment is computed by joining the points (sensitivity, Se, 1-specificity, Sp) at each interval value of the continuous test results and draws a straight line joining the x-axis. This forms several trapezoids and the AUC can be easily calculated directly by summing the area of the trapezoids that are formed below the connected points making up the ROC curve.

According to Bamber(1975), Hanley and McNeil (1982), the area under an empirical ROC curve, when calculated by the trapezoidal rule, is equal to the Mann-Whitney two sample statistics applied to the two samples since the nonparametric analog to the t-test is the Wilcoxon rank-sum test, or synonymously the Mann-Whitney U test. Here, the possible diagnostic test results for each cutoff value c are considered, and the corresponding true

and false positive rates are calculated by

$$\left. \begin{aligned} T P R (c) &= \frac{s_1 (c)}{n_1} \\ F P R (c) &= \frac{s_0 (c)}{n_0} \end{aligned} \right\} \quad 1.28$$

where $s_1(c)$ is the number of subjects with test results greater than or equal to $c (Y \geq c)$ among the diseased subjects and $s_0(c)$ is the number of subjects with test results greater than or equal $c (Y \geq c)$ among the non-diseased subjects. The ROC curve is subsequently created by connecting these points with a straight line (Bamber, 1975; Hanley and McNeil, 1982). The AUC of the nonparametric ROC curve is obtained using trapezoidal rule and is estimated by

$$A \hat{U} C = \frac{1}{n_1 n_0} \sum_{i=1}^{n_1} \sum_{j=1}^{n_0} \psi (Y_{i1}, Y_{j0}) \quad 1.29$$

where n_1 and n_0 are the number of diseased and non-diseased subjects respectively. While Y_{i1} and Y_{j0} are the diagnostic test result of the i th subject with disease and the diagnostic test result of the j th subject without disease respectively and Ψ is a function comparing Y_{i1} and Y_{j0} such that

$$\psi (Y_{i1}, Y_{j0}) = \begin{cases} 1 & \text{if } Y_{i1} > Y_{j0} \\ 0.5 & \text{if } Y_{i1} = Y_{j0} \\ 0 & \text{if } Y_{i1} < Y_{j0} \end{cases} \quad 1.30$$

AUC is estimated by substituting for equation 1.30 in equation 1.29 to yield

$$A \hat{U} C = \frac{1}{n_1 n_0} \sum_{i=1}^{n_1} \sum_{j=1}^{n_0} \left[\Psi \{ Y_{i1} > Y_{j0} \} + \frac{1}{2} \Psi \{ Y_{i1} = Y_{j0} \} \right] \quad 1.31$$

Since the Mann-Whitney U-statistic is based upon an estimate of $P(Y_1 > Y_0)$, which is exactly the AUC, the properties of the Mann-Whitney U-statistic can be used to predict the statistical properties of the AUC (Hanley and McNeil, 1982). The variance of the estimated AUC is computed using Mann-Whitney Statistic (Hanley and McNeil, 1982) as:

$$V(A \hat{U} C) = \frac{A \hat{U} C (1 - A \hat{U} C) + (n_1 - 1) (Q_1 - A \hat{U} C^2)}{n_1 n_0} + \frac{(n_0 - 1) (Q_2 - A \hat{U} C^2)}{n_1 n_0} \quad 1.32$$

Where Q_1 and Q_2 are defined as

$$\left. \begin{aligned} Q_1 &= \frac{1}{n_0 n_1^2} \sum_Y n_0^{-y} \times \left[(n_1^{>y})^2 + n_1^{>y} \times n_1^{-y} + \frac{(n_1^{-y})^2}{3} \right] \\ Q_2 &= \frac{1}{n_0^2 n_1} \sum_Y n_1^{-y} \times \left[(n_0^{<y})^2 + n_0^{<y} \times n_0^{-y} + \frac{(n_0^{-y})^2}{3} \right] \end{aligned} \right\} \quad 1.33$$

where n_0^{-y} is the number of true negative subjects with test results equal to y , n_1^{-y} is the number of true positive subjects with test results equal to y , $n_0^{<y}$ is the number of true negative subjects with test results less than y and $n_1^{>y}$ is the number of true positive subjects with test results greater than y (Hanley and McNeil, 1982).

The trapezoidal approach systematically underestimates the AUC, because of the way all of the points on the ROC curve are connected with straight lines, rather than smooth concave curves which is normally experienced when the Mann-Whitney U statistic approach is used for estimation (Zhou et al, 2002). By increasing the number of the possible cut-off points, the bias of the estimation of AUC can be significantly reduced and make it acceptable for the estimation. Hanley and McNeil (1983) showed that the area computed by trapezoidal rule under an empirical ROC curve is equal to the Mann-Whitney U statistic for comparing correlated AUCs from two samples. In general the interpretation of the AUC is the same regardless of whether trapezoidal rule or the Mann-Whitney U statistic

was used. A way to compare the trapezoidal rule and the Mann-Whitney U statistic in estimating AUC is to compare their respective AUCs. In order to determine if the two AUCs are significantly different, the variances of both AUCs estimates must be taken into account.

CLASSIFICATION MODELS ASSESSMENT

Classification models discussed generally here includes logistic regression, discriminant analysis and dummy variable multiple regression because of their similarity in models and estimation techniques. Classification modeling has the purpose of finding a mathematical relationship between a response, or “dependent” variable and “independent” variables with the purpose of estimating the independent variables through which future values of the response variable could be predicted from the estimated variable(s). We shall verify the significance of AUC in assessing classification models. For instance, based on the previous works by Okeh and Oyeka (2014,2015,2016), on dummy variable regression analysis, predicted probability of positive response can be obtained and applied in calculating AUC for assessing the dummy variable regression model. This assessment is based on the p-value of the AUC calculated. So much have been said about some of these models: Ogun(2003) said that discriminant analysis is a rather powerful statistical tool when many variables (e.g. patients risk factors for diabetes) are to be considered simultaneously while Onyeagu (2003) viewed discriminant analysis as a technique concerned with the problem of classification in the sense that the output generated from these classification models belong to a certain range of values defined by the cut-off value c . This cut-off value if known or assumed can always be applied on the basis of which classification into groups of presence or diseased (coded 1) or negative or non-diseased (coded 0) is made. Applying the cut-off value c in dichotomizing a continuous or discrete outputs into the range $[1,0]$, so that this output $y(x)$, is set such that

$$y(x) = \begin{cases} 1, & \text{if } x \geq c \\ 0, & \text{otherwise} \end{cases} \quad 1.34$$

And can be used in constructing a contingency table from where a set of pairs of sensitivity and 1-specificity can be obtained for constructing the ROC curve and calculating AUC. The alternative method developed here in building a contingency table required having observed outcomes generated by equation 1.34 and predicted outcomes obtained from predicted probability of positive response. This method will be demonstrated as a new method under methodology.

METHODOLOGY

Here we propose a simpler method of calculating AUC from predicted probability of positive response obtained from classification models. We shall also propose an easy to understand method of comparing between two or more diagnostic tests in terms of their AUCs. To define the AUC, this paper adopted the pattern of Mann-Whitney (MW, 1947) statistic approach of calculating AUC based on predicted probability of positive response from classification models and observed outcomes. In particular to specify the observed subjects’ outcomes, let

$$y_{ij} = \begin{cases} 1, & \text{if } x_{ij} \text{ is the sample test result of the } i\text{th subject} \\ & \text{screened and tested positive at } j\text{th diagnostic test} \\ 0, & \text{otherwise} \end{cases} \quad (2.1)$$

for $i = 1, 2, \dots, n$ subjects; $j = 1, 2, \dots, T$ (diag. test)

In other to predict the outcome values or obtain the predicted subjects’ outcomes using the predicted probability of positive response denoted as, $\hat{P}_{ij}^+(1)$, we propose to develop multiple rules, called α -prediction rules which

represents unbiased estimate of an indicator variable $(y_{ij,\alpha})$ used in modeling AUC and this is given as

$$\hat{y}_{ij,\alpha} = \begin{cases} 1, & \text{if } \hat{P}_{ij}^+(1) \geq 1 - \alpha; 0 \leq \alpha \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad 2.2$$

Where $\hat{y}_{ij,\alpha}$ is the predicted outcome (in row) for i th subject at j th diagnostic test, $\hat{P}_{ij}^+(1)$ is the predicted probability of positive response of i th subject at the j th diagnostic test. Also α is some probability real numbers

given the fact that the graph of ROC curve is within this probability range. Equation 2.2 implies that higher values of α means that more predicted outcome values will be 1 while lower values of α indicates that more predicted outcome values will be zero.

CONSTRUCTING ROC CURVE BASED ON THE α PREDICTION RULES

To construct ROC curve, let Y be a random variable representing a continuous diagnostic test result. To dichotomize these test results, we assume that there exist a cut-off value c such that test results of the population of diseased subjects(coded 1) are classified based on $Y \geq c$ and the population of non-diseased subjects(coded 0) are classified based on $Y < c$.

Mathematically, using a cut-off value c to define a dichotomous test result from a continuous diagnostic tests $Y((y \in -\infty, +\infty))$ as diseased ($D=1$) if $Y \geq c$ and non-diseased ($D=0$) if $Y < c$.

To construct ROC curve, let a total random sample of $a+c$ subjects be observed outcomes $(y_{ij} = 1)$ that are known or believed on the basis of diagnosis to actually have the disease in a population and a total random sample of $b+d$ subjects be observed outcomes $(y_{ij} = 0)$ from the same population known or believed on the basis of diagnosis not to actually have the disease. Similarly, let a total of $a+b$ subjects be predicted outcomes $(\hat{y}_{ij,\alpha} = 1)$ that are estimated to have actually tested positive to the disease and let a total of $c+d$ subjects be predicted outcomes $(\hat{y}_{ij,\alpha} = 0)$ estimated to have actually tested negative to the disease. The resulting four fold contingency table for a given α -prediction rule is Table 1

Table 1: Contingency Table notation for a given α -prediction rule

Predicted outcomes	Observed outcomes		Total
	$y_{ij} = 1$	$y_{ij} = 0$	
$\hat{y}_{ij,\alpha} = 1$	a	b	$a + b$
$\hat{y}_{ij,\alpha} = 0$	c	d	$c + d$
Total	$a + c$	$b + d$	$n = n_{..} = a + b + c + d$

Note that a or true positive(TP) is the number of subjects that are observed to have disease $(y_{ij} = 1)$ and predicted to have disease also $(\hat{y}_{ij,\alpha} = 1)$, b or false positive(FP) is the number of subjects that are observed not to have disease $(y_{ij} = 0)$ and predicted to have disease $(\hat{y}_{ij,\alpha} = 1)$, c or false negative(FN) is the number of subjects that are observed to have disease $(y_{ij} = 1)$ and predicted not to have disease also $(\hat{y}_{ij,\alpha} = 0)$, d or true negative(TN) is the number of subjects that are observed not to have disease $(y_{ij} = 0)$ and predicted not to have disease also $(\hat{y}_{ij,\alpha} = 0)$ and $n = n_{..} = a + b + c + d$ is the total number of subjects randomly sampled from the population.

Based on the results in a classification table 1 for each model, we carry out ROC curve analysis to measure the accuracy of the diagnostic test in discriminating between alternative health status by first calculating the sensitivity and 1-specificity since ROC curve is a graph resulting from the plotting of these values. By varying the value of α in Equation 2.2 between 0 and 1 inclusive, we generate so many predicted outcomes that can be represented in there corresponding contingency tables for each model. Based on these tables, we calculate values of sensitivity and 1-specificity that will be used in constructing an ROC curve. Specifically here, each α -rule contributes one point to the ROC curve and so since a pair of sensitivity and 1-specificity values contributes one point to the ROC curve, several pairs will be used in constructing a smooth ROC curve. This method of defining AUC represents a shift from the existing works by Buros and Tubbs (2013); Hanley and McNeil (1982) and Mann-Whitney (1947).

CALCULATING AUC BASED ON PREDICTED PROBABILITY OF POSITIVE RESPONSE

To model the AUC based on the contingency table 1, let y_{1ij} and y_{0ij} be the diagnostic test results of the i th subject at j th diagnostic test who are drawn randomly from the diseased and non-diseased population respectively while n represents the total number of sampled observations of subjects for those responding positive (diseased)

and negative(non-diseased) from the population, then the AUC is defined as

$$A\hat{U}C = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^T \Omega(y_{1ij}, y_{0ij}), i = 1, 2, \dots, n \text{ subjects}; j = 1, 2, \dots, T (\text{diag.test}) \quad 2.3$$

where

$$\Omega(y_{1ij}, y_{0ij}) = \begin{cases} 1, & \text{if } y_{1ij} > y_{0ij} \\ 0.5, & \text{if } y_{1ij} = y_{0ij} \\ 0, & \text{if } y_{1ij} < y_{0ij} \end{cases}$$

Equivalently

$$A\hat{U}C = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^T \left[\Omega(y_{1ij} > y_{0ij}) + 0.5 \Omega(y_{1ij} = y_{0ij}) \right] \quad 2.4$$

The AUC defined by equations 2.3 or 2.4 follows the non-parametric MW approach and it is rather flexible and yields ROC estimates even with a better precision than the MW approach or the trapezoidal rule for calculating the AUC. This method of calculating AUC avoids the computational complex procedures of the maximum likelihood estimation (MLE) and numerical integration methods which not only involves lengthy calculations but also have restrictive assumptions about the distribution of diagnostic test results. It is note worthy that estimates from parametric methods such as the method of MLE are inconsistent thereby giving a misleading picture of the regression relationship (Pepe, 2003). Our method of calculating AUC is unique in so many ways: it incorporates predicted probability of positive response in the construction of ROC curve and indeed AUC, it uses the α prediction rule which enables the construction of very smooth ROC curve because several values of α normally will produce smoothness in the curve, the AUC calculated is also diagnostic test dependent since the test result depends on the test for the subjects and finally the method is not only simpler and straight forward but also it avoids the iteration procedure which is rigorous, time consuming and liable to errorneous results. The new AUC if obtained for two or more diagnostic tests can be compared using a chi-square test statistic proposed for that purpose which approximates continuous distribution to discrete distribution as seen in a contingency table.

APPLICATION TO REAL DATA

The proposed methods can be applied to real data on gestational diabetes mellitus(GDM).This was a retrospective study of test results of pregnant women screened using 1 hour 50g Glucose Challenge Test(GCT) and diagnosed using 75g OGTT as well as 100g OGTT according to WHO(1999) and National Diabetic Data Group(NDDG,1979) criteria. These test results were collected using the simple random sampling method. Medical records showed that out of a total of six thousand and ten (6010) pregnant women registered for antenatal clinic (ANC) who were screened using a universal screening with 1 hour 50g Glucose Challenge Test (GCT) for GDM in the sampled hospitals within the two years (from January 2011 to December 2012) chosen for this study, a total sample of 1113 pregnant women who had positive risk factors (such as positive family history of diabetes, age at least 30 years, BMI ≥ 30 kg/m², previous fetal weight ≥ 4 kg, and positive obstetric history of GDM) and aged between 15-45 years at less than 24 weeks and between 24-28 weeks of gestation tested positive for GDM(indicating the presence of GDM) since their plasma blood glucose level was at least 140 mg/dl after 1 hour. These positively responding women were subsequently recalled for confirmatory diagnostic test using 2-hour 75g OGTT in accordance with the criteria set by WHO (1985) and later repeated using 3-hours 100g OGTT during the later part of their gestation period.

The study protocol was according to the recommendations for universal screening by the Fifth International Workshop Conference on Gestational Diabetes (Metzger et al, 2007).The essence of the repeated tests was to actually determine the status of GDM in them. Since the results of the diagnosis are two, one stands the chance of comparing between their tests results (between 2-hours 75g OGTT and 3-hours 100g OGTT). Women who were known diabetics, or who were suffering from any chronic illness were excluded from the study. After obtaining permission from the hospitals' Research and Ethics Committee, assess was granted into the record units of the antenatal wards of these hospitals where the medical history of the patients were kept in a proforma containing general information on demographic characteristics such as body mass index, maternal age, previous fetal weight and vital clinical histories such as obstetric history of GDM, and family history of diabetes were taken. BMI was calculated by dividing the weight in kilograms by the height in meters squared.

The data for this paper is recorded according to the number of pregnant women, GDM test results (response variable) and their observed risk factors or parent independent variables. This is suitable for fitting the classification models for analysis. Since we have two set of response variable from the two diagnostic tests, we took the average of the test results as observed outcomes. Analysis of this data will yield the predicted probability of positive response needed to be applied in calculating the AUC for the models.

PREDICTED PROBABILITY OBTAINED FROM THE CLASSIFICATION MODELS

We here calculate predicted probability of positive response from the classification models for the purpose of calculating the AUC. Fitting the data to the models, the analysis yielded the results shown in table 2.

Table 2. Predicted probability from the classification models

Classification models	Predicted probability	p-value	F-ratio
Logistic Regression	0.1949	0.0251	10.346
Linear Discriminant Analysis	0.1184	0.0425	6.6421
Dummy variable regression	0.4785	0.0013	15.843

Having obtained the predicted probability of positive response for each of the models as in table 2, we now use each of them to find the predicted outcomes as defined in equation 2.2 as well as the observed outcomes as defined in equation 2.1. Since ROC curve has the ability of evaluating the discriminatory power of a continuous test result (observed outcome) to correctly assign into a two-group classification, the observed outcomes is dichotomized using at least 7.8mmol/l as cut-off value for GDM diagnosis as recommended. Having generated the observed outcomes based on equation 2.1 and predicted outcomes based on equation 3.2, the resulting coded data from these outcomes are cross-classified and will constitute as many tables as contingency table 2.5 for constructing ROC curve for each model. For instance, the contingency table 3 represents a pair of sensitivity and 1-specificity out of 1113 pairs required for constructing the ROC curve for each model.

Table 3. Outcome values for a given α -prediction rule

Predicted Outcomes	Observed Outcomes		
	Diseased	Not diseased	Total
Positive	n_{11}	n_{12}	$n_{1.}$
Negative	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{..} = n$

Suppose each α -rule (for instance, table 3) contributes one point to the ROC curve, the estimates of sensitivity and specificity obtained from this table for generating the point is not enough to be used to obtain sufficient pairs of sensitivity and 1-specificity that will enable for the actual smooth ROC curve analysis for each model. In order to have sufficient estimated pairs of sensitivity and 1-specificity that can generate a smooth curve, we vary the α value from 0 to 1 in the prediction rule of equation 2.2 so as to generate multiple prediction rule capable of being used in obtaining enough contingency tables for the construction of ROC curve. This computation is supported by SAS version 9 soft ware. The difference between one model and another model is the predicted probability of positive response for that model. The number of pairs of sensitivity and 1-specificity equals a sample size 1113 for each model. Using this method, the AUCs are obtained for the models.

ASSESSING THE DISCRIMINATING ABILITY OF CLASSIFICATION MODELS

Assessment of classification models here is based on their discriminating or predictive ability of diagnostic accuracy. Since many models are available for assessment, interpreting and comparing the models using the ROC curves may be erroneous, instead the interpretation and comparison of the discriminatory accuracy of the test will be based on the AUC which summarizes the accuracy of each model. It is vital to note that these models were chosen because of their similar set of modeling techniques. The estimators of AUC for the selected models are estimated using both parametric and non-parametric methods. From Table 4, result shows that little differences exist among the non-parametric estimates than the parametric estimates. The highest difference in AUC can be seen in linear discriminant analysis. This may be due to its strict compliance to the normality assumption as well as equal variance.

Table 4. CLASSIFICATION MODELS AGAINST METHODS OF CALCULATING AUC

	Classification models	Binormal parametric mtd(parametric mtd)	Predicted probability method (nonparametric mtd)	Mann-Whitney U statistic method (nonparametric mtd)	Trapezoidal Rule (nonparametric mtd)
1	Logistic Regression	0.6284832	0.6034563	0.6007869	0.6034212
2	Dummy Variable Regression	0.6194562	0.6004383	0.6003231	0.6007853
3	Linear Discriminant Analysis	0.6384753	0.6013955	0.6002135	0.6011457

NORMALITY TEST FOR THE CLASSIFICATION MODELS

The test of normality for the models shows that after estimating maximum likelihood estimate $\hat{\lambda}$ by power transformation, it was discovered that only discriminant analysis is the distribution (Table 5) that have the normal distribution on the basis of the specification of Shapiro-Wilk normality test say at $\alpha = 0.05$. This test suggests that if the chosen alpha level is 0.05 and the p-value is less than 0.05, then the null hypothesis that the data are normally distributed is rejected. If the p-value is greater than 0.05, then the null hypothesis is accepted. Similarly, two test results such as X(diseased) and Y(non-diseased) may not follow normal distribution according to Krzyśko et al (2008), but the fact that parametric binormal ROC curve suggests that two test results such as X(diseased) and Y(non-diseased) must each follow normal distribution always give good results, due to the fact that ROC curves concerns itself with the relationship between distributions instead of the individual distributions.

Table 5 Test of Normality for the classification models

		GDM STATUS			
		Diseased=1		Non-diseased=0	
		$\hat{\lambda}$	p-value	$\hat{\lambda}$	p-value
1	Logistic Regression	1.22	0.0350	1.19	1.05E-26
2	Dummy Variable Regression	0.52	0.0112	0.74	1.03E-04
3	Linear Discriminant Analysis	0.63	0.1064	0.69	2.16E-07

Information in cells of Table 6 shows the p-values for the difference between two AUCs when the predicted probability method of calculating AUC is used. The null hypothesis of no difference in AUCs is rejected when any of the p-value is smaller than the level of significance ($\alpha = 0.05$). Table 2.8 showed that the AUCs for linear discriminant analysis did not show any significant difference because the assumptions of normality and equal covariance are not met while dummy variable regression analysis followed by logistic regression analysis showed significant differences in AUCs indicating their flexibility in handling data when the assumption of normality and equal variance is violated.

TABLE 6. TESTING THE DIFFERENCE BETWEEN TWO MEASUREMENTS OF AUC USING P-VALUES

	Logistic Regression	Dummy Variable Regression	Linear Discriminant Analysis
Logistic Regression	0.02456234	0.013384476	0.03485746
Dummy Variable Regression		0.012759491	0.01625433
Linear Discriminant Analysis			1.00000000

SUMMARY AND CONCLUSION

The purpose of this paper was to evaluate the methods of calculating AUC and its significance in assessing classification models. In assessing these classification techniques, it was discovered that non-parametric methods of estimating AUC give convergent results in terms of their measurement for AUC while parametric approaches are known generally for the higher values of AUC. Practically, the assumption of normality cannot be achieved for the parametric methods of estimating AUC. This is why the non-parametric methods are recommended. The strength of our method is that it has easy implementation to discriminate diagnostic test procedures even by non-statisticians. The proposed method offers reliable statistical inferences even in small sample problems and circumvent the difficulties of deriving the statistical moments of complex summary statistics.

DISCUSSION

The method of calculating AUC from predicted probability of positive response avoids the computational complex

procedures of the maximum likelihood estimation (MLE) and numerical integration methods which not only involves lengthy calculations but also have restrictive assumptions about the distribution of diagnostic test results since there are parametric methods. It is note worthy that estimates from parametric methods such as the method of MLE are inconsistent thereby giving a misleading picture of the regression relationship (Pepe, 2003). Our method of calculating AUC is unique in so many ways: it incorporates predicted probability of positive response in the construction of ROC curve and indeed AUC, it uses the α prediction rule which enables the construction of very smooth ROC curve because several values of α normally will produce smoothness in the curve, the AUC calculated is also diagnostic test dependent since the test result depends on the test for the subjects and finally the method is not only simpler and straight forward but also it avoids the iteration procedure which is rigorous, time consuming and liable to errorneous results. Computational burden can still be substantial in binormal ROC curve as a method of calculating AUC because a number of iterative procedures that are involved in obtaining estimators, for instance MLE of AUC (Dorfmann & Alf 1969; Metz et al, 1998).

RECOMMENDATIONS

The new method of calculating AUC from predicted probability is recommended because of the reasons given above. It is highly recommended that non-parametric method of calculating AUC such as the predicted probability method be employed because of the fact that it is always distribution free and heavy computational procedures are not required. It is also advised that dummy variable regression be utilized in classifying disease conditions since it not discriminates well, it determines the contributions of the various levels of the parent independent variables.

Acknowledgement

We will not fail to appreciate Dr. Happiness Ilouno of the Department of Statistics, Nnamdi Azikiwe University Awka Nigeria for her encouragement during this research work. Her effort was wonderful.

References

- Aoki, K., J. Misumi, T. Kimura, W. Zhao and T. Xie, 1997. Evaluation of cutoff levels for screening of gastric cancer using serum pepsinogens and distributions of levels of serum pepsinogens I, II and of PG I/PG II ratios in a gastric cancer case-control study. *Journal of Epidemiology*, 7, 143 – 151.
- Bamber D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology* 12, 387-415.
- Buros, Amy and Tubbs, Jack D.(2013). Applying the Jonckheere-Terpstra Statistic to AUC Regression. Department of Statistical Science, Baylor University, Waco, TX 76706. [www.louisville.edu/sphis/bb/srcos-2013/...](http://www.louisville.edu/sphis/bb/srcos-2013/)
- DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L. (1988). Comparing the area under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44(3), 837-845.
- Dorfman DD, Alf JrE. Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals – rating-method data. *Journal of Mathematical Psychology* 1969; 6: 487-496.
- D. D. Dorfman and E. Alf, “Maximum likelihood estimation of parameters of signal detection theory: a direct solution,” *Psychometrika*, vol. 33, no. 1, pp. 117–124, 1968.
- Egan, JP. (1975). Signal detection theory and ROC Analysis. New York: Academic Press.
- Green, DM. and Swets, JA. (1966). Signal detection theory and psychophysics. John Wiley & Sons, Inc. New York.
- Greiner, M., Pfeiffer, D., and Smith, RD. (2000). Principals and practical application of the Receiver Operating Characteristic analysis for diagnostic tests. *Preventive Veterinary Medicine*, 45:23–41.
- J. A. Hanley and B. J. McNeil, “The meaning and use of the area under a receiver operating characteristic (ROC) curve,” *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.
- Hanley JA, McNeil BJ. A method of comparing the area under two ROC curves derived from the same cases. *Radiology* 1983; 148: 839-843.
- Hsiao, J. K., J. J. Barko and W. Z. Potter, 1989. Diagnosing diagnoses: receiver operating characteristic methods and psychiatry. *Archives of General Psychiatry*, 46, 664 – 667.
- Krzyśko M., Wołyński W., Górecki T., Skorzybut M. (2008) Systemy uczące się, Wydawnictwo Naukowo-Techniczne
- Lasko, T. A., J. G. Bhagwat, K. H. Zou and L. Ohno-Machado, 2005. The Use of Receiver Operating Characteristic Curves in Biomedical Informatics. *Journal of Biomedical Informatics*, 38, 404 – 415.
- L’opez-Rat’on, M., Cadarso-Su’arez, C., and Lado, MJ. (2012a). T’ecnicas de estimaci’o n e inferencia de las curvas ROC. Editorial Acad’emica Espa~ nola.
- Mann, H.B & Whitney D.R (1947).On a test whether one of two random variables is stochastically larger than the other.*Ann.Math.Statist*;18,pp.50-60.
- McClish, DK. (1989). Analyzing a portion of the ROC curve. *Medical Decision Making*, 9:190–195.

- Metz CE. Basic Principles of ROC analysis. *Seminars in Nuclear Medicine* 1978; **8**(4): 283-298.
- Metz CE. ROC methodology in radiologic imaging. *Investigative Radiology* 1986; **21**(9): 720-733.
- Metz, C. E., 1989. Some practical issues of experimental design and data analysis in radiological ROC studies. *Investigation Radiology*, **24**, 234 – 245.
- Metzger BE, Buchanan TA, Coustan DR, de Leiva A, Dunger DB, Hadden DR, et al. Summary and recommendations of the fifth international workshop-conference on gestational diabetes mellitus. *Diabetes Care* 2007; **30**: S 251.
- Metz, CE., Herman, BA., and Shen, JH. (1998). Maximum likelihood estimation of Receiver Operating Characteristic (ROC) curves from continuously-distributed data. *Statistics in Medicine*, **17**:1033–1053.
- National Diabetes Data Group. Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. *Diabetes* 1979; **28**: 1039–1057
- Ogum, G. E. O. (2003). *Introduction to methods of multivariate analysis*. Afri Towers Ltd. Aba. Nigeria.
- Okeh U.M and Oyeka I.C.A (2014) Two Factor Analysis of Variance and Dummy Variable Regression Models. *Global Journal of Science Frontier Research: F Mathematics and Decision Sciences*. Volume 14 Issue 6 Version 1.0. Online ISSN: 2249-4626 & Print ISSN: 0975-5896.
- Okeh U.M and Oyeka I.C.A (2015) One Factor Analysis of Variance and Dummy Variable Regression Models. *Global Journal of Science Frontier Research: F Mathematics and Decision Sciences*. Volume 15 Issue 7 Version 1.0. Online ISSN: 2249-4626 & Print ISSN: 0975-5896.
- Okeh UM and Oyeka ICA (2016). Dummy Variable Multiple Regression Analysis of Matched Samples. *Biometrics & Biostatistics International Journal*. Volume 3 Issue 5.
- Onyeagu, S. I. (2003). *A first Course in Multivariate analysis*. MegaConcept, Nigeria.
- Pepe, M.S. (2003). *The statistical evaluation of medical test for classification and prediction*. Oxford: Oxford University Press.
- Pepe, MS. (2004). *The statistical evaluation of medical tests for classification and prediction*. 1st ed. Oxford University Press, USA.
- P.K. Sen, On some convergence properties of U statistics, *Calcutta Statist. Assoc. Bull.* **10** (1960), pp. 1{18.
- Swets, J. A. and R. M. Picket, 1982. *Evaluation of diagnostic systems: methods from signal detection theory*. Academic Press, New York.
- WHO. Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications. Report of a WHO consultation. Part 1: Diagnosis and Classification of Diabetes Mellitus. Geneva: Department of Non-communicable Disease Surveillance, World Health Organization, 1999.
- Zou, KH., Hall, WJ., and Shapiro, DE. (1997). Smooth non-parametric Receiver Operating Characteristic (ROC) curves for continuous diagnostic tests. *Statistics in Medicine*, **16**:2143–2156.
- Zhou, X. H., N. A. Obuchowski and D. K. McClish, 2002. *Statistical methods in diagnostic medicine*. Wiley, New York.