IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 13, Issue I Ver. VI (Jan. - Feb. 2017), PP 25-36
www.iosrjournals.org

Identification Radio-Lucent and Radio-Opaque Kidney Stones

Dr.Fayyadh Abdulla Ali* Dr.TareqAzeezSalih**
** Asst. Prof.-Statistics Department –college of Administration and economics- university of Wasit-Iraq*
*** lecturer-Statistics Department –college of Administration and economics- university of Wasit-Iraq*

Abstract: *The receiver operating characteristic (ROC) curve is a graphical representation of the relationship between false positive and true positive rates. It is a widely used statistical tool for describing the accuracy of a diagnostic test and applied in measuring discriminatory ability of diagnostic or prognostic tests. This makes ROC analysis one of the most actively research areas in medical statistics. In this research we used ROC analysis in radiological tests as a method for distinction between the two types of kidney stones (radio opaque and radio lucent) . The sample of this research was 183 patients (119male and 64 female) their ages between 18-80 having kidney stones its size more than or equal (10mm). This research done between September 2014 to August 2016 in* ***Al-Emammain Al-Kadhymain*** *medical city, Baghdad, Iraq. In all patients X-Ray of Kidney,Uretare,bladder(KUB) done after preparation and then non contrast Computed Tomography(CT) was performed to all patients. two parameters were studied which are appearance of stone on KUB and Hounsfield unit for each stone was measured in Computed Tomography (CT). We classified the stones according to their appearance on KUB to Radio-Opaque stone* ***(128 stones) and Radio-Lucent stone (55 stones) .*** *By statistical analysis we found that the cut-off value of HU was* ***543 with sensitivity 97.7% and specificity 92.7% .*** *Findcut-off value of Hounsfield unit(HU) is of value in the classification ofkidney stoneaccording their appearance.*

Keywords: *Sensitivity; Specificity; ROC analysis; Medical decision making; Radiology*

I. Introduction

Receiver operating characteristics (ROC) analysis is a methodology (technique) for visualizing,organizing, evaluating, comparing and selecting classifiers on the basis of their predicting performance.First known application of ROC analysis took place during Second World War in the early 1950s for the analysis of RADAR signal detection. In the 1960s, Dr Lee Lusted was the first to recognize a possible role for ROC analysis in medical decision making .

Later its use began in the signal detection theory for illustrating the compromise between hit rates and false alarm rates of classifiers . Other fields to which ROC analysis has been introduced include psychophysics ,medicine (various medical imaging techniques for diagnostic purposes, including computed tomography, mammography, chest x-rays and magnetic resonance imaging , and also diverse methods in epidemiology) and social sciences[12].

ROCanalysishasbeenextendedforuseinvisualizingandanalyzingthebehaviorofdiagnosticsystems(Swets,1988).Themedicaldecisionmakingcommunityhasanextensiveliteratureontheuseof ROCgraphsfordiagnostictesting(Zou,2002).Swets,DawesandMonahan(2000a)
recentlybroughtROCcurvestotheattentionofthewiderpublicwiththeirScientificAmericanarticle.

Recent years have seen an increase in the use of ROC graphs in the machine learning community[5].

Kidney stones(renal calculi),are solid masses made of crystals. Kidney stones originate in your kidneys, but can be found at any point in your urinary tract. The urinary tract includes the kidneys, ureters, bladder, and urethra. [10]

Renal calculus remains to be a common problem in the hospital. It is the third most common urological problem after urinary tract infection and prostate disease[14] .

Computed tomography (CT) has a superior sensitivity and specificity over all other modalities in diagnosis ofrenal stone in determining thesize and number ofkidney stonesno matter how small it is, also it helps in the identificationof Hounsfield unit(**HU)** andthus determine thecomposition ofkidney stone.

II. Aim of Research

The aim of this research is to differentiated between radio opaque and radio lucent stone by using ROC depending on attenuation measurements Hounsfield unit(HU)and to finding cutoff value depending on Hounsfield unit in computed tomography(CT).

III. Hounsfield Scale

The **Hounsfield scale** or **CT numbers**, named after Sir Godfrey New bold Hounsfield, is a quantitative scale for describing radio density.

The Hounsfield unit (HU) scale is a linear transformation of the original linear attenuation coefficient measurement into one in which the radio density of distilled water at standard pressure and temperature (STP) is defined as zero Hounsfield units (HU), while the radio density of air at STP is defined as -1000 HU. In a voxel with average linear attenuation coefficient μ, the corresponding HU value is therefore given by:

$$HU = 1000 \times \frac{\mu - \mu_{water}}{\mu_{water} - \mu_{air}}$$

Whereμ_{water} and μ_{air} are respectively the linear attenuation coefficients of water and air.

Thus, a change of one Hounsfield unit (HU) represents a change of 0.1% of the attenuation coefficient of water since the attenuation coefficient of air is nearly zero.

It is the definition for CT scanners that are calibrated with reference to water.[6],[7]

IV. Sensitivity, Specificity, Accuracy and AUC[12],[15]

The traditional measures to quantify the diagnosticaccuracy of a test are sensitivity and specificity. Theseparameters describe the fractions of patients (diseasedand non-diseased) that are classified correctly. The sensitivityor true positive fraction (TPF) describes thefraction of diseased patients that actually has a positivetest result. The specificity or true negative fraction(TNF) describes the probability of a negative test resultin non-diseased individuals. Sensitivity and specificitydescribe the results of a test in a dichotomous way: atest result is either positive or negative[4].

The"rawdata"producedbya classificationschemeduringtestingarecountsofthecorrectandincorrectclassifications fromeachclass.Thisinformationisthen normallydisplayed ina*confusionmatrix.*Aconfusionmatrixisaformofcontingencytableshowingthedifferences between the true and predicted classes forasetoflabeledexamples,asshowninTable1.

Table 1 diagnostic test results in relation to true disease status in A 2 × 2 Table(Aconfusionmatrix)

Diagnostic test result	Disease status		Total
	Present (Positive)	Absent (Negative)	
Present (Positive)	True positive (TP)	False positive (FN)	All test positive (CP)
Absent (Negative)	False negative (FP)	True negative (TN)	All test negative (CN)
Total	Total with disease (RP)	Total without disease (RN)	Total sample size(N)

In Table 1, TP and TN are the number of true positives and true negatives respectively, FP and FN are the numbers of false positives and false negatives respectively

The row totals, CP and CN, are the number of *truly* positive and negative, and the column totals, RP and RN, are the number of *predicted* positive and negative, although the confusion matrix shows *all* of the information about the classifier's performance, more meaningful measures can be extracted from it to illustrate certain performance criteria.

The values described are used to calculate different measurements of the quality of the test. The first one is sensitivity, which is the probability of having a positive test among the patients who have a positive diagnosis. Specificity, is the probability of having a negative test among the patients who have a negative diagnosis.

Accuracy, is the proportion of the total number of predictions that were correct.

$$Sensitivity(1-\beta) = \frac{TP}{TP+FN} = \frac{TP}{CP} = P(TP) \ldots..(1)$$

$$Specificity(1-\alpha) = \frac{TN}{TN+FP} = \frac{TN}{CN} = P(TN) \ldots..(2)$$

$$Accuracy(1-Error) = \frac{TP+TN}{TP+TN+FN+FP} = \frac{TP+TN}{CP+CN} = P(C) \ldots..(3)$$

$$Positive\ predictive\ value = \frac{TP}{TP+FP} = \frac{TP}{RP} \ldots..(4)$$

$$Negative\ predictive\ value = \frac{TN}{FN+TN} = \frac{TN}{RN} \ldots..(5)$$

High sensitivity often implies low specificity and vice versa.

Sensitivity and specificity are the basic measures of the accuracy of a diagnostic test. Theydescribe the abilities of a test to enable one to correctly diagnose disease when disease isactually present and to correctly rule out disease when it is truly absent.

As suggested by above equations, sensitivity is the proportion of true positives that are correctly identified by a diagnostic test. It shows how good the test is at detecting a disease. Specificity is the proportion of the true negatives correctly identified bya diagnostic test. It suggests how good the test is at identifying normal (negative) condition.

A test can be very specific without being sensitive, or it can be very sensitive without being specific. Both factors are equally important. A good test is a one has both high sensitivity and specificity.

Accuracy is the proportion of true results, either true positive or true negative, in a population. It measures the degree of veracity of a diagnostic test on a condition.

Sensitivity [P (Tp)] can be increased with little loss in specificity [P (Tn)], or they may not. This means that the comparison of two systems can become ambiguous. Therefore, there is a need for a single measure of classifier performance [often termed accuracy, but not to be confused with P(C)] that is in variant to the decision criterion selected,priorprobabilities,andiseasilyextendedtoincludecost/benefitanalysis.Thispaperdescribestheresultsofanexperimentalstudytoinvestigatetheuseoftheareaunderthe ROCcurve(AUC)assuchas ameasureofclassifierperformance.

Several summary indices are associated with the ROC curve. One of the most popular measures is the area under the ROC curve (AUC) .AUC is a combined measureof sensitivity and specificity. AUC is a measure of the overall performance ofa diagnostic test and is interpreted as the average value of sensitivity for all possible values of specificity. It can take on any value between 0 and 1, since both the x and y axes have values ranging from 0 to 1. The closer AUC is to 1, the better the overall diagnostic performance of the test, and a test with an AUC value of 1 is one that is perfectly accurate.

The area under the ROC curve is a measure for thediagnostic accuracy of a test and is often used to makecomparisons between diagnostic tests or observers.

The total area under the ROC-curve is a measure of the performance of the diagnostic test since it reflects the test performance at all possible cut-off levels. The area lies in the interval [0.5, 1] and the larger area, the better performance. Assume that a high value from the method indicates that diagnosisis positive and a low value indicates that diagnosis is negative. The area is then a measurement of the probability that the distribution of the positive diagnosis is statistically larger than the distribution of the negative diagnosis.

There are several ways to calculate the area under a ROC curve.**First**, the trapezoidal rule can be used but gives an underestimation of the area. **Second**, it is possible to get a better approximation of the curve by fitting the data to a binormal model with maximum-likelihood estimates. After that it is possible to get a better estimate of the area. This is done, for example, in the program Rockit[13][Rockit, 2002]. A **third** way to calculate the area is to use the Mann-Whitney U statistic (also known as the non-parametric Wilcoxon statistic). That is, no assumptions on the distributions of the data are done since Wilcoxon is a distribution-free statistic [1] [Bamber, 1975, Hanley and McNeil,[8] 1982]. The program Rockit also presents a Wilcoxon area-estimation.

Equal AUCs of two tests represents similar overall performance of medical tests but this does not necessarily mean that both the curves are identical. They may cross each other.

Area under the ROC curve[8] is considered as an effective measure of inherent validity of a diagnostic test. This curve is useful in ***(i)*** finding optimal cut-off point to least misclassify diseased or non-diseased subjects, ***(ii)*** evaluating the discriminatory ability of a test to correctly pick diseased and non-diseased subjects; ***(iii)*** comparing the efficacy of two or more tests for assessing the same disease; and ***(iv)*** comparing two or more observers measuring the same test (inter-observer variability).

WhenthedecisionthresholdisvariedandanumberofpointsontheROCcurve $[P(Fp) = \alpha, P(Tp) = 1 -$ $\beta]$havebeenobtainedthesimplestwaytocalculatetheareaundertheROCcurveistousetrapezoidal integration,

$$AUC = \sum_i \left\{(1 - \beta_i . \Delta\alpha) + \frac{1}{2}[\Delta(1 - \beta) . \Delta\alpha]\right\} \dots \dots . . (6)$$

Where $\Delta(1 - \beta) = (1 - \beta_i) - (1 - \beta_{i-1}) \dots \dots . . (7)$

$\Delta\alpha = \alpha_i - \alpha_{i-1} \dots \dots . . (8)$

Itisalsopossibletocalculate theAUCbyassumingthat theunderlyingprobabilitiesofpredictingnegativeorpositiveareGaussian. TheROC curve will then haveanexponentialformandcanbefittedeither: **directly**usinganiterativeMaximumLikelihood(ML)estimation givingthedifferencein meansandtheratioofthevariancesofthe positiveandnegativedistributions;**or,**iftheROCcurveisplottedondoubleprobabilitypaper,astraightlinecanbefittedtothepointsontheROCcurve.The slopeandinterceptof this fitted line arethenusedtoobtainanestimateoftheAUC.

The trapezoidal approachsystematicallyunderestimatestheAUC.ThisisbecauseofthewayallofthepointsontheROCcurveareconnectedwith straightlinesratherthansmoothconcavecurves.However,providingthereareareasonablenumber ofpointsontheROCcurvetheunderestimationoftheareashouldnotbetoosevere.Thetrapezoidalapproachalsodoes notrelyonanyassumptionsastotheunderlyingdistributionsofthepositiveandnegativeexamplesareexactly the same quantity measured usingtheWilcoxontestofranks[12]

TheStandardErroroftheAUC(SE$(\hat{\theta})$)isofimportanceifwewishtotestthesignificanceofoneclassificationschemeproducingahigherAUCthananother.Conventionallytherehavebeenthreewaysofcalculatingthisvariability associatedwiththeAUC:

1. From the confidence intervals associated with themaximum likelihoodestimateofAUC, $(\hat{\theta})$;
2. From the standard error of the Wilcoxon statistic,SE(W);and

3 FromanapproximationtotheWilcoxon statisticthatassumes that the underlying positive and negative distributions are exponential in type.

V. Receiver Operating Characteristics (Roc) Analysis[6]

Receiver operating characteristic (ROC) curve is the plot that depicts the trade-off between the sensitivity and (1-specificity) across a series of cut-off points when the diagnostic test is continuous or on ordinal scale (minimum 5 categories). This is an effective method for assessing the performance of a diagnostic test.

The terms “ROC curve”, “ROC graph” and “ROC analysis” are sometimes used interchangeably, though the “ROC analysis” is the most general.

An ROC graph for original two-class problems is defined as a two-dimensional plot which represents TPR (sensitivity) on y-axis in dependence of FPR (= 1-specificity) on x-axis.

An ROC curve is a curve on an ROC graph with start point in (0,0) and end point in (1,1). Drawing procedure for this curve depends on the type of classifiers we want to evaluate.

An example of an ROC graph with four different ROC curves each representing one classifier is given in Fig. 1. Classifier A is by far better than the other three classifiers. ROC curves of classifiers B and C cross – each of these two is superior to the other for some deployment contexts (i.e. combinations of class distribution and misclassification costs). Classifier D is of no use as its performance is no better than chance.

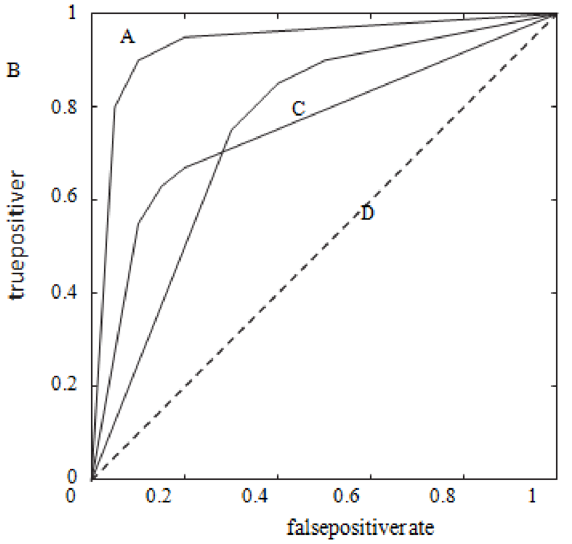


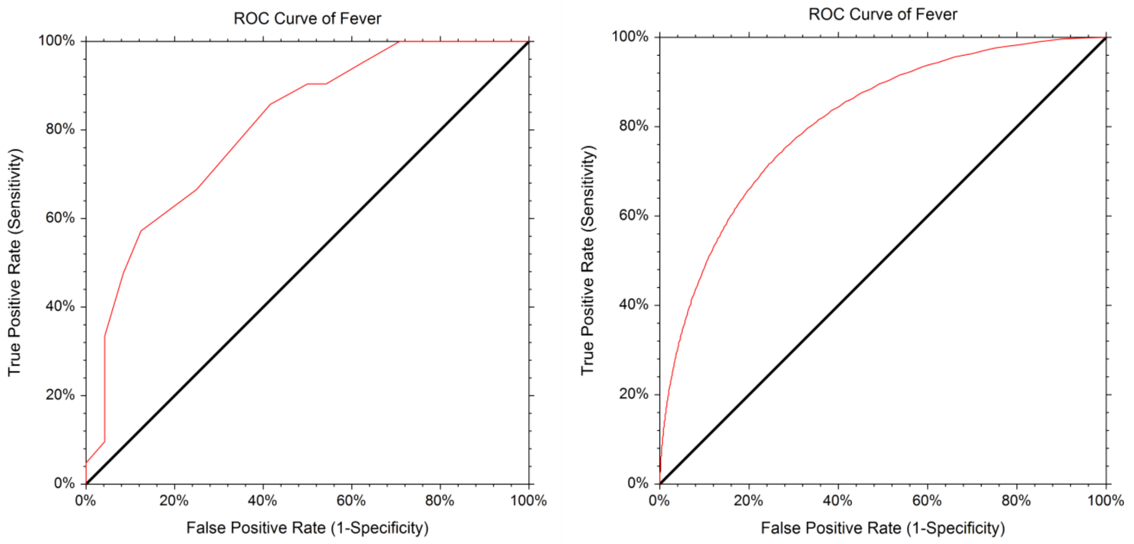
Fig.1.AnROCgraphwithfourROCcurves.

A diagnostic test yields a measurement (criterion value) that is used to diagnose some condition of interest such as a disease. (In the sequel, we will often call the ‘condition of interest’ the ‘disease.’) The measurement might be a rating along a discrete scale or a value along a continuous scale.

A positive or negative diagnosis is made by comparing the measurement to a cutoff value. If the measurement is less (or greater as the case may be) than the cutoff, the test is negative. Otherwise, the test is positive. Thus the cutoff value helps determine the rates of false positives and false negatives.

A receiver operating characteristic (ROC) curve shows the characteristics of a diagnostic test by graphing the false-positive rate (1-specificity) on the horizontal axis and the true-positive rate (sensitivity) on the vertical axis for various cutoff values.

Each point on the ROC curve represents a different cutoff value**.**



An empirical ROC curve A binormal ROC curve

Fig.2: ROC Space

Cutoff values that result in low false-positiverates tend to result low true-positive rates as well. As the true-positive rate increases, so does the false positiverate. Obviously, a useful diagnostic test should have a cutoff value at which the true-positive rate is high and thefalse-positive rate is low.

Several methods have been proposed to generate ROC curves. These include the binormal and the empirical(nonparametric) methods.

Likelihood Ratio

The likelihood ratio statistic measures the value of the test for increasing certainty about a positive diagnosis. It is calculated as follows:

$$\mathrm{LR} = \frac{\mathrm{Pr(posotivetest|disease)}}{\mathrm{Pr(posotivetest|No\ disease)}} = \frac{\mathrm{Sensitivity}}{1 - \mathrm{Specificity}}$$

ROC curve is defined as a plot of test sensitivity(TPR) as the y coordinate versus its1-specificity or false positive rate (FPR) as the x coordinate forvarying cut-off points of test values. For a given diagnostic test, the true positive rate (TPR) against false positive rate (FPR) can be measured, where

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP + FN}} \dots \dots ..(9)$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{TN + FP}} \dots \dots \dots .(10)$$

Is an effective method of evaluating the quality or performance of diagnostic tests, and is widely used in radiology to evaluate the performance of many radiological test

As we can see from theaboveequations,TPRisequivalenttosensitivityandFPRisequivalentto(1–specificity).AllpossiblecombinationsofTPRandFPRcomposeaROCspace. The receiver operating characteristic (ROC) curve[11] is the plot that displays the full picture of trade-off between the sensitivity and (1- specificity) across a series of cutoff points.OneTPRandoneFPRtogetherdetermineasingle pointintheROCspace,andthepositionofapointintheROCspaceshowsthetradeoffbetweensensitivityand specificity,i.e.theincreaseinsensitivityisaccompaniedbyadecreaseinspecificity.Thusthelocationofthepointin theROCspacedepictswhetherthediagnosticclassificationisgoodornot.Inanidealsituation,apointdetermined bybothTPRandFPFyieldsacoordinates(0,1),orwecansaythatthispointfallsontheupperleftcorneroftheROCspace.Thisideapointindicatesthediagnostictesthasasensitivityof100%andspecificityof100%.Itisalso calledperfectclassification.Diagnostictestwith50%sensitivityand50%specificitycanbevisualizedonthediagonal determinedbycoordinate(0,0)andcoordinates(1,0).Theoretically,arandomguesswouldgiveapointalongthis diagonal.Apointpredictedbyadiagnostictestfallintotheareaabovethediagonalrepresentsagooddiagnosticclassification,otherwiseabadprediction.Agraphicpresentationofwhatdescribedaboveisshowninfigure1.

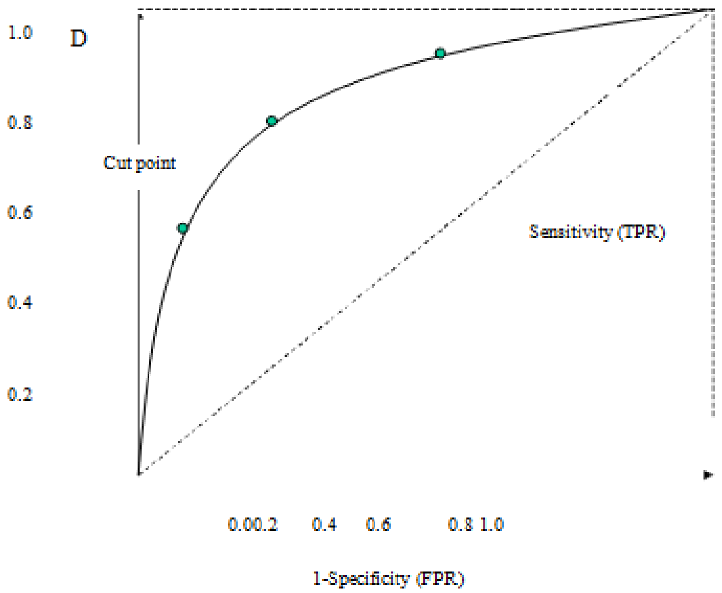


Fig.3: ROC Space

Several points in ROC space are important to note. The lower left point (0, 0) represents the strategy of never issuing a positive classification, such a classifier commits no false positive errors but also gains no true positives. The opposite strategy, of unconditionally issuing positive classifications, is represented by the upper right point (1, 1).The point (0, 1) represents perfect classification. D's performance is perfect as shown. Asinglecut-pointofadiagnostictestdefinesonesinglepointintheROCspace;however,differentpossiblecut-pointsofadiagnostictestdetermineacurveinROCspace,whichisalsocalledROCcurve.Likeasinglepointinthe ROC space, ROC curve is often plotted by using true positive rate (TPR) against false positive rate (FPR) for different cut-points of a diagnostic test, starting from coordinate (0, 0) and ending at coordinate (1, 1). FPR (1 – specificity) is represented by x-axis and TPR (sensitivity) is represented by y-axis. Thus, ROC curve is a plot of a test’s sensitivity vs. (1-specificity) as well. The interpretation of ROC curve is similar to a single point in the ROC space, the closer the point on the ROC curve to the ideal coordinate, the more accurate the test is. The closer the points on the ROC curve to the diagonal, the less accurate the test is. In addition, (1) the faster the curve approach the ideal point, the more useful the test results are; (2) the slope of the tangent line to a cut-point tells us the ratio of the probability of identifying true positive over true negative, i.e. likelihood ratio (LR) for the test value: LR = sensitivity/(1-specificity), if the ratio is equal to 1, the selected cut-point doesn’t add additional information to identify true positive result. If the ratio is greater than 1, the selected cut-point help identify true positive result. If the ratio is less than 1, it decreasesdiseaselikelihood(3)theareaunderROCcurve(AUC)providesawaytomeasuretheaccuracyofa diagnostictest.Thelargerarea,themoreaccuratethediagnostictestis.AUCofROCcurvecanbe measured by the followingequation,Wheret=(1–specificity)andROC(t)issensitivity.

$$AUC = \int_{0}^{1} ROC(t)dt$$

CommonlyusedclassificationusingAUCforadiagnostictestissummarizedintable2:

Table2 accuracy classification by AUC for a diagnostic test

AUC range	Classification
0.9<AUC<1.0	Excellent
0.8<AUC<0.9	Good
0.7<AUC<0.8	worthless
0.6<AUC<0.7	Not good

Inshort, ROC curveisagoodtooltoselectpossibleoptimalcut-pointforagivendiagnostictest.

VI. Area under an ROC Curve (AUC)

Let Y denote a random variable representing a continuous diagnostic test result. The diagnosis according to any cutoff value c is positive if $Y \geq c$and negative 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 cutoff value c, TP(c), and FP(c) are

$TP(c) = P(Y \geq c|D_1)$

$FP(c) = P(Y \geq c|D_0)$

The ROC curve is denoted by

$ROC(t) = 1 - F_1(F_0^{-1}(1-t))$

where $TP(c) = F_1(c), FP(c) = F_0(c)$, and $t$is the all possible FP rates according to the varying c values in $(-\infty, \infty)$

The area under an ROC curve (AUC) is a popular measure of the accuracy of a diagnostic test. Other things beingequal, the larger the AUC, the better the test is a predicted the existence of the disease. The possible values ofAUC range from 0.5 (no diagnostic ability) to 1.0 (perfect diagnostic ability).

A statistical test of usefulness of a diagnostic test is to compare it to the value 0.5. Such a statistical test can be made if we are willing to assume that the sample is large enough so that the estimated AUC follows the normal distribution. The statistical test is

$$z = \frac{\hat{A} - 0.5}{\sqrt{var(\hat{A})}} \dots\dots\dots..(11)$$

Where$\hat{A}$is the estimated AUC and var($\hat{A}$)is the estimated variance of$\hat{A}$.

Two methods are commonly used to estimate the AUC. The first is the *binormal*method presented by Metz (1978) and McClish (1989). This method results in a smooth ROC curve from which both the complete and partial AUC may be calculated. The second method is the empirical (nonparametric) method by DeLong et al (1988). This method has become popular because it does not make the strong normality assumptions that the binormal method makes. The above z test may be used for both methods, as long as an appropriate estimate of var($\hat{A}$) is used .

6.1. The AUC of a Single Binormal ROC Curve [9]

The formulas that we use here come from McClish (1989). Suppose there are two populations, one made up of individuals with the disease and the other made up of individuals without the disease. Further suppose that the value of a criterion variable is available for all individuals. Let Y_1refer to the value of the criterion variable in the diseased population and Y_0refer to the value of the criterion variable in the non diseased population. The binormal model assumes that both $Y_1 and$ $Y_0$are normally distributed with different means and variances. That is,

$\mathrm{Y}_1 \sim \mathrm{N}(\mu_1, \sigma_1^2), \mathrm{Y}_0 \sim \mathrm{N}(\mu_0, \sigma_0^2)$

The ROC curve is traced out by the function

$$FP(c) = \Phi\left(\frac{\mu_1 - c}{\sigma_1}\right) \quad , TP(c) = \Phi\left(\frac{\mu_0 - c}{\sigma_0}\right) \quad -\infty < c < \infty$$

Where $\Phi(z)$is the cumulative normal distribution function.

The area under the whole ROC curve is

$$A = \int_{-\infty}^{\infty} TP(c)F\acute{P}(c)dc$$

$$A = \int_{-\infty}^{\infty} \left[\Phi\left(\frac{\mu_1 - c}{\sigma_1}\right)\phi\left(\frac{\mu_0 - c}{\sigma_{0x}}\right)\right]dc$$

$$A = \Phi\left[\frac{a}{\sqrt{1+b^2}}\right] \dots\dots\dots.(12)$$

Where

$$a = \frac{\mu_1 - \mu_0}{\sigma_1} \; , b = \frac{\sigma_0}{\sigma_1} \quad \dots\dots\dots(13)$$

The area under a portion of the AUC curve is given by

$$A = \int_{c1}^{c2} TP(c)F\acute{P}(c)dc$$

$$A = \frac{1}{\sigma_0}\int_{c2}^{c1} \left[\Phi\left(\frac{\mu_1 - c}{\sigma_1}\right)\phi\left(\frac{\mu_0 - c}{\sigma_0}\right)\right]dc$$

The partial area under an ROC curve is usually defined in terms of a range of false-positive rates rather than the criterion limits c_1 and c_2 . However, the one-to-one relationship between these two quantities, given by

$$\hat{A} = \Phi\left[\frac{\hat{a}}{\sqrt{1+\hat{b}^2}}\right] \ldots\ldots..(14)$$

Note that for ease of reading we will often omit the use of the *hat* to indicate an MLE in the sequel.

The variance of $\hat{A}$is derived using the method of differentials as

$$Var(\hat{A}) = \left(\frac{\partial A}{\partial \Delta}\right)^2 var(\hat{\Delta}) + \left(\frac{\partial A}{\partial \sigma_0^2}\right)^2 var\left(s_0^2\ \right) + \left(\frac{\partial A}{\partial \sigma_1^2}\right)^2 var\left(s_1^2\ \right) \ldots\ldots\ldots\ldots.(15)$$

$$\frac{\partial A}{\partial \Delta} = \frac{E}{\sqrt{2\pi(1+b^2)\sigma_1^2}}[\Phi(\widetilde{c_1}) - \Phi(\widetilde{c_0})] \ldots\ldots\ldots(16)$$

$$\frac{\partial A}{\partial \sigma_0^2} = \frac{E}{4\pi(1+b^2)\sigma_0\sigma_1}[\mathrm{e}^{-k0} - \mathrm{e}^{-k1}] - \frac{abE}{2\sigma_0\sigma_1\sqrt{2\pi(1+b^2)^{3/2}}}[\Phi(\widetilde{c_1}) - \Phi(\widetilde{c_0})] \ldots\ldots(17)$$

$$E = exp\left[-\frac{a^2}{2(1+b^2}\right] \ldots\ldots\ldots.(18)$$

$$\Delta = \mu_1 - \mu_0$$

$$\frac{\partial A}{\partial \sigma_1^2} = -\frac{a}{2\sigma_1}\left(\frac{\partial A}{\partial \Delta}\right) - b^2\left(\frac{\partial A}{\partial \sigma_0^2}\right) \ldots\ldots.(19)$$

$$\widetilde{c_i} = \left[\Phi^{-1}(FP_i) + \frac{ab}{1+b^2}\right]\sqrt{1+b^2} \ldots\ldots(20)$$

$$k_i = \frac{\widetilde{c_i}^2}{2}$$

$$Var(\hat{\Delta}) = \frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1} \ldots\ldots.(21)$$

$$Var(s_0^2) = \frac{2\sigma_0^4}{n_0 - 1} \ldots\ldots(22)$$

$$Var(s_1^2) = \frac{2\sigma_1^4}{n_1 - 1} \ldots\ldots(23)$$

where n_1 and n_0 are the numbers of diseased and nondiseased study subjects,respectively.

Using the following transformation which results in statistics that are closer to normality and ensures confidence limits that are outside the zero-one range. The transformation is

$$\widehat{\Psi} = \frac{1}{2}ln\left(\frac{1+A}{1-A}\right) \ldots..(24)$$

The variance of $\widehat{\Psi}$is estimated using

$$var\left(\widehat{\Psi}\right) = \frac{4}{\left(1-\hat{A}^2\right)^2} var(\hat{A}) \ldots\ldots(25)$$

An $100(1-\alpha)\%$ confidence interval for $\widehat{\Psi}$may then be constructed as

$$L, U = \widehat{\Psi} \mp z_{1-\alpha/2}\sqrt{var\left(\widehat{\Psi}\right)} \ldots..(26)$$

Using the inverse transformation, the confidence interval for A is given by the two limits

$$\frac{1-e^{-L}}{1+e^{-L}}\ and\ \frac{1-e^{-U}}{1+e^{-U}} \ldots..(27)$$

6.2. The AUC of a Single Empirical ROC Curve

The empirical (nonparametric) method by DeLong et al (1988) is a popular method for computing the AUC. This method has become popular because it does not make the strong normality assumptions that the binormal method makes. The formula for computing this estimate of the AUC and its variance are given later in the section on comparing two empirical ROC curves.

Nonparametric ROC. In our study, the empirical method was used for the nonparametric ROC analysis. Thismethod is popular because it does not make any distributional assumptions about the diagnostic test measurements[3].

In this approach, the possible diagnostic test results for each cutoff value c are considered, and the corresponding true and false positive rates are calculated by

$$TP(c) = \frac{s_1(c)}{n_1} \ldots\ldots.(28)$$

$$FP(c) = \frac{s_0(c)}{n_0} \ldots\ldots(29)$$

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 . The *AUC* of the nonparametric ROC curve is obtained using trapezoidal rule and is estimated by

$$\hat{A} = \frac{1}{n_1 n_0} \sum_{i=1}^{n_1} \sum_{j=1}^{n_0} \Psi(Y_{i1} Y_{j0}) \ldots..(30)$$

$$\Psi(Y_{i1} Y_{j0}) = \begin{cases} 1 & if\ Y_{i1} > Y_{io} \\ \frac{1}{2} & if\ Y_{i1} \ = Y_{io} \ \ldots \ldots..(31) \\ 0 & if\ Y_{i1} \ < Y_{io} \end{cases}$$

and $Y_{i1}$and $Y_{j0}$are the diagnostic test results for the diseased and non diseased individuals, respectively. The variance of the estimated *AUC* is computed using Mann-Whitney Statistic :

$$var(A) = \frac{\hat{A}(1-\hat{A}) + (n_1 - 1)(Q_1 - \hat{A}^2) + (n_0 - 1)(Q_2 - \hat{A}^2)}{n_1 n_0} \ldots..(32)$$

$Q_1$and $Q_2$are defined as

$$Q_1 = \frac{1}{n_0 {n_1}^2} \sum_Y n_0^{=y} \times \left[\left(n_1^{>y}\right)^2 + n_1^{>y} \times n_1^{=y} + \frac{\left(n_1^{=y}\right)^2}{3} \right] \ldots \ldots (33)$$

$$Q_2 = \frac{1}{{n_0}^2 n_1} \sum_Y n_1^{=y} \times \left[\left(n_0^{<y}\right)^2 + n_0^{<y} \times n_0^{=y} + \frac{\left(n_0^{=y}\right)^2}{3} \right] \ldots.(34)$$

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*

An ROC curve is a two-dimensional depiction of classifier performance. To compare classifiers we may wantto reduce ROC performance to a single scalar value representing expected performance. A common method is tocalculate the area under the ROC curve, abbreviated AUC (Bradley, 1997; Hanley & McNeil, 1982). Since the AUCis a portion of the area of the unit square, its value will always be between 0 and 1.0. However, because randomguessing produces the diagonal line between (0, 0) and (1, 1), which has an area of 0.5, no realistic classifier shouldhave an AUC less than 0.5.

The AUC has an important statistical property: the AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance. This is equivalent to the Wilcoxon test of ranks (Hanley & McNeil, 1982). The AUC is also closely related to the Gini index (Breiman, Friedman, Olshen, & Stone, 1984), which is twice the area between the diagonal and the ROC curve.

Hand and Till (2001) point out that $Gini + 1 = 2 \times AUC.$

VII. Application

7.1. Study Sample

The study was conducted in the department of diagnostic radiology in Al-Emamain Al-Kadhmain medical city , Baghdad , Iraq. 183 patients (119 males and 64females) all the patient included in this study are send from urology department and they have renal stones equal or more than 10mm in diameter as shown by U/S examination.

7.2. Exclusion Criteria

1. Patient age (less than**18** and more than **80** years).
2. Stone size (more than or equal **10** mm).
3. pregnancy
4. Patient with Barium contrast study within **3**days before examination .

7.3. Data collection

Full clinical history was taken from all patients included in this study .All the patients were submitted to the plain X-Ray of the abdomen(KUB) and non enhanced CT of the abdomen

7.4. Plain X-Ray of the abdomen(KUB-(kidney, ureters, bladder))

KUB was performed to all patients after preparation which is:

1. Before 1 day from examination the patient was asked to take early dinner and 1-2 spoons of castor oil to avoid gasses shadow in the KUB X-Ray.
2. The examination is performed supine position and ask the patient to take a deep breath .The X-Ray field extend from pubic symphysis to the superior aspect of kidney .the exposure factor was - (80-100 kv,60MAs) , and this exposure factor according to the patient built.

The following parameters were studied by KUB examination :
1. Size of stone .
2. Appearance of stone (radio- lucent or radio -opaque) .

The study focused on the renal stone that measures 10mm or more.

7.5. statistical analysis

Statistical analysis was carried out using the Statistical Package for Social Sciences (SPSS for Windows, version 17.0). Demographic data of the cases included in the study were collated and graphed. Confidence intervals with receiver operating characteristic (ROC) curve was constructed to determine the best cut-off value for determining what Hounsfield value at which a calculus can be classified as a Radio-opaque or Radio-Lucent by CT . Finally, the sensitivity, specificity, positive predictive value,negative predictive value were obtained . All data analyses were conducted at 0.05 level of significance or at 95% confidence interval

This study included 183 patients (**119** male ;**64**famale)the age range between **18-80** years ,the Radio-Opaque stones in KUB was **128** and the Radio-Lucent stones was **55 .**

The mean of Radio-Opaque stones was 902.05 whereas the mean of Radio-Lucent stones was460.47.The standard deviation of radio-opaque stones was (208.043)while The standard deviation of radio-Lucent stones was (141.517) respectively .

The minimum value of (HU) for Radio-Opaque Stones and Radio-Lucent stones was (397,307) respectively whereas The maximum value of (HU) for Radio-Opaque Stones and Radio-Lucent stones was (1437,930) respectively , as shown in table(3) below:

Table(3) Descriptives

	type of stone			Statistic	Std. Error
value of HU	radio-lucent	Mean		460.47	19.082
		95% Confidence Interval for Mean	Lower Bound	422.22	
			Upper Bound	498.73	
		5% Trimmed Mean		443.17	
		Median		430.00	
		Variance		2.003E4	
		Std. Deviation		141.517	
		Minimum		307	
		Maximum		930	
		Range		623	
		Interquartile Range		116	
		Skewness		2.388	.322
		Kurtosis		5.866	.634
	radio-opaque	Mean		902.05	18.389
		95% Confidence Interval for Mean	Lower Bound	865.66	
			Upper Bound	938.43	
		5% Trimmed Mean		897.19	
		Median		887.00	
		Variance		4.328E4	
		Std. Deviation		208.043	
		Minimum		397	
		Maximum		1407	
		Range		1010	
		Interquartile Range		200	
		Skewness		.483	.214
		Kurtosis		.374	.425

In order to find optimal HU cut-off value we use receiver operatingcharacteristic (**ROC**) we find the cut point is **543** HU with sensitivity of **97.3%** and specificity of **92.7%.** The negative predictive value was **94.4%** while positive predictive value was **96.9%.**and the accuracy value was 96.2%,The area under the curve was **0.944**.

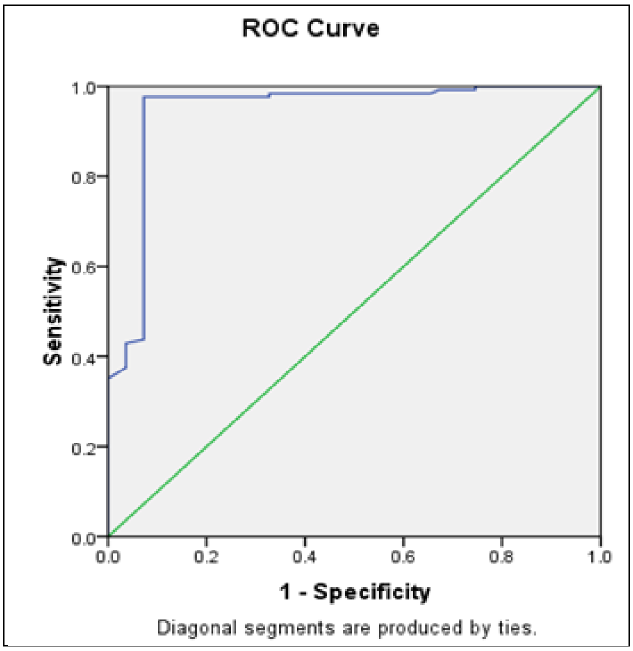


Table (4) Area Under the Curve				
Test Result Variable(s):value of HU				
Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.944	.022	.000	.901	.988

The test result variable(s): value of HU has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.
a. Under the nonparametric assumption
b. Null hypothesis: true area = 0.5

7.6.Conclusions

- ROCgraphsareaveryusefultoolforvisualizingandevaluatingclassifiers in medical research as a graphical display for the relationship between sensitivity and specificity for a continuous test variable.
- A ROC graph gives a global view on the test performance, but can not be applied directly to clinical practice. For that purpose a cut-off value of the test variable should be chosen, a choice that depends on the disease to be diagnosed and the consequences of false positive and false negative test results.
- Based on the constructed ROC curve, a threshold value of 543 in CT was established as a cut-off value in determining whether a calculus is a Radio-Opaque Or Radio-Lucent .

References

[1] Bamber, D. The Area above the Ordinal Dominance Graph and the Area below the Receiver Operating Characteristic Graph. J. Math Psychol, 12:387-415,1975

[2] BRADLEY ,ANDREWP., THEUSEOFTHEAREAUNDERTHEROC CURVEINTHE EVALUATIONOFMACHINELEARNINGALGORITHMS, *PatternRecognition*,Vol.30,No.7,pp.1145-1159,1997

[3] Colak, Ertugrul et al.,Comparison of Semiparametric, Parametric, and Nonparametric ROC Analysis for Continuous Diagnostic Tests Using a Simulation Study and Acute Coronary Syndrome Data, Computational and Mathematical Methods in Medicine Volume 2012, Article ID 698320, 7 pages

[4] Erkel , Arian R. van , Peter M. Th. Pattynama, Receiver operating characteristic (ROC) analysis: Basic principles and applications in radiology, European Journal of Radiology 27 (1998) 88–94

[5] Fawcett, Tom, ROC Graphs: Notes and Practical Considerations for Data Mining Researchers ,2003 .

[6] Feeman, Timothy G. "Mathematics of Medical Image" ,April 21,2011 www.fpnotebook.com/read/CT/hounsfield unit .html

[7] G,Motley,DalrympleN,KeeslingC,Fischer J and Harmon W,Hounsfield unit density in the determination of urinary stone composition.Urlogy,2001:170-3

[8] Hanley, J. A. and McNeil, B. J. (1982). The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. Radiology, 143(1):29-36.

[9] https://www.coursehero.com/file/13690591/ROC-Curvespdf/

[10] http://www.healthline.com , Written by health line Editorial team/puplished on march 21 2013 , kidney stones: types, testing and treatments

[11] Kumar, Rajeev and AbhayaIndrayan,receiver operating characteristic (roc) curve for medical researchers , Indian Pediatrics , Volume 48-April 17, 2011 .

[12] Majnik, Matjaz and ZoranBosnic , ROCanalysisofclassifiersinmachine learning:Asurvey , Intelligent Data Analysis 17 (2013) 531–55 .

[13] Rockit,ROC analysis, http://www-radiology.uchicago.edu/krl/toppage11.htm, 2002.

[14] Starwolf,J,Jr,MD,Facs,Nephrolithiasis,April 8,2014,http://emidicine.medscape.com/article /437096-overview.

[15] Zhu ,Wen , Nancy Zeng , Ning Wang , Sensitivity,Specificity,Accuracy,AssociatedConfidenceIntervalandROCAnalysiswithPracticalSAS Implementations,2010.