International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Statistical Study to Classify β-Thalassaemia diseases in Erbil City at Thalassaemia Center by Using ROC Curve Analysis

Aras Jalal Mhamad Karim

Sulaymaniyah University, School of Administration & Economics,
Statistics Department, Tasluja–Rapareen Street 1, zone 501, 46001, Kurdistan Region – Iraq

Abstract: ***Receiver operating characteristics (ROC) curve analysis is accepted as the most commonly used statistical analysis technique in the medical field to determine a cut-off value for a clinical test and is useful for organizing classifiers. In this paper, an ROC Curve is proposed to classify β-thalassaemia diseases in Erbil City, Iraq as the research goal, in the classification of diseases, there are, four possible outcomes. If the diseased patient is positive, the case is classified as positive, that means it is counted as a true positive; but if it is classified as negative, that means it counts as a false negative. Should the diseased patient be negative and it is classified as negative, it means that it is country as a true negative; if it is classified as positive, it counts as a false positive. This analysis is applied specifically to a case study in Erbil in Kurdistan Region of Iraq. Results for statistical analysis show that the test (Red_Blood_Cell_Count) has a greater area under curve than the other diagnostic tests after comparison.***

Keywords: ROC Curve Analysis, Classification Tool, Medical Test Distinguish, Likelihood Ratio, Criterion (Thresholds) in optimality.

1. Introduction

ROC analysis divides performance into true and false positive rates. Different ROC profiles are more or less desirable under different class distributions and different error cost functions [16]. A receiver operating characteristics (ROC) graph is a method to imagine, arrange and choose classifiers on the basis of their performance. ROC curve has been used for a long time in signal detection theory to indicate the tradeoff between hit rates and false alarm rates of classifiers. ROC analysis has been also used to visualize and analyze the behavior of diagnostic systems. The community of medical decision makers has produced ample literature on the use of ROC graphs for diagnostic testing [6].

2. Materials and Methods

2.1 Explanation of model classification (classifiers)

The classification model (classifiers) depends on the creation which is the number of cases that have desirable property or those that have undesirable property, that have been classified correctly or wrongly [13]. In the classification of diseases, there are, four possible outcomes. If the diseased patient is positive, the case is classified as positive, that means it is counted as a true positive; but if it is classified as negative, that means it counts as a false negative. Should the diseased patient be negative and it is classified as negative, it means that it is counts as a true negative; if it is classified as positive, it counts as a false positive [6]. Classification tool can be used for the analysis of several statistics as seen by [4], [7].

1- Sensitivity (tp rate):

$$SE = \text{tp rate} = \frac{TP}{P} \quad (1)$$

2- Specificity:

$$SP = \frac{TN}{FN + TN} \quad (2)$$

3- fp rate (1- specificity):

$$\text{fp rate} = \frac{FP}{N} \quad (3)$$

4- The proportion of correct classification:

$$\text{Correct Classification} = \frac{EF}{Total} = \frac{TP + TN}{P + N} \quad (4)$$

2.2 ROC Curve

The ROC curves have two dimensional graphs that are visual depictions of the performance and performance trade-off of a classification model [6], ROC curves began as tools within the theory of communication to provide visual determination of optimal operating points in signal discriminators [8]. The ROC curve is a graph of TPF versus FPF, which are both independent of disease prevalence. Basically, a traditional ROC curve defines the possible compromises between TPF and FPF - thus among the relative frequencies of positive is true, positive is false, negative is true, and false negative decisions – due to the variability of decision thresholds [11].

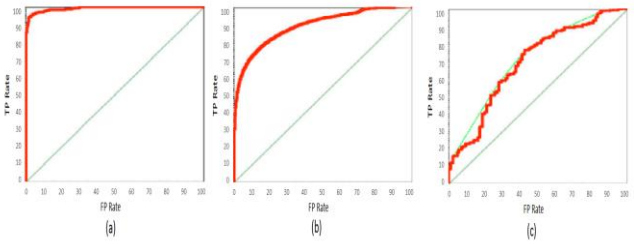

Figure 1: ROC curves: (a) an almost perfect classifier (b) a reasonable classifier (c) a poor classifier.

Additionally, Fig.1 presents some normal cases of ROC curves. Part (a) shows the ROC curve of a nearly perfect

classifier with \the performance curve nearly touching the 'perfect performance' point in the top left corner. Part (b) and part (c) show ROC curves of subordinate classifiers, and this level provides a simple visual representation of how various models perform making it easy to identify optimal versus sub-optimal models [8].

2.2.1 ROC Space and Important point on ROC Curve

A ROC graph shows the relative tradeoffs between true positives and false positives. Fig. 2 depicts an ROC graph with six classifiers ranging from X_1 to X_6. There are, many significant points in ROC space to be noted. The lower left point (0, 0) that is indicated by X_1 and always represents the strategy of negative classification; such as a classifier makes no false positive errors but also does not benefit from any true positives. The opposite strategy is a positive classification, represented by the upper right point (1,1) denoted by X_5 in Figure 2. The point (0,1) denotes perfect performance. This means that X_4 is perfect classification as indicated in Figure 2. Informally one point in ROC space is superior to another when the point is closer to the top left corner (northwest), i.e increasing tp rate, decreasing fp rate [6], [8]. The diagonal line from the bottom left corner to the top right corner denotes random classifier performance [8]. Where classifiers near the left-hand side of an ROC graph, may be thought of as ''conservative'': they have low true positive rates and they produce a few false positive errors. Classifiers on the upper right-hand side of an ROC graph may be deemed as ''liberal'': they group most positives appropriately, but they tend toward high false positive rates. In Fig. 2, X_3 is more conservative than X_2. It is possible to rank classifiers mapped onto a ROC graph based on their distance to the 'perfect performance' point (X_4). In Fig. 2, classifier X_3 can be considered to be better than a hypothetical classifier X_2 as X_3 is nearer the top left corner [6], [8]. The random classifier will generate an ROC point that ''slides'' back and forth on the diagonal depending on its frequency in guessing the positive class. For example, if a classifier randomly guesses the positive half the time class, that means it can be expected to be correct with half the positives and half the negatives; and yield the point (0.5,0.5) in ROC space. If the correct rate is 90% of the time in guessing the positive class then the false positive rate will also increase to 90%, yielding (0.9,0.9) in ROC space. In Fig. 2, X_6 performs much worse than random. The classifier on the diagonal has no any information about the class. On the other hand, a classifier below the diagonal can be said to possess beneficial knowledge, but its application of the knowledge or information is incorrect [6].

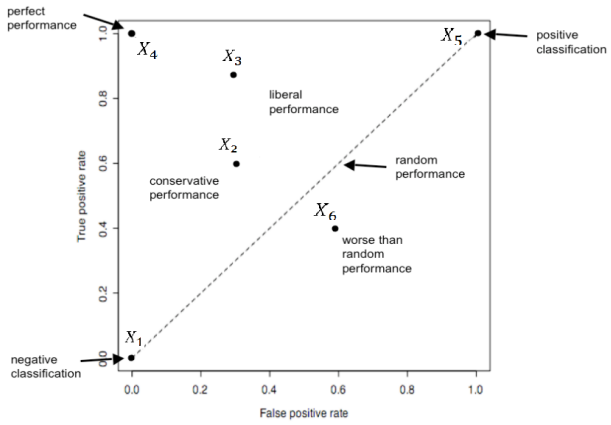

Figure 2: Shows six classifier in ROC curves

2.3 Likelihood Ratio

The likelihood ratio (LR) is described as the ratio between the probability of a defined test result with the presence of a disease and the probability of the same test result with the absence of a disease [3]:

$$LR = \frac{a}{b} \quad (5)$$

Where: a is (probability of a test result among the diseased persons), and b is (probability of the same test result among the non-diseased persons). The likelihood ratio is beneficial in clinical decision-making as it is able to provide the ratio of odds of disease among persons with a given test result (post-test) to the odds of disease among all persons (pre-test). Should a test generate results on a continuous scale, then it is [possible that a likelihood ratio can in theory be defined for each test value (x) [3]:

$$LR_{(x)} = \frac{c}{d} \quad (6)$$

Where: c is (probability of a test result x among the diseased persons), and b is (probability of the same test result x among the non-diseased persons). In theory it is possible for the likelihood ratio to be estimated by employing an infinitesimally small interval for the test result on the depend on two criteria - for example to test value of 100, we need two criteria (100.1 and 99.99). In practical terms, the likelihood ratio for a single test value is not easy to estimate, except when the an researcher is dealing with an exceptionally large sample. Thus, likelihood ratios are more frequently computed for test results on one side of a particular criterion (dichotomous test). If a test produces dichotomous results (i.e., positive or negative, by using a specific criterion for positivity), then two likelihood ratios can be defined - i.e., the likelihood ratio for a positive test (LR+) and the likelihood ratio for a negative test (LR -) [3]:

$$LR + = \frac{e}{f} \quad (7)$$

Where: e is (probability of a positive test among the diseased persons), and b is (probability of a positive test among the non-diseased persons).

$$LR - = \frac{g}{h} \quad (8)$$

Where: g is (probability of a negative test among the diseased persons), and b is (probability of a negative test among the non-diseased persons).

2.4 Area under the ROC Curve

The total area under the ROC curve is a, indication of how the diagnostic test performs as it is a reflection of the test performance at all possible cut-off levels. The assumption is that a high value from the method is indication of a positive diagnosis while a low value points to a negative diagnosis. 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 [17]. In the comparison of classifiers there may be a wish to reduce ROC performance to a single scalar value that represents expected performance. The traditional approach is to compute the area under the ROC curve (AUC) [2]. As the AUC is a part of the area of the unit square, therefore its value will invariably be between 0 and 1. It generates the diagonal line from (0, 0) to (1, 1) due to random guessing. 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 disease higher than a randomly chosen negative disease [6].

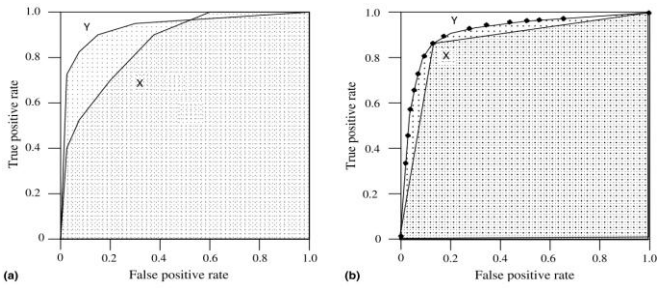

Figure 3: Shows two ROC graphs. (a) Shows the area under two ROC curves. (b) Shows the area under the curves of classifier (X) and a probabilistic classifier (Y)

The (Y) has better average performance classifier than (X) because it has a greater area such as in Figure (3-a) that shows the area under two ROC curves. So Figure (3-b) clarifies the area under the curve of both classifiers (X) and (Y). Classifier X characterizes the performance of Y when it is paired with an individual, fixed threshold. Also the performance of the two is equal at the fixed point (As threshold), as performance deteriorates to Y further from this point [6]. To the question of what is a good value for the area under the curve, one answer is to conduct an examination of what some of the likelihood ratios would be for various areas. For a test to be considered good it should have at least the LR+ equal to 2.0 and LR- is 0.5 or less, to correspond to an area of approximately 0.75. A test would be better with a likelihood ratio equal to 5 and 0.2, respectively, to correspond to an area of roughly 0.92. Better still would be likelihood ratios equal to 10 and 0.1, to correspond approximately to an area of 0.97 [10]. There is a need to make comparisons of the ROC curves to determine the best method. Comparisons between curves are based on the area under the ROC curve [5]. In such cases it is important to consider the correlation between the areas that are induced by the data into account, which would reduce the standard error and increase the power of the comparison. When comparing two areas the critical ratio is defined by [9].

$$z = \frac{A_1 - A_2}{\sqrt{SE_1^2 + SE_2^2 - 2rSE_1SE_2}} \quad (9)$$

where:A1 and A2 are the two areas and SE1 and SE2 the corresponding standard errors and r represents the correlation between the two areas due to working on the same set of data [17].

2.5 Thresholds (Cut-off Value) and the optimal cut-off value

The ROC-curve shows a sequence of cut-off values (thresholds). A single cut- point of a diagnostic test is a single point in the ROC space; but there are various possible cut-points of a diagnostic test that determine a curve in ROC space, that is called ROC curve [18].

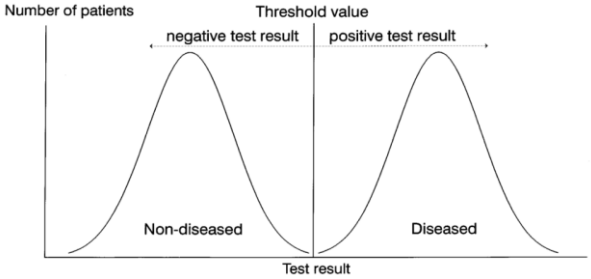

Figure 4: The probability distributions of diagnostic test results according to the threshold value

Figure 4 shows the probability distributions of the results of a hypothetical perfect diagnostic test. The results of the diseased and non-diseased individuals do not indicate any overlap and the selected threshold value is in between these distributions. A higher test result than the threshold value, may indicate a positive test. A test result below the threshold value indicates a negative test, with all diseased and non-diseased patients classified correctly. The probability distributions of diseased and normal over- lap. Any threshold value will result in the misclassification of some diseased patients as normal, or of some normal individuals as diseased, or both (Fig. 5).

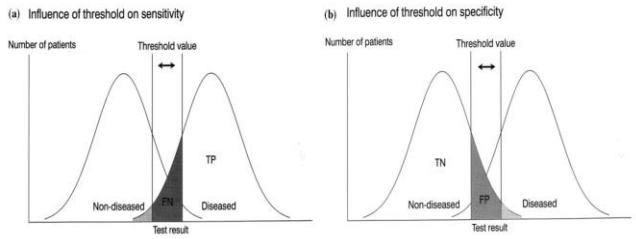

Figure 5: Impact of threshold on sensitivity and specificity

Applying a lower threshold value reduces the incidence of false-negative results (higher sensitivity; Fig. (5-a), but increases the incidence of false positives (lower specificity; Fig. (5-b). Raising the threshold value, however, will increase the incidence of false negatives (lower sensitivity; Fig. 5-a)) and decrease the incidence of false positives (higher specificity; Fig. 5-b)). A higher sensitivity is related to a reduction in specificity and a lower sensitivity with an

enhanced specificity. The ROC curve graphically represents this reciprocal relationship between sensitivity and specificity, computed for all possible threshold values (Fig.6).

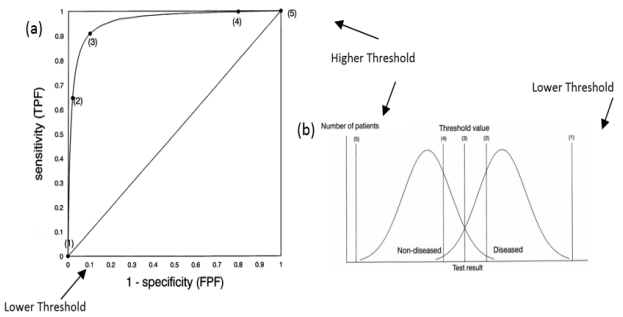

Figure 6: The ROC curve (a) graphically shows the relationship between sensitivity and specificity for all possible threshold values (b)

The sensitivity or TPF is indicated by the vertical axis of the graph above while the horizontal axis is represented by the false-positive fraction (FPF-1specificity). Every operating point on the ROC curve is a representation of the combination of sensitivity and specificity at a particular threshold value. At threshold value that are unrealistically high \ all patients are deemed to be normal, thus resulting in a TPF of 0 and a FPF of 0 (specificity = 1) and corresponding to the operating point in the lower left-hand corner of the ROC graph. If the threshold is lowered it will cause an increase in the TPF and FPF (lower specificity). At the lowest possible threshold, the TPF and FPF are both 1(specificity = 0), and this corresponds to the upper right-hand corner of the ROC graph [12], [5]. The important question is whether there is an optimal value for t or not, consider the Youden index:

$$y = \mathrm{Max}(tp - fp) \quad (10)$$

$$y = \mathrm{Max}(tp + tn - 1) \quad (11)$$

i.e., the maximum value of the sum of the sensitivity (tp) and specificity (tn) is minus 1. This index, like the AUC, is a descriptive measure of the ROC-curve. The optimal value of the cut-off point t is thus obtained when the sum (tp + tn) is at its maximum [14].

3. Data Analysis and Results

The data set used in this study consists of a sample of 234 observations (patents) and was obtained from the PhD dissertation of Dr. Dhahir Tahir Ahmad. The subjects were selected from the thalassaemia center in Erbil city. The data search was done manually by Dr. Dhahir [1]. The variables which are determined in this study are; Group variable; it is the variable to detect a group of thalassaemia patients and consists of two groups, the first group for major and the second group for minor with control group, while the other variables are; Red_Blood_Cell_Count, Hemoglobin and Red_Blood_Cell_ Diameter, which are scale variables.

Table 1: Displaying the sample size and positive, negative group

Variable	*Classific-ation variable*	*Sample size*	*Positive group*	*Negative group*	*Disease prevalence*
			Group = 1	*Group = 0*	
Red_Blood_Cell _Count	Group	234	147	87	62.8

Table (1) presents a sample size, which is based on the suggestion that meaningful qualitative conclusions can be drawn from ROC experiments conducted with a total of 234 observations or more. In this case the positive group (with disease) numbered 147 and negative (without disease) were 87. The percentage positive group or disease prevalence in this study is 62.8 (147/234), and clinically, the disease prevalence the probability of disease prevalence before performance of the test.

Table 2: Area under the ROC curve (AUC)

Area under the ROC curve (AUC)	*Standard Error*	*95% Confidence interval*	*Statistic test*	*Significance level P (Area=0.5)*
0.945	0.0167	0.907 to 0.970	26.625	<0.0001

In the table (2) this area (0.94) is quite good; it is near the perfect value of 1.0 and more sizeable than worst case value at 0. Which, means that an individual selected at random from the positive group has a test value larger than that for a randomly selected individual from the negative group in 94% of the time. When the variable under study is unable to distinguish one group from the other such as in Fig 7, i.e. Where the two distributions appear to be no different, the area in this case is equal to 0.5 (which indicates the ROC curve coincides with the diagonal). On the other hand, in the case of an ideal separation, of the two groups' values, i.e. The area under ROC curve will be equal to 1 (the ROC curve will reach the upper left corner of the plot). The 95% Confidence interval is the interval in which the true (population) area under the ROC curve is with 95% confidence. The P-value is the probability that the sample area under the ROC curve (0.945) is found when in fact, the true (population) area under the ROC curve is 0.5 (null hypothesis: Area = 0.5). The P is low (P = 0.0001<0.05) then it can be concluded that the Area under the ROC curve is significant, which means it is different from 0.5 and there is evidence that the laboratory test does have an ability to distinguish between the two groups.

Table 3: Criterion values and coordinates of the ROC curve

Criterion	*Sensi-tivity*	*95% CI*	*Speci-ficity*	*95% CI*	*+LR*	*-LR*
≤3	49.66	41.3 - 58.0	95.4	88.6 - 98.7	10.8	0.53
≤3.37	72.79	64.8 - 79.8	95.4	88.6 - 98.7	15.8	0.29
≤3.4	73.47	65.6 - 80.4	94.25	87.1 - 98.1	12.8	0.28
≤3.49	79.59	72.2 - 85.8	94.25	87.1 - 98.1	13.9	0.22

≤3.5	79.59	72.2 - 85.8	91.95	84.1 - 96.7	9.89	0.22
≤3.59	82.99	75.9 - 88.7	91.95	84.1 - 96.7	10.3	0.18
≤3.6	84.35	77.5 - 89.8	90.8	82.7 - 95.9	9.17	0.17
≤3.68	86.39	79.8 - 91.5	90.8	82.7 - 95.9	9.4	0.15
≤3.7	87.76	81.3 - 92.6	89.66	81.3 - 95.2	8.48	0.14
≤3.87*	92.52	87.0 - 96.2	89.66	81.3 - 95.2	8.94	0.08
≤3.9	92.52	87.0 - 96.2	87.36	78.5 - 93.5	7.32	0.09
≤3.91	93.2	87.8 - 96.7	87.36	78.5 - 93.5	7.37	0.08
≤3.93	93.2	87.8 - 96.7	86.21	77.1 - 92.7	6.76	0.08
≤3.98	94.56	89.6 - 97.6	86.21	77.1 - 92.7	6.86	0.06
≤4	94.56	89.6 - 97.6	82.76	73.2 - 90.0	5.48	0.07

It is clear from table (3) that the specificity of the model in the classification of β-Thalassaemia diseases is 89.66, and the sensitivity value is 92.52. These values correspond to the optimal cut-off value (3.87) as shown in Fig 7 and, the criterion value is indicated with a (*) sign indicating the value corresponding to the highest average of sensitivity and specificity (highest accuracy). To be considered a test should have a LR+ of at least 2.0 and a LR- of 0.5 or less. It is clear from table 3 that the value of LR+ is 8.94 and for LR- is equal to (0.08), which can be interpreted as the area under the curve being closer to very good area in this study.

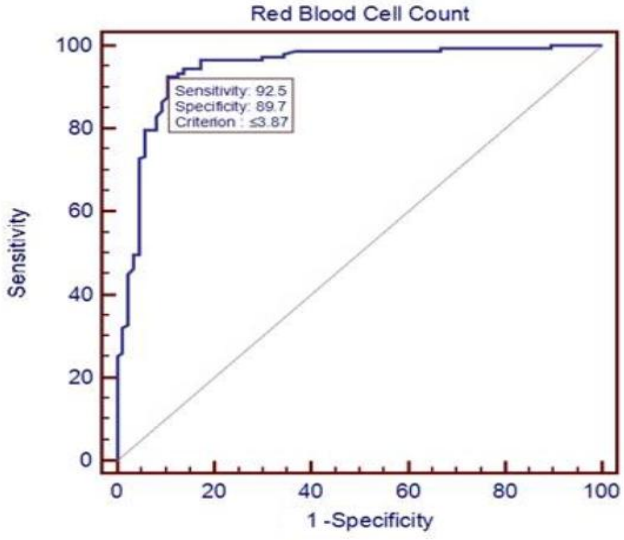

Figure 7: ROC Curve graph for the test Red Blood Cell Count

Fig 8 shows the relationship between sensitivity and specificity by using different cut values, and from it, one can determine the optimal cut-off value that is equal to 3.87.

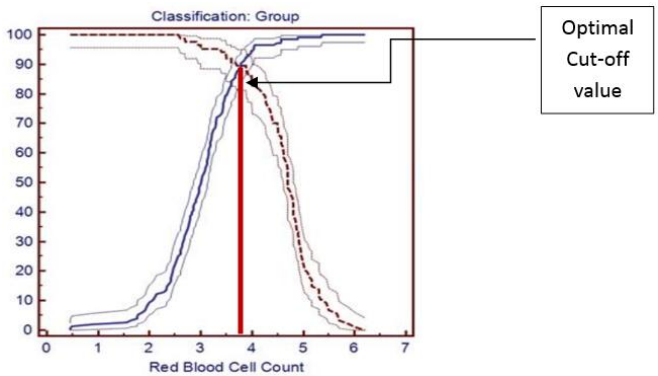

Figure 8: Relationship between sensitivity and specificity by using different cut values

Of the ROC curves of three diagnostic tests, test one is the results from the Red_Blood_Cell_Count and test two is the Hemoglobin, and test three is the results from the Red_Blood_Cells_Diameter, all these tests being applied on the same subjects to classify the same disease and ROC analysis is used to determine if there is any difference between the three diagnostic tests as shown in Fig 9.

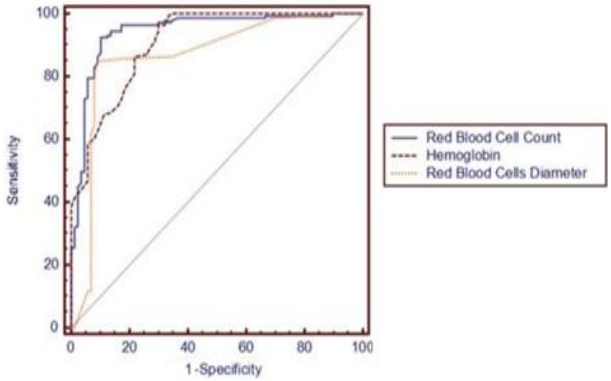

Figure 9: Three different diagnostic tests

The above figure depicts three different ROC curves according to three different tests. The areas under the curve (AUC) are shown in table (4), the AUC for test one is (0.945) and for test two it is 0.908; , also, the AUC for test three is 0.865, while test one has higher AUC value than curves the curves of test two and test three. This means that test one is superior to both tests two and three, and the curve is closer to the perfect discrimination.

Table 4: AUC for three different diagnostic tests

Variables	*AUC*	*SE*	*95% CI*
Red_Blood_Cell_Count	0.945	0.0167	0.907 to 0.970
Hemoglobin	0.908	0.0198	0.863 to 0.941
Red_Blood_Cells_Diameter	0.865	0.0285	0.814 to 0.906

If there is a wish to obtain an overall view of the performances of these three different diagnostic tests, it is possible to make a comparison of the area under the ROC curves. The overall performance of test one is better than test two test three at all the threshold points.

Table (5-a): Pairwise comparison of ROC curves between Red_Blood_Cell_Count V.S Hemoglobin

Difference between areas	*Standard Error*	*95% Confidence interval*	*Statistic test*	*Significance level*
0.0369	0.0166	0.00437 to 0.0694	2.223	P = 0.0262

The results are as shown in table (5-a) where the difference area between Red_Blood_Cell_Count and Hemoglobin is 0.0369, and the upper limit of this value of the confidence interval is 0.0694 while the lower limit value is 0.00437 with standard error of 0.0166. , Also, the area test statistic is z = 2.223 and the two-tailed p-value is 0.0262, which means there is a significant difference between the two tests. This means to say that if is a much better performance than the Hemoglobin on this data set.

Table (5-b): Pairwise comparison of ROC curves between Red_Blood_Cell_Count V.S Red_Blood_Cells_Diameter

Difference between areas	*Standard Error*	*95% Confidence interval*	*Statistic test*	*Significance level*
0.0796	0.0308	0.0192 to 0.140	2.582	P = 0.0098

It is clear from Table (5-b), that the difference between Red_Blood_Cell_Count area and Red_Blood_Cell_Diameter area is 0.0796, then the upper and lower limits values of the confidence interval are 0.140 and 0.0192 respectively, with standard error of 0.0308, while the area statistic test is 2.582 and the significant level is (p-value = 0.0098), meaning that there is a high significant difference between the two tests above. In other words, when two tests are compared, the Red_Blood_Cell_Count test has a high significantly better performance than the Red_Blood_Cell_Diameter.

Table (5-C): Pairwise comparison of ROC curves between Hemoglobin V.S Red_Blood_Cells_Diameter

Difference between areas	*Standard Error*	*95% Confidence interval*	*Statistic test*	*Significance level*
0.0428	0.0316	-0.0192 to 0.105	1.353	P = 0.1760

Table (5-C) shows the difference between Hemoglobin test and Red_Blood_Cell_Diameter test, and the difference between the two tests area is 0.0428 while the upper and lower limit value of this difference are -0.0192 and 0.105 respectively. The standard error is 0.0316, and the statistic test area is 1.353 while the p-value is 0.176, which, means there is no statistically significant difference between the two tests.

4. Conclusions (Discussion)

In this study, the methodology provided a powerful technique to distinguish between diagnostic tests in β-thalassaemia diseases. The main aim of the conducted study was to analyze the comparison between tests, ROC curves of three diagnostic tests (Red_Blood_Cell_Count, Hemoglobin, and Red_Blood_Cells_ Diameter) and these tests being applied on the same subjects to classify the same disease and to use ROC analysis to examine if there is any difference between them. The result shows that the first test has higher AUC value than curves of the second and third tests, so the difference areas between Red_Blood_Cell_Count and Hemoglobin) is 0.0369 , and the area test statistic is 2.223 with the p-value ar 0.0262, which means if the tests are compared at all cut-off points at once, the Red_Blood_Cell_Count test would show a much better performance than the Hemoglobin on this data set. In other words, the difference between Red_Blood_Cell_Count area and Red_ Blood_ Cell_ Diameter area is 0.0796, and also the area statistic test is 2.582 and the significant level is 0.0098, which means that when two tests are compared, the Red_Blood_Cell_Count test has a significantly better performance than the Red_Blood_Cell_Diameter. The other result shows the difference between Hemoglobin test and Red_ Blood_ Cell_ Diameter test, while the difference between two tests area is 0.0428. Additionally, statistic test area is 1.353 and the p-value is (0.176), which means there is no statistically significant difference between the two tests. The percentage positive group or disease prevalence in this study is 62.8, clinically, the disease prevalence is the same as the probability of disease being present before the test is performed. The most interesting results was that the Red_Blood_Cell_Count test was shown to have a greater area under curve than the other diagnostic tests. This area is 0.94, which is almost the perfect value of 1.0 and a lot larger than worst case value equal to 0.5, which means that for 94% of the time, a randomly chosen individual from the positive group has a test value higher than that for a randomly selected individual from the negative group. The P-value for the sample area under the ROC curve (0.945) is 0.0001and less than significant level (0.5), so we can conclude that the area under the ROC curve is significant, meaning that it is different from 0.5 and there is proof that the laboratory test can distinguish between the two groups. Finally, the specificity of the model in the classification of β-Thalassaemia diseases according to Red_Blood_Cell_Count test is 89.66, and the sensitivity value is 92.52. These values correspond to the optimal cut-off value of 3.87, and the value of LR+ is 8.94 while LR- is equal to 0.08, which, can be interpreted as the area under the curve being closer to very good area in this study.

References

[1] Ahmad Dahir Tahir, "Estimation and Evaluation of Some Biophysical Parameters in β – Thalassaemia Patients in Erbil City". A Ph.d thesis , college of medicine, hawler medical university, 2009.
[2] Bradley, A.P., "The use of the area under the ROC curve in the evaluation of machine learning algorithms". Pattern Recogn, 1997.
[3] Choi C. K. Bernard. "Slopes of a Receiver Operating Characteristic Curve and Likelihood Ratios for a Diagnostic Test". American Journal of Epidemiology, The Johns Hopkins University, School of Hygiene and Public Health, Vol. 148, No. 11, 1998.
[4] Cizek, Gregory J. & Fitzgerald, Shawn M. "Methods, Plainly Speaking: An Introduction to Logistic regression. Measurement & Evaluation in Counseling and Development". Vol.31, January, 1999.
[5] Eng ,John "Receiver Operating Characteristic Analysis: A Primer", Academic Radiology, Vol 12, No 7, July 2005.
[6] Fawcett, Tom. "An introduction to ROC analysis". Institute for the Study of Learning and Expertise, USA. Pattern Recognition Letters 27 (2006) 861–874.
[7] Fraas, John W. and Newman, Isadore, "Ordinary Least Squares Regression, Discriminant Analysis, and Logistic Regression: Questions Researchers and Practitioners Should Address When Selecting an Analytic Technique". Paper Presented at the Annual Meeting of the Eastern Educational Research Association (Hilton Head Island, GA, February 26-March 1, 2003).
[8] Hamel, Lutz. "Model Assessment with ROC Curves". Department of Computer Science and Statistics, University of Rhode Island, USA, 2007.

[9] Hanley, J.A., McNeil, B.J., "The meaning and use of the area under a receiver operating characteristic (ROC) curve". Radiology 143, 29–36.1982.
[10] http://www.childrensmercy.org/stats/ask/roc.asp, Stats: "ROC curve" (August 18, 1999).
[11] Metz, E. Charles. "Basic Principles of ROC Analysis". Seminars in Nuclear Medicine, Vol. VIII, No. 4 (October), Department of Radiology, University of Chicago, and the Franklin McLean Memorial Research Institute, Chicago, Ill, 1978.
[12] R. A.van Pattynama T. M. P. Erkel, "Receiver operating characteristic (ROC) analysis: Basic principles and applications in radiology". European Journal of Radiology 27 (1998) 88–94.
[13] Soderstrom, Irina R. and Leitner, Dennis W. "The Effects of Base Rate, Selection Ratio, Sample Size, and Reliability of Predictors on Predictive Efficiency Indices Associated with Logistic Regression Models". Paper Presented at the Annual Meeting of the Mid-Western Educational Research Association (Chicago, IL, October 15-18, 1997).
[14] Steyn S. H. "Using ROC-analysis to determine good screening tests based on continuous variables". Statistical Consultation Service, 2012.
[15] Swets, J., "Measuring the accuracy of diagnostic systems". Science 240, 1285 –1293, 1988.
[16] Webb I. Geoffrey and Ting Ming Kai. "On the application of ROC analysis to predict classification performance under varying class distributions". School of Computer Science and Software Engineering and Gippsland School of Computing and Information Technology, Monash University, Australia. Prepublication draft of paper accepted for publication in Machine Learning, roc-preprint.tex; 20/04/2004.
[17] Westin Kallin Lena. "Receiver operating characteristic (ROC) analysis Evaluating discriminance effects among decision support systems". Department of Computing Science, Umeå University, SE-90187 Umeå, Sweden, UMINF 01.18, 2002.
[18] Zhu Wen, Zeng Nancy and Wang Ning, "Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS Implementations". Health Care and Life Sciences, NESUG 2010.