Original Research Article

Investigation of Diagnostic Test Performance Using Receiver Operating Characteristic and Fundamental Concepts of Information Theory

ABSTRACT

Aims: Receiver operating characteristic(ROC) curve and a fundamental concept of information theory is directly applicable to evaluation of diagnostic test performance. In this study,the performance of the two diagnosis tests on the field of rheumatic disorder is analyzed byusing receiver operating characteristic and fundamental concepts of information theory. The aims of this study to investigate which diagnosis tests has better performance and to demonstrate which test can be an alternative to gold standard test by carrying out ROC and fundamental concepts of information theory.

Study design: ROC analysis and fundamentals concept of information theory(entropy, conditional entropy, mutual information).

Place and Duration of Study: Department of Statistics, between July 2012 and July 2013.

Methodology: Anti-Streptolysin O (ASO) is a value which is used to learn whether the patients have group "*A*" beta-hemolytic streptococcal infection which causes rheumatic disorder diseases. ASO values of 68 subjects who applied to Thoracic and Cardiovascular Surgery Training and Research Hospital in Istanbul for the diagnosis of rheumatic disorder were used. ASO values were evaluated according to Turbidimetric tests of two different firms. These tests were called as I. Turbidimetric test and II. Turbidimetric test. Both ROC and Information Theory analyses were applied to the data. Therefore, both firms' Turbidimetric test diagnostic test performances were evaluated and which diagnostic test had better performance was determined.

Results: According to Roc curve results, Area Under curve(AUC) is calculated 0.98 for I. Turbidimetric test and 0.90 for II. Turbidimetric test . On account of information theory analysis; the entropy value is the same but mutual information values are different. According to the result of mutual information, I. Turbidimetric test provides more diagnostic information than II. Turbidimetric test. Therefore I. Turbidimetric test dominates II. Turbidimetric test. Based on these results, it can be verified that mutual information value is parallel to AUC value. Another result is found for threshold values of tests. According to results an alternative threshold values for tests can be obtained by using mutual information.

Conclusion: The Turbidimetric tests' performances are examined using ROC and information theory. With regard to ASO values, it is concluded that I. Turbidimetric test is more likely to show the similarity to Nefelometric test in comparison with II. Turbidimetric test. Using I. Turbidimetric test has financial benefits to clinicians, since it is less expensive in contrast with Nefelometric test.

Keywords: Diagnostic tests, receiver operating characteristic (ROC), entropy, conditional entropy, mutual information.

1. INTRODUCTION

Diagnostic tests are widely used in many areas. In particular, these tests have huge importance in medicine sector. By courtesy of early and accurate diagnosis, the morbidity

and mortality of disease can be reduced. For this reason, it is important to compare various diagnostics test with each other under specific clinical conditions in order to determine which one is the best to use.

One of the approaches used to analyze the performance of diagnostic tests is ROC theory. The roots of ROC theory are laid on statistical decision theory. ROC theory is related to many fields. It was not only used in the 1950's for radio signals, but also used in the 2010's for predictions of land changes, species distributions and ecological niches [1]. In particular, the usage of ROC analysis was canalized to the medicine sector after 1960's. Since that time, ROC played an essential role in medicine sector and it is still widely used in this sector. ROC curves became the standard approach to summarizing diagnostic test performance after published a medical application of this method as [2].

The other approach which is used to analyze the performance of diagnostic tests in recent years is information theory. Information theory was developed by Claude Shannon (1948). In Shannon's theory,[3], the information is associated with uncertainty. This theory of knowledge and uncertainty for the measurement is based on a mathematical basis. Metz, Goodenough and Rossmann[4] developed a formula used in assessing the performance of diagnostic tests by using information theory. After this work, Mossman and Somoza [5] developed a new mathematical and graphical method to evaluate and compare the performance of diagnostic tests for the value of any prevalence by using the properties of the ROC analysis and information theory approach. In [6] obtained the distance between patients and healthy distributions by using the concept of relative entropy. Benish[7] investigated the concept of relative entropy with a different perspective.

In this study,the performance of the two diagnosis tests on the field of rheumatic disorder is analyzed by using receiver operating characteristic and fundamental concepts of information theory. ASO values are used for the diagnosis of rheumatic disorder. ASO is a value which is used to learn whether the patients have group "*A*" beta-hemolytic streptococcal infection which causes these diseases. In this article, ASO values are measured by using Turbidimetric tests which belong to two different firms. These tests were called as I. Turbidimetric test and II. Turbidimetric test. The aims of this study to investigate which Turbidimetric test has better performance and to demonstrate which test can be an alternative to gold standard (Nefelometric test) test.

2. MATERIAL AND METHODS

2.1.RECEIVER OPERATING CHARACTERISTIC(ROC) CURVE

The ROC curve is a fundamental tool for diagnostic test evaluation. When it is considered the results of a particular test in two populations, one population with a disease, the other population without the disease, it will be rarely observed a perfect separation between the two groups. Indeed, the distribution of the test results will overlap, as shown in the following Figure 1.

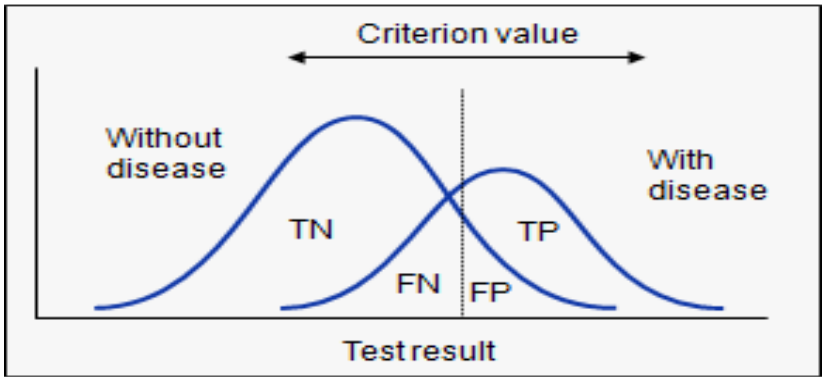

Fig.1. Two distributions of diseased and non-diseased group

For every possible cut-off point or criterion value (threshold value) is selected to discriminate between the two populations, there will be some cases with the disease correctly classified as positive (TP = True Positive fraction), but some cases with the disease will be classified negative (FN = False Negative fraction). On the other hand, some cases without the disease will be correctly classified as negative (TN = True Negative fraction), but some cases without the disease will be classified as positive (FP = False Positive fraction). The different fractions (TP, FP, TN, FN) are represented in Table 1 [8].

Table 1.Different fractions(TP, FP, TN, FN)

Test Results	**Truth**	
	Positive	**Negative**
Positive	TP	FP
Negative	FN	TN

There are some criteria to measure the performance of the diagnostic test. Sensitivity, specificity and efficiency (accuracy) are some of the performance criteria. These criteria are based on the Table 1. Sensitivity is a probability that a test result will be positive when the disease is present. It is equal to TP fraction. Specificity is a probability that a test result will be negative when the disease is not present. It is equal to one minus FP fraction. Efficiency is calculated by total number of TP and FP over sample size. It gives a clue about the accuracy of the diagnostic test.

In a receiver operating characteristic (ROC) curve the true positive fraction (TP or sensitivity) is plotted in function of the false positive fraction (FP or 1- specificity) for different cut-off points. Each point on the ROC curve represents sensitivity and one minus specificity pair corresponding to a particular decision threshold. FP fraction amounts to costs and TP fraction amounts to benefits. A test with perfect discrimination (no overlap in the two distributions) has a ROC curve that passes through the upper left (northwest) corner (100% sensitivity, 100% specificity). Therefore the closer the ROC curve is to the upper left (northwest) corner, the higher the overall accuracy of the test [9]. The most commonly used global index of diagnostic accuracy is the area under the ROC curve (AUC) [10]. The total area under the curve(AUC) is a measure of the performance of the diagnostic test since it reflects the test performance at several possible cut-off levels. The area lies in the interval [0,1] and the larger area, the better performance.

2.2.BASIC CONCEPTS OF INFORMATON THEORY

The performance of a diagnostic test is frequently described in terms of the amount of information it provides. A fundamental concept of information theory, entropy and mutual information, is directly applicable to evaluation of diagnostic test performance. In this section we introduce most of the basic definitions in information theory required for evaluation of diagnostic test performance[9].

The entropy of a random variable is a measure of the uncertainty of the random variable. It is the number of bits on average required to describe the random variable. Let X be a discrete random variable, taking a finite number of possible values $X_1, X_2, \dots, X_n$ with respective

probabilities $p_i \geq 0$ *for* $i = 1, \ldots, n$ *and* $\sum p_i = 1$. The Shannon entropy $H(X)$is defined by [11].

$$H(X) = -\sum_x p(x) log_2 p(x) \quad (1)$$

If $(X, Y) \sim p(x, y)$, the conditional entropy $H(X\backslash Y)$ is defined as

$$H(X\backslash Y) = -\sum_{x,y} p(x, y) log_2 \frac{p(x,y)}{p(y)} \quad (2)$$

Mutual information is a measure of the distance between two probability distributions. The mutual information of two random variables is a quantity that measures the mutual dependence of the two variables. The interpretation is that when mutual information is absent, marginal distributions are independent and their entropies add up to total entropy. The mutual information $I(X;Y)$is the reduction in the uncertainty of Xdue to the knowledge of Y. $I(X;Y)$ is calculated by the formula given below

$$I(X;Y) = H(X) - H(X\backslash Y) \quad (3)$$

2.3.INFORMATION-BASED MEASURES OF DIAGNOSTIC TEST PERFORMANCE

The performance of a diagnostic test is identified with respect to amount of the reduction of disease's uncertainty. The purpose of this section is to demonstrate how basic concepts in information theory apply to the problem of quantifying diagnostic test performance.
While evaluating the performance of the diagnosis test using the information theory, we need to explain the concepts of test results and disease statement. Disease statement is denoted by D. On the condition that there are two statements such as the existence or the non-existence of a disease, we can specify the disease statement as follows,

$$D = \{Di\} \quad i: \{+, -\}$$

$D+ =$ Get ill before diagnosis test
$D- =$ Get no ill before diagnosis test
$T+=$ Positive test results
$T-=$ Negative test results

The probability distribution of the disease statement before the test is defined with $P(D+)$ and $P(D-)$ values. In this case, the entropy before the test is calculated as below.

$$H(D) = P(D+) log_2 P(D+) + P(D-) log 2 P(D-) \quad (4)$$

After the diagnosis test is applied, the uncertainty of the disease statement changes. On the condition that the diagnosis test results are known, the entropy of the disease statement is called conditional entropy and is calculated according to the formula below.

$$H(D\backslash T) = P(T+)[P(D+\backslash T+) log_2 P(D+\backslash T+) + P(D-\backslash T+) log_2 P(D-\backslash T+)] + P(T-)[P(D+\backslash T-) log_2 P(D+\backslash T-) + P(D-\backslash T-) log_2 P(D-\backslash T-)] \quad (5)$$

If $H(D)$ is defined as pretest entropy, we need to define $H(D\backslash T)$ as the expected value of posttest entropy [8,11]. Besides, the difference between $H(D)$ and $H(D\backslash T)$ is called as mutual information and it is denoted by $I(D;T)$. Mutual information is the reduction in the

uncertainty of D due to the knowledge of T. It is the general criterion of what the diagnosis test will tell us. $I(D;T)$isdefined as

$$I(D;T) = H(D) - H(D\backslash T) \quad (6)$$

2.4.APPLICATION

Turbidimetric test and Nefelometric test are used for the diagnosis of rheumatic disorder. Both tests are based on the principal of impurity in the blood. Nefelometric test is accepted as the gold standard in the analysis of plasma protein with micro molecule of which molecule massiveness is measured with milligram. If Nefelometric test results are in the range of 0-200 IU/ml reference interval, the diagnosis is resulted as healthy for the person. If Nefelometric test results are over [0, 200] IU/ml reference interval, the diagnosis is resulted as ill for the person [12].

New Turbidimetric tests are alternatives to Nefelometric test and they are becoming more precise day by day for the specific proteins such as ASO, which is used for the diagnosis of rheumatic disorder [12]. Furthermore, while the unit cost of the Nefelometric test is more than the unit cost of the Turbidimetric test, there are disadvantages to the Nefelometric test such as the requirement of more space in the laboratory, occupying additional personnel and orientation of them. There are no practical differences between those two tests with regard to the duration of test results. Each laboratory is required to decide to work with whether Turbidimetric test and Nefelometric test due to its substructure, patient potential and establishment requirement.

In this study, ASO values being the first phase of the diagnosis of rheumatic disorder are measured by using Turbidimetric tests which belong to two different firms. These tests were called as I.Turbidimetric test and II.Turbidimetric test. The aims of this study to investigate which Turbidimetric test has better performance and to demonstrate which test can be an alternative to Nefelometric test.

3. RESULTS AND DISCUSSION

Diagnosis values, I. Turbidimetric test results and II. Turbidimetric test results are coded as vectors in R programme. After coding process, ROC curves of the both tests are generated in the Figure 2.

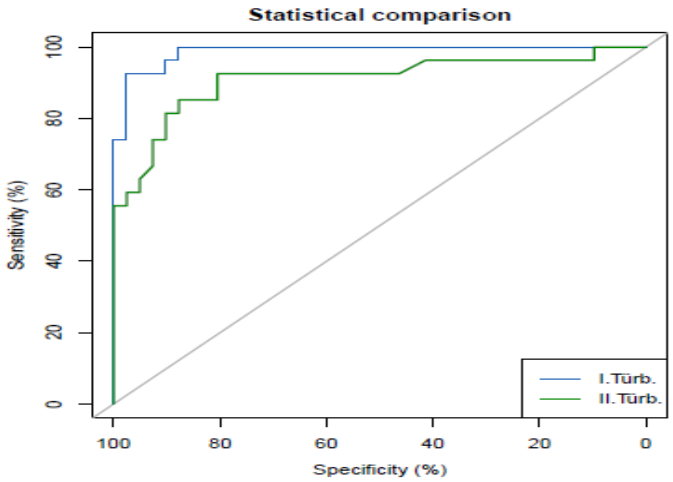

Fig.2. Statistical Comparison for Two ROC Curves

In Figure 2, it is observed that I. Turbidimetric test dominates II. Turbidimetric test for all sensitivity and specificity values. According to these results, AUC is calculated 0.98 for I. Turbidimetric test and 0.90 for II. Turbidimetric test.

The sensitivity(SE), specifity(SP) and efficiency(Eff) of both test for different threshold values are given in Table 2. Using sensitivity results, the threshold values 165 and 48 are chosen for I.Turbidimetric and II. Turbidimetric test respectively. Probabilities of detecting actually ill people on these threshold values are the greatest among all threshold values.

Table 2. The sensitivity(SE), specifity(SP) and efficiency(Eff) of both test for different threshold values

Test	Threshold V.	Sensitivity	Specificity	Efficiency
I.Turbidimetric	165	0.96	0.90	0.91
	173	0.92	0.97	0.95
	197	0.77	0.97	0.89
II.Turbidimetric	48	0.96	0.41	0.63
	101	0.92	0.78	0.83
	202	0.51	1.00	0.80

According to the specificity(SP) results in Table 2, the threshold values 173 and 202 are chosen for I.Turbidimetric and II. Turbidimetric test respectively. Probabilities of detecting actually healthy people on these threshold values are the greatest among all threshold values. Using the efficiency value(Eff) in Table.2, 173 is chosen the largest percent correct value which maximizes efficiency.

Diagnostic performance is measured in units of information for I. Turbidimetric and II. Turbidimetric test. Entropy, conditional entropy and mutual information values of both test are given in Table 3.

Table3. Entropy, conditional entropy and mutual information of two tests

Tests	$H(D)$	$H(D\backslash T)$	$I(D;T)$
I. Turbidimetric	0.96	0.50	0.46
II. Turbidimetric	0.96	0.63	0.33

On account of information theory analysis; if the disease statement is taken as the random variable, the random variable is indicated either as the presence or the absence of the disease before the diagnostic test. Under this circumstance the entropy of the disease is only affected with the possibility of disease existence or disease non existence. Since these possibilities are equal in both tests, the entropy of the disease is the same. In Table 3, the entropy value is the same but mutual information values are different. According to the result of mutual information, I. Turbidimetric test provides more diagnostic information than II. Turbidimetric test. Therefore I. Turbidimetric test dominates II. Turbidimetric test. Based on these results, it can be verified that mutual information value is parallel to AUC value.

Another result of information theory analysis is the measurement of mutual information values for all threshold values. While mutual information can be measured for all threshold values, AUC isn't measured for all threshold values. Because of this reason, it is preferred mutual information value rather than AUC value. Table 4 represents four threshold values maximizing mutual information for each test. Table 4 doesn't contain the threshold 165 of I. Turbidimetric test and the threshold 197 of I. Turbidimetric test. These threshold values have the highest sensitivity and specificity, but they do not have the highest mutual information values. These results prove that, for the overall quality, neither sensitivity nor specificity but the results of mutual information should be examined. Alternative threshold values can be obtained by using information theory.

Table 4. Mutual information of two tests for different threshold values

Threshold Values		$I(D;T)$	
I. Turbidimetric	**II. Turbidimetric**	**I. Turbidimetric**	**II. Turbidimetric**
173	102	0.70	0.41
142	118	0.67	0.41
185	124	0.64	0.40
171	101	0.64	0.39

4. CONCLUSION

In this study; ROC which is a long-standing method for the evaluation of the diagnostic test performance and information theory which has been used recently to evaluate the diagnostic test performance are presented in detail.

This study aims to investigate which Turbidimetric test has better performance. This performing test is going to be conducted during the study in order to demonstrate whether it can be an alternative to Nefelometric test which is currently the gold standard for the diagnosis of rheumatic disorder. Because of this reason, AUC of Nefelometric test is equal to 1. The Turbidimetric tests' performances are examined using ROC and information theory. With regard to AUC values, it is concluded that I. Turbidimetric test is more likely to show the similarity to Nefelometric test in comparison with II. Turbidimetric test. Using I. Turbidimetric test has financial benefits to clinicians, since it is less expensive in contrast with Nefelometric test.

As a result of Information Theory analysis, the threshold value of 173 is the largest percent correct value which maximizes mutual information. Based on this largest percent correct threshold value, it can be deduced that 0-200 UI/ml reference interval which is mentioned in

the medicine literature for Nefelometric test can be replaced with a "*new*" 0-173 UI/ml reference interval. The use of this new reference interval provides more accuracy and leads to less error in the diagnosis of ASO values. As a conclusion of the study, it is recommended to the clinicians to implement I. Turbidimetric test with a new reference interval for the diagnosis of rheumatic disorder.

It is aimed that this study will hopefully give various points of view to the researchers who want to make research on this subject by explaining how the tests used for the diagnosis of various diseases are evaluated with this way.

REFERENCES

[1]Pontius Jr RG. & Parmentier B. 2014. Recommendations for using the Relative Operating Characteristic (ROC). Landscape Ecology 29(3): 367-382.

[2]Lusted LB. 1971. Signal detectability and medical decision-making. Science 171: 1217-1219.

[3]Shannon CE. 1948. A Mathematical Theory of Communication. Bell System Technical Journal 27:379-423, 623-656.

[4]Metz CE, Goodenough DJ & Rossmann K. 1973. Evaluation of receiver operating characteristic curve data in terms of information theory, with applications in radiography. Radiology 109:297-303.

[5]Mossman D. & Somoza E. 1989. Maximizing diagnostic information from the dexamethasone suppression test: An approach to criterion selection using receiver operating characteristic analysis. Archives of General Psychiatry 46*:* 653-660.

[6]Lee WC. 1999. Selecting diagnostic tests for ruling out or ruling in disease: the use of the Kullback-Leibler distance. International Journal of Epidemiology 28: 521-525.

[7]Benish WA. 2002. The use of information graphs to evaluate and compare diagnostic tests. Methods Inf Med 41: 114-118.

[8]Ege Oruç Ö. & Kanca A. 2011. Evaluation and comparison of diagnostic test performance based on information theory. International Journal of Statistics and Applications 1(1): 10-13

[9]Zweig MH & Campbell G.1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine*.* Clinical Chemistry 39: 561-577.

[10]Faraggi D. & Reiser B. 2002 Estimation of the area under the ROC curve. Statistics In Medicine Statist. Med. 21: 3093–3106.

[11]Cover TM. & Thomas JA. 2006. Elements of Information Theory 2nd Edition,New Jersey, USA: John Wiley & Sons, Inc. Pp.13-22.

[12]http://www.randox.com/brochures/PDF%20Brochure/LT008.pdf