INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XI, November 2025

Evaluation, Agreement and Interpretation of Independent Radiologist Assessments - ROC Analysis

Srikanth Chivukula; Dr. Uma Shankar; Dr. Raghunath Reddy

Rayalaseema University

DOI: https://doi.org/10.51583/IJLTEMAS.2025.1411000040

Received: 10 November 2025; Accepted: 20 November 2025; Published: 06 December 2025

ABSTRACT:

ROC curves are useful for comparing two independent observations and the outcomes of their assessments for the same disease in a given study. In general, the test with the higher AUC may be considered better and is an effective way to summarize the overall diagnostic accuracy of the test. AUC scores are convenient to compare multiple classifiers. Nonetheless, it is also important to check the actual curves especially when evaluating the final model. In the ROC curve analysis, the choice of the optimal cutoff value depends both on probabilistic and clinical considerations. From a probabilistic standpoint, one can use the coordinates of the ROC curve to identify the cutoff that maximizes the discrimination between true-positive rate and false-positive rate(1).

The name "Receiver Operating Characteristic" came from. ROC analysis is part of a field called "Signal Detection Theory" (3) developed during World War II for the analysis of radar images. It was not until the 1970's that signal detection theory(4) was recognized as useful for interpreting medical test results.

Key Words: Sensitivity, Specificity, ROC curve, Area under the curve (AUC), Kappa.

INTRODUCTION:

ROC analysis originated in the early 1950's with electronic signal detection theory . The applications of ROC methodology in diagnostic radiology dates back to the early 1960's. The first ROC curve in diagnostic radiology was calculated by Lusted (1960) who re-analyzed the previously published data on the detection of pulmonary tuberculosis and showed the reciprocal relationship between the percentage of false positive and of false negative results from the different studies of chest film interpretations (2). Since then, several authors have used ROC methodology to diagnostic imaging systems. The work of Dorfman and Alf (1968) was a pioneering step toward objective curve fitting and the use of computerized software in ROC analysis (2). A maximum likelihood approach under binormal assumption was developed in 1968 by Metz(6).

In the 1970s and 80s, the technique had considerable relevance to medical test evaluation and decision making, and the decades followed seen much development and use of the technique in areas such as Life sciences, radiology, cardiology, clinical chemistry, and epidemiology.

Now the applications have a much wider range and features in subjects as diverse as Clinical Research, Medical devices, sociology, experimental psychology, atmospheric and earth sciences, finance, machine learning, and data mining, among others.

The derived summary measure of accuracy of a diagnostic performance, one can compare individual tests or judge whether the various combination of tests (e.g. combination of imaging techniques or combination of readers) can improve diagnostic accuracy.

RESEARCH AND METHODS

In the context of the ROC curve analysis, the test with the higher AUC is considered better and is an effective way to summarize the overall diagnostic accuracy of the test. The practical usefulness of ROC curves in

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XI, November 2025

comparing two independent observations and the outcomes of their respective assessments for the same disease will be discussed. . Further, AUC scores are usually treated as the convenient way to compare multiple classifiers. Nonetheless, it is also important to check the actual curves especially when evaluating the final model. In the ROC curve analysis, the choice of the optimal cutoff value depends both on probabilistic and clinical considerations. From a probabilistic standpoint, one can use the coordinates of the ROC curve to identify the cutoff that maximizes the discrimination between true-positive rate and false-positive rate.

In this section, a real data scenario is considered to exhibit the practical applicability of comparing two independent ROC curves. Breast Cancer is one of the chronic diseases, which is rapidly growing and noticed among women population across the world. Although the survival rate of breast cancer patients has improved, survival remains poor for advanced stage patients, especially for patients with locally advanced breast cancer.

Simulated data similar to the outcome response measures from a randomized, open-label, clinical study with indication as Metastatic Breast Cancer was considered. This simulated data has been utilized to study the patterns and the differences in the evaluations of the independent radiologists' observations. Several methods used to evaluate the chemotherapeutic response of breast cancer patients however, MRI is accepted as the best imaging modality for monitoring the response to NAC. Specific reports have shown that dynamic contrast-enhanced MRI can reflect the tumor pathophysiologic response to NAC before any changes occur in the tumor volume.

Two sets of response evaluation criteria have been considered for solid tumors; the response evaluation criteria in solid tumor is based on (RECIST) criteria. This criteria helped to convert radiologic imaging observations into a quantitative assessment of a tumor's response to therapy. The assessments per RECIST criteria were obtained from two independent investigator assessments. The assessment reports had included the following:

(a) the measurements of the lesions and the count

(b) assessment of pathologic lymph nodes

(c) criteria for disease progression and definition for minimal diameter

(d) Comments on new lesions included in the target lesions.

The aim of this study is to compare the performance of both assessments and evaluating the response of breast cancer patients to evaluate the differences, if any. To evaluate efficacy and safety of a generic recombinant human monoclonal antibody(MAB) as treatment therapy in patients with Metastatic Breast Cancer. Evaluations and outcomes of the treatment were recorded indicating the onset of a PR or CR was done at the end of 6 months. The primary end point of the study was to assess efficacy as Objective Response Rate (Complete Response and Partial Response) assessed by the criteria

Tumor response and lesions were categorized into target lesions and non-target lesions and new lesions were evaluated by CT scan. Two Independent assessor's verified the assessments) of tumor response and based on the subjects overall response final assessment were considered for evaluation of efficacy.

The outcomes from the data included 72 patients with a pathologically proven diagnosis of locally advanced breast cancer, who were treated between May 2013and May 2014. Patients ranged in age from 28 to 63 years, with the mean age of 49.0 ± 9.6 years.

STATISTICAL METHODS AND DISCUSSIONS

Two radiologists, who were experienced in evaluating radiological finding of the breast and unaware of the Histopathological outcomes, interpreted all the cases in this study. According to RECIST criteria a lymph node with a short axis were considered measurable, and these lymph nodes were assessed as target lesions. A maximum of two target lesions were assessed. The longest diameter of tumor masses or the short axis of lymph nodes were measured. After chemotherapy, the longest diameter of the tumor and the short axis of the visible

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XI, November 2025

lymph nodes were measured. If the lesions did not disappear completely, but still could not be precisely measured, then it was assigned a value of 5mm.If the lesion was totally absent after therapy, then it was assigned a value of 0 mm. If the target lesion was split in to fragments after chemotherapy, the longest diameterof fragments were added to the target lesion sum, as the RECIST recommendation (7).

The sensitivity, specificity, and the P value after the comparisons, were calculated with respect to the response evaluation, using the pathologic results as a reference. The software used to derive the results are from the open source (5)

Additionally, ROC analysis was performed on the rating data to assess and compare the diagnostic performance of both the assessments. To summarize the overall performances, the areas under the ROC curves (AUC) were calculated and compared. Statistically significant differences between the AUC values are reported in terms of the 95% confidence interval (CI). An Interactive Web-tool for ROC Curve Analysis Using R Language Environment is used to analyse both the datasets. The results are presented below.

The ROC curve for the first dataset evaluated by the first radiologist is shown below in Figure 1.

In Table 1, below shows the TPR at 78.57% and the Area under the Curve (AUC) has covered 72.88%. However, the Confidence interval 61.19 % - 84.57%, which is wider than, expected. The Cut off values are determined based on Youden optimal Cut off Method.

The TPR = 78.57% and FPR = 0.3953 at Rating 2 has the highest positive rate

Table 1: TPR and FPR rates for the first Investigator

Marker	**Cutpoint**	**FPR**	**TPR**
Rating	-Inf	1.0000	1.0000
Rating	1	1.0000	1.0000
Rating	2	0.3953	0.7857
Rating	3	0.2093	0.5357
Rating	4	0.1395	0.4286
Rating	5	0.0698	0.2500
Rating	Inf	0.0000	0.0000

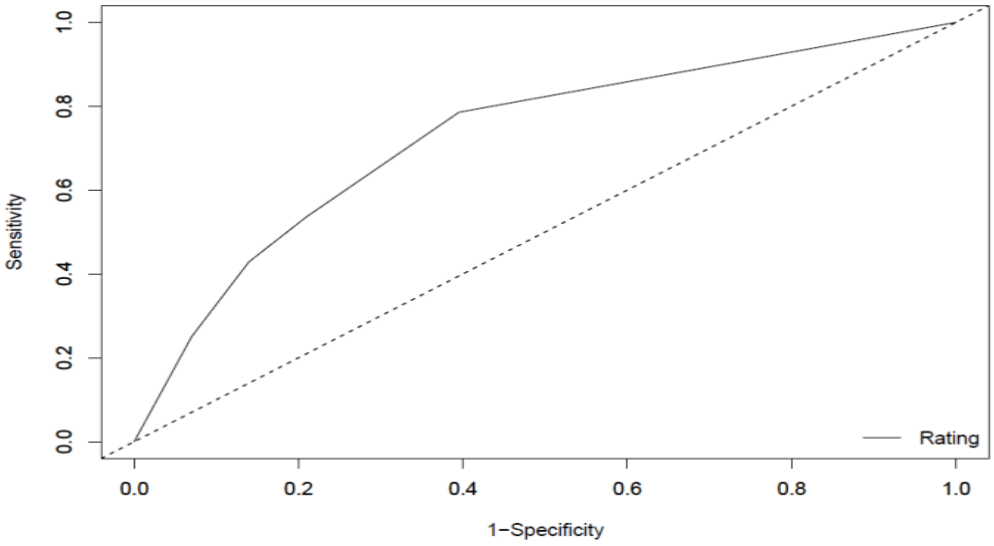
The TPR = 78.57% and FPR = 0.3953 at Rating 2 has the highest positive rate.

Table 1.1 AUC with upper and lower limits

Marker	AUC	SE.AUC	Lower Limit	Upper Limit	Z	p-value
Rating	0.72882	0.05964	0.61193	0.84571	3.83672	0.00012

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XI, November 2025

Figure 1: ROC for the evaluation done by first radiologist



Second Investigator Results are presented below

The ROC curve for the second dataset evaluated by the second radiologist is shown below in Fig 2. In the table below, rating 2 shows the TPR at 71.88% and the Area under the Curve (AUC) has covered 65.70%. However, the Confidence interval 53.35 % - 78.05% which is wider than expected.

Table 2: TPR and FPR rates for the Second Investigator

Marker	Cutpoint	FPR	TPR
Rating2	-Inf	1.0000	1.0000
Rating2	1	1.0000	1.0000
Rating2	2	0.4103	0.7188
Rating2	3	0.2308	0.4688
Rating2	4	0.1795	0.3438
Rating2	5	0.1282	0.1562
Rating2	Inf	0.0000	0.0000

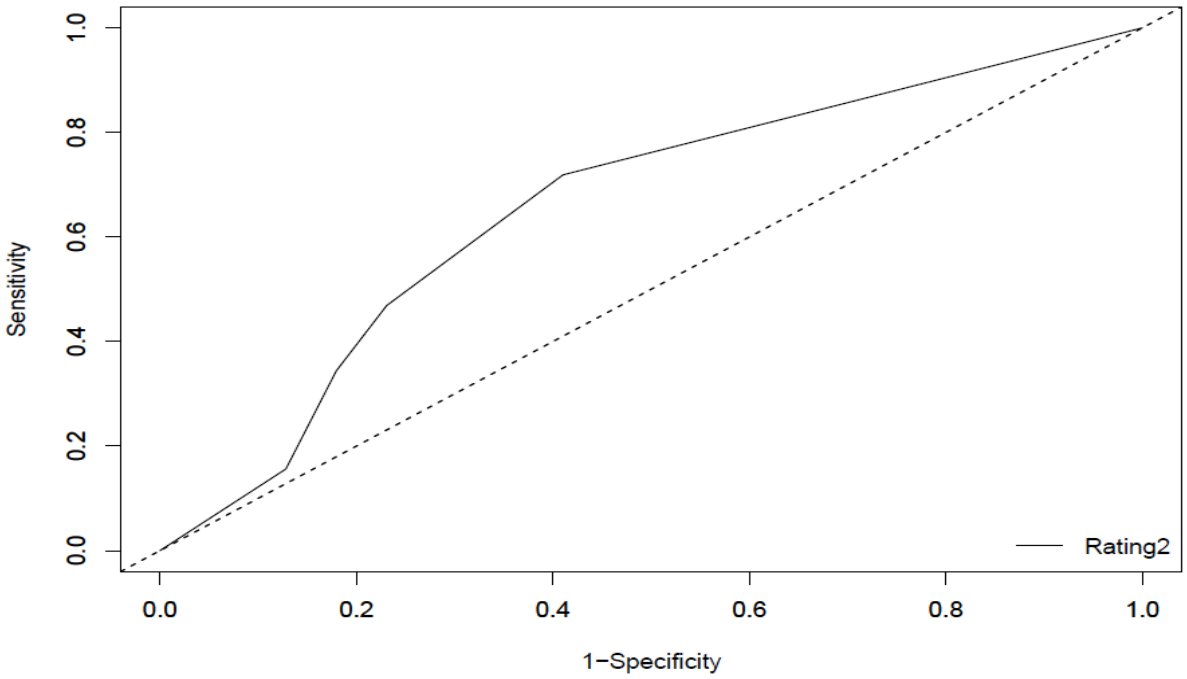
The TPR = 71.88% and FPR = 0.4103 at cutpoint 2 has the highest positive rate.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XI, November 2025

Table 2.1 AUC with upper and lower limits

Marker	AUC	SE.AUC	Lower Limit	Upper Limit	z	p-value
Rating	0.6570513	0.06302699	0.5335207	0.7805819	2.49181	0.0127094

Figure 2: ROC for the evaluation done by second radiologist



After plotting both the individual datasets, comparison of both the datasets are considered in the webtool and it has provided the results below.

Table 3.1 Comparison of the AUC of the both the investigators observations

Marker	AUC	SE.AUC	Lower Limit	Upper Limit	Z	p-value
Rating	0.7116948	0.05609518	0.6017503	0.8216393	3.77385	0.0001607472
Rating2	0.6570513	0.06302699	0.5335207	0.7805819	2.49181	0.0127093971

Table 4.1 Comparing the significance of the ratings and the outcome

Marker1 (I)	Marker2 (J)	AUC(I)	AUC(J)	|I - J|	SE(|I - J|)	z	p-value
Rating	Rating2	0.7117	0.6571	0.0546	0.0844	0.6476	0.5172

The ROC curve after plotting both the datasets for comparison is shown below in Fig 3. In the table above, rating 2 shows the TPR at 84.62% and the Area under the Curve (AUC) has covered 71.16%.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XI, November 2025

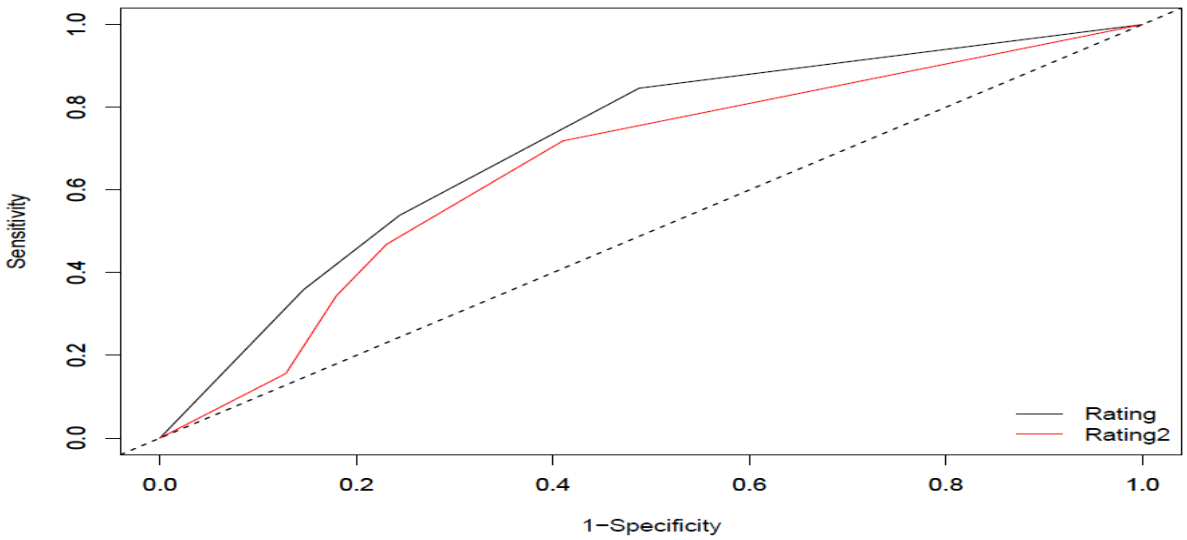


Figure 3: ROC for both the evaluation on a single panel

However, it concludes that there is no difference between both the evaluation patterns and the p value is 0.5172, hence the evaluations and interpretations of both the radiologists concur.

CONCLUSION:

On observing Figure 3, the ROC curves of two evaluation patterns are closer/almost parallel across the pairs of co-ordinates. This sets a mandate scenario to have a hypothetical framework to validate verify the two curves differ to each other or not. For addressing this, a webtool support is taken and outcomes are reported in Table (3.1 and 3.2). From the results, it is evident that two curves do not differ ($p>0.05$) and this can be supported with the difference of AUC's i.e., 0.054, which is almost least value. Now, this can be correlated with Figure 3, that is the distance between co-ordinates of two curves are not that large and has not witnessed a maximum distance at any point. Moreover, an ROC curve can be identified as a better one if it's 1-Speicificity should be lesser than the other. With the present comparing scenario both ROC curves have same pattern of having almost similar 1-Specificity, so this can also be viewed in augmenting the result of p-value that the two curves are having similar efficacy in detecting the lesions

REFERENCES:

1. Swets JA. Indices of discrimination or diagnostic accuracy: their ROCs and implied models. Psychol Bull. 1986;99:100–17. [PubMed]
2. "Logical Analysis in Roentgen Diagnosis": Published in the journal *Radiology*, Key Publications by Lusted in 1960 ; Lusted LB. Logical analysis in roentgen diagnosis. Radiology. 1960;74:178–93. [PubMed]
3. Dorfman, D. D., & Alf, E., Jr. (1968). Maximum likelihood estimation of parameters of signal detection theory—A direct solution. *Psychometrika*, *33*(1), 117–124.
4. Green DM, Swets JA. Signal detection theory and psychophysics. First ed. New York: John Wiley & Sons; 1966.
5. Dorfman DD, Alf EJR. Maximum likelihood estimation of parameters of signal detection theory - a direct solution. Psychometrika. 1968;33:117–24. [PubMed]
6. FORTRAN programs ROCFIT, CORROC2, LABROC1 and LABROC4, ROCKIT. Available at:http://www.radiology.uchicago.edu/krk/KRL_ROC/software_index6.htm.
7. Metz CE. Basic principles of ROC analysis. Semin Nucl Med. 1978;8:283–98. [PubMed]
8. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, SargentD, Ford R, et al. New response evaluation criteria in solidtumours: revised RECIST guideline (version 1.1). *Eur J Cancer*2009;45:228-247
9. The R Journal: article published in 2016, volume 8:2; Dincer Goksuluk, Selcuk Korkmaz, Gokmen Zararsiz and A. Ergun Karaagaoglu , The R Journal (2016) 8:2, pages 213-230.