Using Receiver Operating Characteristic (ROC) Curves to Evaluate Digital Mammography

Julie M. Pickard

University of Texas Health Science Center, San Antonio, Texas, USA

Abstract
Receiver operating characteristic (ROC) curves are frequently used to compare the accuracy of two or more imaging modalities. This paper addresses the use of ROC analysis to evaluate the speed and accuracy of digital mammography, as compared to conventional film-screen mammography.

1 Introduction
In the United States, breast cancer is the second leading cause of death among women [1]. Mammography is one of several screening tools recommended for the early detection of breast cancer [2, 3]. This procedure consists of taking radiological images of a compressed breast. The radiologist reviews the films for abnormal tissue and classifies the diagnosis into discrete categories, i.e. normal (negative) or suspicious (positive).

Even with recent technological advances in digital radiological imaging systems, screen film mammography (SFM) is still the current “gold standard” for breast cancer screening [4]. The advantages of SFM include ease of use, low cost, and acceptable image quality. On the other hand, these “conventional” systems use one medium to detect, display, and store the image. This does not allow image manipulation after acquisition. Digital systems can separate these stages and provide the radiologist with a means to directly manipulate the image. In addition, digital mammography eliminates film-processing time, has a higher dynamic range, and has the potential to improve image quality. However, at this time, the spatial resolution of digital systems is not as high as conventional systems [4, 5]. Therefore, the need to accurately evaluate the benefits of each system from one another exists.

One method used to define the accuracy of an imaging system is to evaluate the system’s sensitivity and specificity. Sensitivity, also called true positive fraction (TPF), is the probability of diagnosing the presence of disease when it is actually present. Specificity, also called true negative fraction (TNF) is the probability of identifying the absence of disease when it is not present. A graphical representation of these factors is called the receiver operating characteristic (ROC) curve.

Originally developed in the 1950’s, the ROC curve was used to decipher radar signals from noise. Today ROC analysis is frequently used to compare the accuracy of two or more imaging modalities. This paper addresses the use of ROC analysis to evaluate the speed and accuracy of digital mammography, as compared to conventional film-screen mammography.

2 Basic Principles of ROC Analysis
When a patient is evaluated for disease by a radiological method, a radiologist reviews the radiographs to determine if the disease is present or not. Even though most radiographs contain detailed information, the diagnostic test result is typically evaluated as either positive or negative. In order to make the diagnosis, the radiologist
has a pre-determined threshold value (or cut-off point) for what a diseased image and non-diseased image looks like. It is important to note that radiologists have considerable variation in how they interpret diseased states from one another and they might use different threshold cut-off values depending upon patient history. For instance, a patient being screened for breast cancer who has breast cancer in her family medical history might be screened as positive and require additional testing if the radiologist notices a typically non-suspicious tissue abnormality on the radiographs. On the other hand, a patient with a similar tissue abnormality that does not have breast cancer in her family medical history might be screened as negative.

The threshold value that the radiologist uses highly influences the sensitivity and specificity of the method. For an ideal model with two Gaussian distributions, as shown in Figure 1 [6], the distribution of sensitivity and specificity do not overlap, thus resulting in a “perfect” diagnostic test. In reality, most diagnostic tests are not perfect, thus leading to gray areas in the diagnosis of diseased and non-diseased states. This uncertainty is represented by an overlap between the distributions.

Fig. 1. Probability distributions of an ideal system. From reference 6.

The influence on sensitivity and specificity by the threshold cut-off value is shown in Figure 2 [6]. A lower threshold cut-off results in a higher sensitivity and lower specificity. In other words, the number of false-negatives decreases and the number of false-positives increases. Inversely, a higher threshold cut-off value results in a lower sensitivity and higher specificity.

The ROC curve is a graphical plot of the true positive rate (along the vertical axis) against 1 minus the false positive rate (along the horizontal axis). ROC curve “comes from the idea that, given the curve, we, the receivers of the information, can use (or operate at) any point on the curve by using the appropriate cut point [7].” Thus, the ROC curve can be used to determine the optimal threshold cut-off value between sensitivity and specificity.

Fig. 2. Probability distributions of a realistic diagnostic system. (a) Shows the influence of the threshold value on sensitivity. (b) Shows the influence of the threshold value on specificity. From reference 6.

3 Using ROC Analysis to Compare Digital and Film Mammography Systems

Since a ROC curve represents a system’s complete sensitivity and specificity range, the graph offers a visual comparison that can be used to compare multiple systems without the influence of a threshold value.

Typically, the area under a ROC curve represents the system’s accuracy. The area can range from 0.0 to 1.0. As shown in Figure 3 [7], a “perfect” system, which does not have overlap
between the two distributions, has an area of 1. A “chance” system, which has complete overlap between the two distributions, has an area of 0.5 [7]. The “chance” curve is the practical limit the system. Most systems have a ROC curve in-between the “chance” and “perfect” curves.

![ROC curves](image)

Fig. 3. A graph showing three ROC curves with different values of area under the curve. The perfect test has an area of 1.0 under the curve. The chance test has an area of 0.5 under the curve. The moderate test lies between the perfect and chance curves. From reference 7.

It is important to note that since the ROC curve contains all of the threshold value combinations, choosing a system based on area alone might be misleading [6-8]. As shown in Figure 4 [7], two ROC curves with the same area can cross over one another. Overall the two systems have the same accuracy, but they are not identical. In order to determine the optimal system from the two, the user would need to know the clinical requirements of the system. For instance, if the clinical system needs a high sensitivity, then the user would pick test system B. Typically, it is best to compare systems in “relevant sensitivity and specificity ranges [6].”

4 ROC Analysis in Digital Mammography

In order to obtain Food and Drug Administration (FDA) approval to market digital mammography systems [9], it is suggested that manufacturers provide a “comparative feature analysis study.” Thus, ROC analysis can be used to show substantial equivalence between digital and film screen mammography systems. Specifically, ROC analysis has been used to compare lesion detectability in phantoms and compare the clinical accuracy of these systems.

![ROC curves](image)

Fig. 4. The graph of crossing ROC curves which have the same area under their curves. From reference 7.

4.1 ROC Analysis to Compare Lesion Detectability in Phantoms

Yip et al. [10] used ROC analysis to compare lesion detectability between conventional spot mammography and digital spot mammography. In this study, four radiologists evaluated one hundred breast phantoms for normal or abnormal regions. Out of the one hundred sample evaluations, twenty-five had phantom lesions randomly placed on top of them. The images (analog and digital) were acquired and read on different days. The images were scored on a scale from 1 (abnormality definitely not preset) to 5 (abnormality definitely present). ROC curves were plotted for all four radiologists and pooled together for comparison (Figure 5). The individual plots showed that three out of the four radiologists “appeared to perform better” with digital images, however these results were determined not to be statistically significant. For all of the ROC evaluations in this study, the authors evaluated the plots by calculating the area under the curves. Overall the results were not statistically significant, however the ROC curves did show that the digital mammography
system was slightly superior to the conventional system.

![ROC curves from pooled data of four radiologists comparing lesion detectability with digital and spot mammography](image1)

**Fig. 5.** ROC curves from pooled data of four radiologists comparing lesion detectability with digital and spot mammography. From reference 10.

### 4.2 ROC Analysis to Compare Clinical Accuracy

Several studies have used ROC analysis to compare the accuracy between full-field digital and conventional mammography in a clinical setting [11-13]. Lewin et al. [11] performed 6736 mammography examinations using both full-field digital and film mammography systems. As shown in Figure 6, the area under the ROC curve for the conventional film system was slightly higher than the digital system, however the results were not statistically significant. Similarly, Kuzmiak et al. [13] compared conventional and digital mammography systems and did not find a statistical difference either. In addition, Kuzmiak et al. used ROC analysis to show intrareader variability between film and digital mammography. The intrareader variability in the area under the ROC curves for film and digital mammography (Figures 7 and 8) were 12% and 6% respectively.

Pisano et al. [14] used ROC analysis to compare the speed and accuracy of digital versus film mammograms. Eight trained radiologists interpreted 63 digitally displayed mammograms versus printed versions. The authors found that the time required to analyze film copies was longer than for digital images. The ROC curve areas and the sensitivity and specificity of each reader were determined not statistically significant. However, the author concluded that overall the sensitivity was slightly better for film display and the specificity was slightly better for digital display. Unfortunately, these authors did not provide the ROC curve plots for this study; therefore, additional interpretation is limited.

![ROC curves for screen-film mammography (dotted line) and full-field digital mammography (solid line)](image2)

**Fig. 6.** ROC curves for screen-film mammography (dotted line) and full-field digital mammography (solid line). From reference 11.

![ROC curves of six readers using film mammography](image3)

**Fig. 7.** ROC curves of six readers using film mammography. From reference 13.

![ROC curves of six readers using digital mammography](image4)

**Fig. 8.** ROC curves of six readers using digital mammography. From reference 13.
5 Limitations

Using ROC analysis to compare conventional and digital mammography systems has become common practice, however, limitations of these comparisons need to be considered.

Even though research to date has not shown that the accuracy of digital mammography systems are statistically better than conventional mammography systems, the variability in the methods used to take the images can create difficulties in accurately comparing them. For instance, ROC analysis cannot take into account the variability of patient exposure during digital versus conventional imaging. Digital imaging has a larger dynamic range than conventional film imaging; therefore, digital imaging has the potential to limit patient exposure. In addition, conventional systems sometimes require exposure at different settings for various regions of the breast.

Second, as mentioned in the background section, researchers should consider the region of comparison used in ROC analysis. All of the studies mentioned in this paper used the total area under of the curve to evaluate each system. Van Erkel and Pattynamna [6] said, “comparing the ROC curves in the relevant sensitivity or specificity ranges is to be preferred over comparing the total area under the curve.” Researchers should consider the cutoff range used in evaluation of the diagnostic images such as detecting breast cancer, thus comparing the sensitivities at a particular false positive range or the partial area under the curve [8].

Lastly, ROC analysis does not take into account all the potential benefits of using digital mammography. ROC analysis does not include the benefits of using electronic media, such as ease of storage and faster transmission and retrieval rates. In addition, ROC analysis does not account for the potential increase in accuracy for typically uninterpretable film cases due to transmission of digital media to several centers for diagnosis.

Therefore, ROC analysis is a great comparison tool, but additional criteria should always be considered when comparing the advantages and disadvantages of different technologies.

6 Conclusion

With the advent of high tech computing systems, digital imaging has come to the forefront of diagnostic imaging. ROC analysis has been used to compare these new imaging modalities, such as digital versus conventional film-screen mammography.

In the last decade, FDA approved several digital mammography systems. Once the use of these digital mammography systems increases, the use of computer assisted detection (CAD) systems will also increase. Studies using ROC analysis to assess the accuracy of these systems has already begun, emphasizing the importance of understanding ROC analysis and its use in digital mammography.
References


