Stereoscopic vs. monoscopic detection of masses on breast tomosynthesis projection images

Gautam S. Muralidhar*[a], Tejaswini Ganapathi[b], Alan C. Bovik[b], Mia K. Markey [a,c], Tamara Miner Haygood [d], Tanya W. Stephens [d], Gary J. Whitman [d]
[a]Dept. of Biomedical Engineering, The University of Texas at Austin, Austin, TX, USA 78712;
[b]Dept. of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, USA 78712;
[c]Dept. of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA 77030;
[d]Dept. of Diagnostic Radiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA 77030

ABSTRACT

The goal of this study was to assess if stereoscopic viewing of breast tomosynthesis projection images impacted mass detection performance when compared to monoscopic viewing. The dataset for this study, provided by Hologic, Inc., contained 47 craniocaudal cases (23 biopsy proven malignant masses and 24 normals). Two projection images that were separated by 8 degrees were chosen to form a stereoscopic pair. The images were preprocessed to enhance their contrast and were presented on a stereoscopic display. Three experienced breast imagers participated in a blinded observer study as readers. Each case was shown twice to each reader – once in the stereoscopic mode, and once in the monoscopic mode in a random order. The readers were asked to make a binary decision on whether they saw a mass for which they would initiate a diagnostic workup or not, and also report the location of the mass and provide a confidence score in the range of 0-100. The binary decisions were analyzed using the sensitivity-specificity measure, while the confidence scores were analyzed using the Receiver Operating Characteristic curve (ROC). We also report a statistical analysis of the difference in partial AUC values greater than 95% sensitivity between the stereoscopic and monoscopic modes.

Keywords: Breast tomosynthesis, stereoscopic display, 3D perception, low dose projections

1. INTRODUCTION

Breast tomosynthesis was recently approved by the U. S. Food and Drug Administration for clinical use. In breast tomosynthesis imaging, 15-30 low dose x-ray projection images of the breast are acquired and multiple slices depicting the 3-dimensional anatomy of the breast are synthesized from these projection images via a reconstruction step [1-3]. The main advantage offered by breast tomosynthesis over traditional mammography is in the reduction of overlapping out-of-plane tissue structures that have been shown to hinder the detection of breast cancer on mammography. Indeed, research and clinical studies have demonstrated the potential of breast tomosynthesis in improving the detection of non-calcified lesions in dense breasts and reducing patient recalls. In the studies that have been conducted so far, breast tomosynthesis cases have been interpreted by viewing the reconstructed slices in either a slice-by-slice or a cine mode [4, 5]. Additionally, the projection images have been reviewed in a way similar to how mammographic images are reviewed in clinical practice [4-6]. The main disadvantage of these two viewing modes is that neither reveals the true 3D structure of the breast. Even though the reconstructed tomosynthesis data provide multiple slices depicting the 3D anatomy, the slices actually provide a quasi-3D view of the anatomy due to the limited number of low dose x-ray projections from which the data are reconstructed. Further, interpreting the projection images is similar to interpreting low-dose mammographic images, and the same problems associated with overlapping out-of-plane tissue structures persist. It is not clear from the studies conducted thus far as to which reading mode is the best for a reliable and efficient interpretation of breast tomosynthesis data. Efficiency of image interpretation is going to be a key factor when

*gautam.s.muralidhar@gmail.com; phone 1 512 471-8660; fax 1 512 471-0616; bme.utexas.edu

Medical Imaging 2012: Image Perception, Observer Performance, and Technology Assessment,
edited by Craig K. Abbey, Claudia R. Mello-Thoms, Proc. of SPIE Vol. 8318, 831807

tomosynthesis is routinely integrated in the clinical workflow since the volume of data that could be generated is potentially very large, thereby placing an additional burden on the ever-increasing workload of the interpreting radiologists.

The imaging geometry of breast tomosynthesis presents a case for stereoscopic (stereo) interpretation of the low-dose tomosynthesis projection images. Multiple x-ray projections are acquired during a tomosynthesis examination with the breast held fixed in its initial position. By selecting two projection images that are separated by a reasonable angle, a stereo pair can be formed that could be viewed with the aid of a stereo display. Stereo acquisition and viewing of mammograms (also called stereo mammography) has been shown previously to make a significant difference in the sensitivity and specificity of breast cancer detection [7]. However, in the most widely publicized stereo mammography configuration [7], the mammograms are acquired at the same x-ray dose as regular mammograms, while tomosynthesis projection images are acquired at a fraction of the x-ray dose used for mammography. Hence, the low-dose tomosynthesis projection images have a poor signal to noise ratio when compared to regular mammograms. However, in a study by Webb et al. [8], it was shown that radiologists detected simulated masses better on the projection images, which were viewed stereoscopically, than on regular mammograms, which were viewed monoscopically. The main purpose of this study was to assess if stereo viewing of breast tomosynthesis projection images impacted mass detection performance when compared to monoscopic (mono) viewing of the projection images.

2. MATERIALS AND METHODS

2.1 Dataset

The dataset for this study, provided by Hologic, Inc. (Bedford, MA), contained 47 craniocaudal cases. There were 23 cases, each containing a single biopsy-proven, malignant mass, and 24 normal cases. Each case was comprised of 15 projection images spanning an angular range of approximately 15 degrees. A stereo pair was formed by selecting two projection images that were separated by approximately ± 4 degrees from the 0 angle projection resulting in a net separation of 8 degrees. The net angle of separation used in our study was consistent with the angle of separation used in previous studies with stereo mammography (6-10 degrees). An experienced breast imager (who we will refer to as the truth radiologist) with more than 20 years of experience verified the truth information for the masses on the projection images of the stereo pair by using the truth information provided by Hologic, Inc. for the reconstructed slices. The truth radiologist certified each stereo pair as having sufficient contrast for interpretation and also graded the breast tissue density using the BI-RADS® categorization: mainly fatty, scattered fibroglandular densities, heterogeneously dense, and extremely dense [9]. Table 1 shows the number of cases according to breast density in the dataset.

Table 1. Number of cases according to breast density category in the dataset.

Number of cases: fatty breasts	Number of cases: breasts with scattered densities	Number of cases: heterogeneously dense breasts	Number of cases: extremely dense breasts
10	21	13	3

The stereo display used in our study was a Planar PL2010M stereo display (maximum resolution 1600x1200) manufactured by Planar Systems, Inc. (Beaverton, OR, USA). This was a research workstation and not a medical grade display. The stereo display was comprised of two monitors mounted on top of one another at an angular separation of about 100-110 degrees. A glass plate with a half-silver coating bisected the angle between the two monitors. The glass plate transmitted the image displayed on the lower monitor, while its top surface reflected the image displayed on the upper monitor. A human with reasonable stereo acuity can fuse the stereo pair by wearing lightweight passive cross-polarized glasses that allow the left eye to see only the reflected image from the upper monitor, and the right eye to see only the transmitted image from the lower monitor. The fused stereo pair then enables perception of depth throughout the image. Figure 1 depicts the display used in our study. The Stereoscopic PlayerTM (3dtv.at, Linz, Austria) software was used to load the stereo pairs on the stereo display.

2.2 Data preprocessing

We manually preprocessed the stereo pairs of the raw projection images to enhance their contrast. This was achieved by manually adjusting the DICOM window width/window level parameter to obtain satisfactory contrast in the two images of the stereo pair. The DICOM window width/window level parameter was set to the same value in both images of the stereo pair. Subsequently, we corrected the histograms of the two images by shifting the median of one histogram with respect to the other such that they were identically matched in shape. To perceive stereoscopic effect, the images of the stereo pair had to be rotated by 90 degrees in the clockwise direction such that the left breast was displayed as though the patient was in a prone position, while the right breast was displayed as though the patient was in a supine position. This display of images, while different from the display of conventional mammographic images, provides horizontal parallax, which in turn induces stereoscopic perception. The stereo display mode used in our study was similar to that used by Getty et al. in their stereo mammography study [7]. For the mono display, the same projection image was loaded on both

Figure 1. Planar PL2010M stereo display used in our study.

the monitors of the stereo display, which provided no horizontal parallax. Finally, a white dot was randomly placed on either the left or the right side of the non-tissue region of the images. The white dot was used as a reference by readers while commenting on the location of the abnormality found on the image. Figure 2 illustrates an example of a processed stereo pair of tomosynthesis projection images of the left breast used in our study. The preprocessing described here could also be automated as shown by Webb et al. [8] in their study.

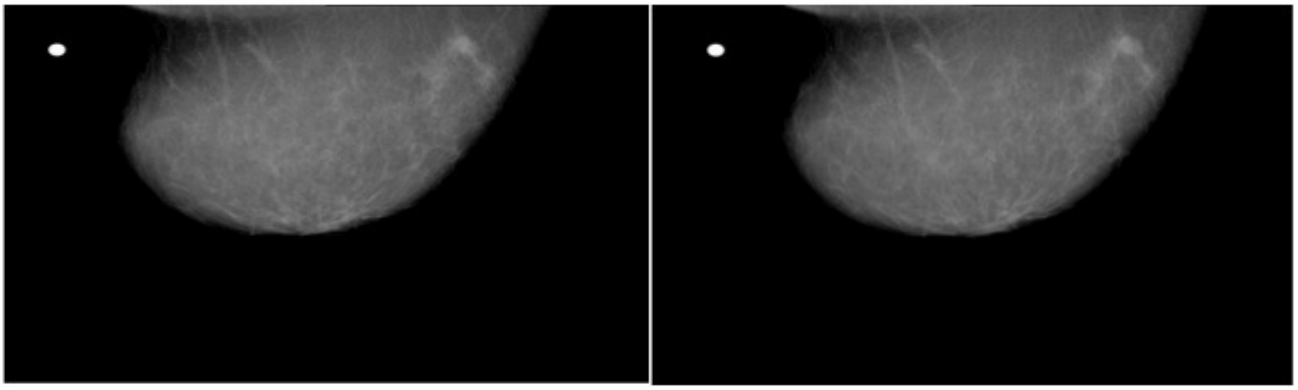

Figure 2. Processed stereo pair of tomosynthesis projection images of the left breast.

2.3 Reader Study

Three experienced breast imagers with a collective experience of more than 25 years in interpreting screening mammograms participated in the study as blinded readers. The Randot stereovision test was administered on each reader (including the truth radiologist) to determine that the readers were not stereo blind. The study was held over two sessions, with a good mix of true positive and true negative cases shown in a random order. A total of 23 cases were shown in one session, and 24 cases in the other session. In each session, the reader was shown each case twice – once in the stereo mode, and once in the mono mode in a random order. The reader was not told what the current viewing mode was. The study sessions were all held under the same ambient lighting conditions and each case was presented on the display till the readers had completed interpreting the case. The cross polarized stereo glasses were used by the reader throughout the session, even while viewing the images monoscopically. Further, the reader was not allowed to modify display parameters such as magnification and contrast.

The readers were asked to provide a binary decision on whether they saw a mass for which they would initiate a diagnostic workup or not, and provide a confidence score in the range of 0-100 that indicated their confidence in the presence of the mass. A confidence score of 0 meant that the readers were 100% certain that there was no mass, while a confidence score of 100 meant that the readers were 100% certain that there was a mass. We also collected the location information of the perceived mass using the breast maps illustrated in Figure 3. However, in this study, we have not used the location information while analyzing the readers' performance, as each abnormal case comprised only a single mass. The binary decisions were collected, as they closely resembled how the readers would operate in an actual clinical setting. The confidence scores, which indicated the reader's confidence in the presence of an abnormality, were collected for analyzing the readers' performance in an experimental setting using the Receiver Operating Characteristic curve (ROC). Previous works (e.g., [10]) on analyzing binary and continuous/multi-category ratings in mammography observer studies have demonstrated that these two decision-making processes yield similar reader performances, even though the binary true positive fraction (TPF) and false positive fraction (FPF) operating points do not always lie on the ROC curves but in their vicinity.

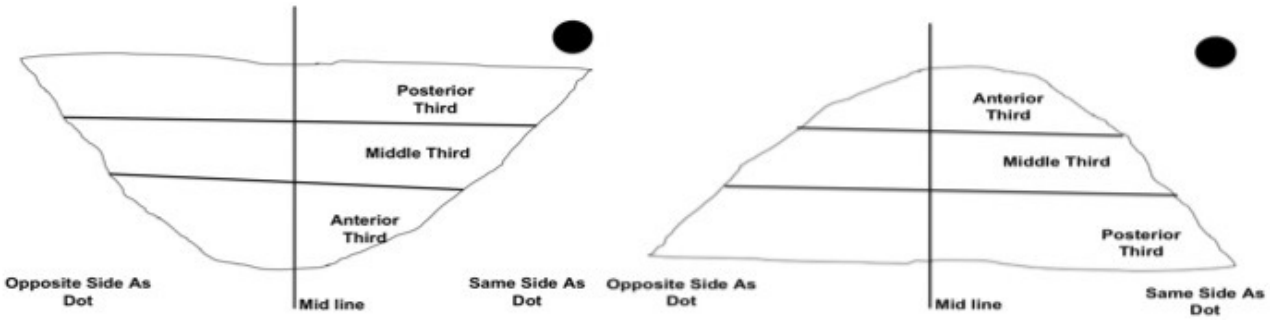

Figure 3. Map of the left breast (left) and right breast (right) used by the readers for indicating the location of the mass.

2.4 Statistical Analysis

From the binary decisions, we computed the overall binary TPF and FPF for each reader and each viewing mode. Similarly, the empirical ROC curves were generated using the confidence scores for each reader and each viewing mode. We first analyzed whether the binary decisions and the continuous ratings resulted in similar reader performances. A bootstrap analysis was carried out to ascertain this. This analysis was similar to what was described by Gur et al. [10] in their work comparing mammography reader performances under binary and continuous/multi-category ratings. We summarize the key steps:

Let us suppose that the i^{th} reader is denoted by R_i and that for this reader, and a given viewing mode (stereo or mono), the binary TPF and FPF obtained were $(TPF_{R_i}^{bin}, FPF_{R_i}^{bin})$. We then evaluated the TPF at the binary FPF, $FPF_{R_i}^{bin}$, from

the corresponding ROC curve. We used linear interpolation for values of $FPF_{R_i}^{bin}$ that were not present in the list of FPF values used to generate the empirical ROC curve. Let us denote the linearly interpolated TPF value from the ROC curve by $TPF_{R_i}^{roc}$. The signed vertical difference $TPF_{R_i}^{bin} - TPF_{R_i}^{roc}$ yields a measure that is indicative of how similar the reader performances are under the binary decision and continuous confidence scores. To asses whether the signed difference in sensitivities was significant or not, we performed bootstrap sampling. For each reader, we separately re-sampled the mass cases and the normal cases independently to ensure that the final sample had the same number of mass and normal cases as in the original dataset. It is important to note that we did not re-sample cases across readers, and each reader's data was analyzed independently of the others. We generated 5000 bootstrap samples, and for each sample the signed vertical difference between the binary and the ROC TPF values were computed as described above. The mean-subtracted bootstrap difference distribution was then used to evaluate the two-sided bootstrap p-value with the test statistic being the signed vertical difference $TPF_{R_i}^{bin} - TPF_{R_i}^{roc}$ computed from the observed data.

The area under the ROC curve (AUC) and the standard deviation in AUC were also computed from the observed data for the stereo and the mono viewing modes. Further, since breast imaging radiologists usually operate at sensitivities greater than 90%, the partial AUC values for 90% and 95% sensitivities were also computed using the pROC software package [11]. The differences in partial AUC values for 90% and 95% sensitivities were statistically assessed [11].

3. RESULTS

Figures 4 and 5 illustrate the ROC curves for the three readers along with the binary operating points for each reader for the mono and the stereo viewing modes, respectively. As can be seen from Figures 4 and 5, the binary operating points do not necessarily lie on the ROC curves. Table 2 lists the bootstrap p-values comparing the reader performance using the binary and continuous rating scales for the mono and the stereo viewing modes. Out of the 6 experimental conditions (3 readers, 2 viewing modes) a statistically significant difference in performance between the binary and continuous ratings was observed for only one condition: the mono viewing mode of reader 2. For the remaining experimental conditions, no significant differences in performance were observed between the binary and the continuous confidence scores. Table 3 lists the AUC values and their standard deviations, the partial AUC values at 90% and 95% sensitivities, and p-values indicating whether the differences in partial AUC values at 90% and 95% sensitivities were significant or not for the mono and the stereo viewing modes. From Table 3, we see that the partial AUC values at 90% and 95% sensitivities were significantly better for readers 1 and 2 under the stereo viewing mode than when compared to the mono viewing mode. For reader 3, the difference in the partial AUC values at 90% ad 95% sensitivities was not statistically significant.

Table 2. Bootstrap p-values (two sided) comparing the difference in performance between the binary decision and continuous confidence scores ($p < 0.05$ indicates statistical significance).

Reader number	Bootstrap p-value (Mono viewing mode)	Bootstrap p-value (Stereo viewing mode)
1	0.133	0.053
2	*0.036*	0.169
3	0.466	0.194

4. CONCLUSIONS

Stereoscopic viewing could benefit the interpretation of breast tomosynthesis cases. The results of our pilot study with three readers suggests that stereo viewing might yield better detection performance, particularly at sensitivities greater than 90%. We also evaluated the reader performance under the two viewing conditions using both the binary decisions and the continuous confidence rating scores that were collected. Our findings suggest that the reader performance was similar when using the binary decisions and the continuous rating scales for 5 out of the 6 experimental conditions. We

are actively recruiting additional readers for this study to demonstrate the potential of stereo viewing for reliable and efficient interpretation of breast tomosynthesis projection images.

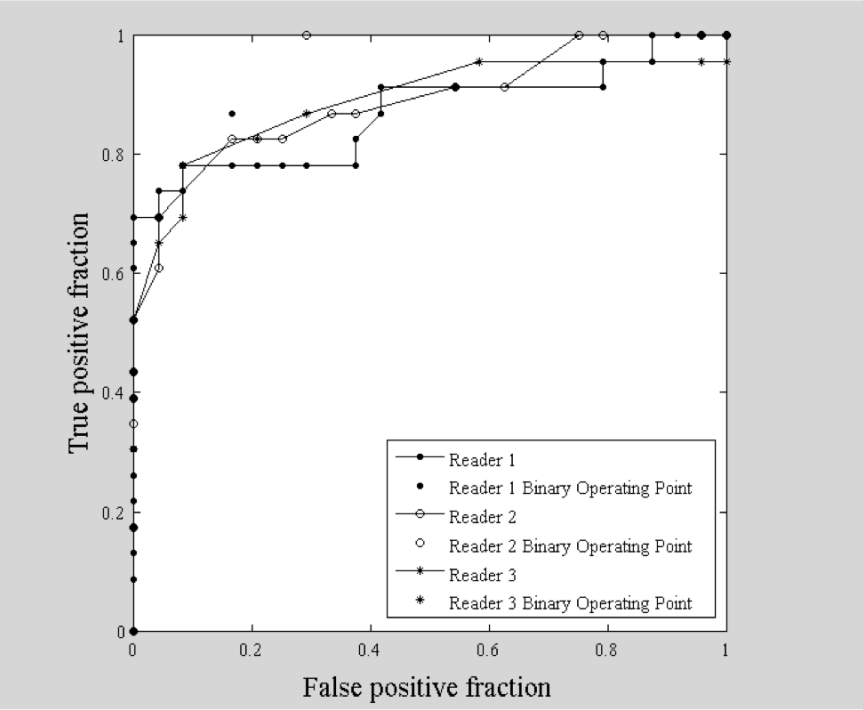

Figure 4. ROC curves and the corresponding binary operating points depicting the performance of the three readers under the mono viewing mode.

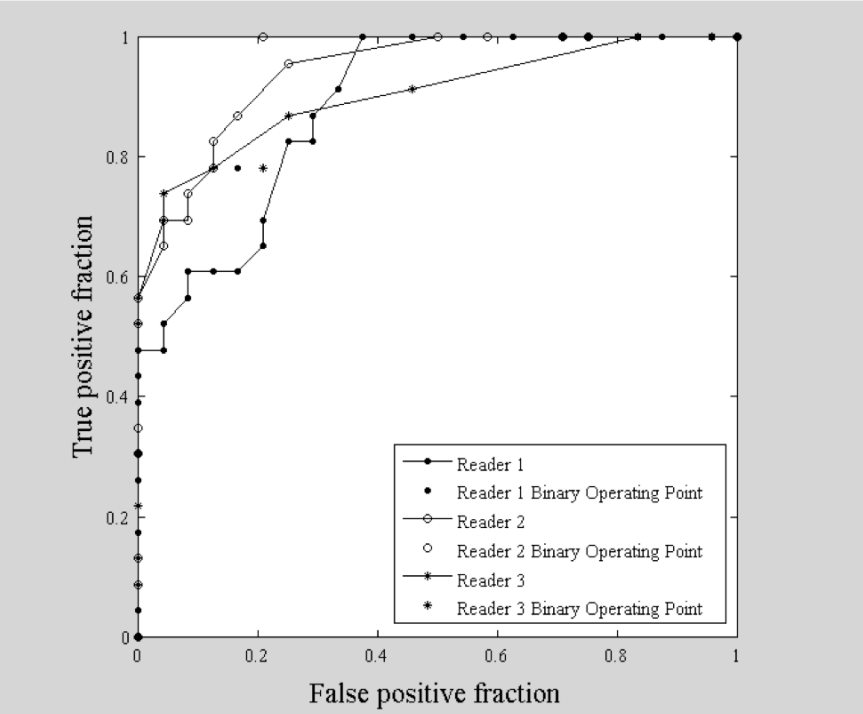

Figure 5. ROC curves and the corresponding binary operating points depicting the performance of the three readers under the stereo viewing mode.

Table 3. AUC and partial AUC evaluated at 90% and 95% TPF along with the results of the statistical test between the partial AUC values obtained under the mono and stereo viewing modes ($p < 0.05$ indicates statistical significance).

Reader number	Mono viewing mode AUC	Stereo viewing mode AUC	Partial AUCs at 90% and 95% TPF (Mono viewing mode)	Partial AUCs at 90% and 95% TPF (Stereo viewing mode)	p-value comparing partial AUC at 90% TPF between mono and stereo viewing modes	p-value comparing partial AUC at 95 % TPF between mono and stereo viewing modes
1	0.870±0.05	0.887±0.04	0.022, 0.006	0.064, 0.031	*0.037*	0.013
2	0.888±0.04	0.942±0.03	0.033, 0.014	0.071, 0.032	0.026	0.037
3	0.889±0.05	0.904±0.04	0.028, 0.0002	0.038, 0.013	0.636	0.302

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of Hologic, Inc. (Bedford, MA, USA) in providing data for this project. The authors would also like to acknowledge Dr. Loren Niklason and Dr. Ashwini Kshirsagar at Hologic, Inc. for providing assistance with the breast tomosynthesis data.

REFERENCES

[1] Niklason, L. T., Christian, B. T., Niklason, L. E., et al., "Digital tomosynthesis in breast imaging," Radiology, 205(2), 399-406 (1997).
[2] Dobbins, J. T. and Godfrey, D. J., "Digital x-ray tomosynthesis: current state of the art and clinical potential, " Physics in Medicine and Biology, 48(19), R65-106 (2003).
[3] Ren, B., Ruth, C., Wu, T., et al., "A new generation FFDM / tomosynthesis fusion system with selenium detector," Proc. SPIE. 7622, 76220B-76220B-11 (2010).
[4] Poplack S. P., Tosteson T. D., Kogel, C. A., et al., "Digital breast tomosynthesis: initial experience in 98 women with abnormal digital screening mammography," American Journal of Roentgenology, 189(3), 616-623 (2007).
[5] Gur, D., Abrams, G. S., Chough, D. M., et al., "Digital breast tomosynthesis: observer performance study," American Journal of Roentgenology, 193(2), 586-591 (2009).
[6] Good, W. F., Abrams, G. S., Catullo, V. J., et al., "Digital breast tomosynthesis: A pilot observer study," American Journal of Roentgenology, 190(4), 865-869 (2008).
[7] Getty, D. J., D'Orsi, C. J. and Pickett, R. M., "Stereoscopic digital mammography: Improved accuracy of lesion detection in breast cancer screening," Lecture Notes in Computer Science, 5116, 74-79 (2008).
[8] Webb, L. J., Samei, E., Lo, J. Y., et al., "Comparative performance of multiview stereoscopic and mammographic display modalities for breast lesion detection," Medical Physics, 38(4), 1972-1980 (2011).
[9] D'Orsi, C. J., Bassett, L. W., Berg, W. A., et al., [BI-RADS: Mammography] American College of Radiology, Reston (VA)(2003).
[10] Gur, D., Bandos, A. I., King, J. L., et al., "Binary and multi-category ratings in a laboratory observer performance study: a comparison," Medical Physics, 35(10), 4404-9 (2008).
[11] Robin, X., Turck, N., Hainard, A., et al., "pROC: an open-source package for R and S+ to analyze and compare ROC curves," BMC Bioinformatics, 12, 77 (2011).