Three Digital Mammography Display Configurations: Observer Performance in a Pilot ROC Study

László K. Tabár[1], Alfredo S. Niño[2], Eric Schreiber[3], Richard Ellis[4], Cynthia Goralnik[5], Steven Scislaw[6], Raja Cheruvu[7], Deepa Masrani[8], Daoxian Zhang[9,*], Steve Muller[10], and Patrick Heffernan[9]

[1] School of Medicine, University of Uppsala, Sweden
[2] Banner Good Samaritan Medical Center, AZ, USA
[3] Cleveland Clinic Foundation, OH, USA
[4] Gundersen Lutheran Medical Center, WI, USA
[5] SimonMed Imaging, AZ, USA
[6] Saint Agnes Medical Center and Physicians Diagnostic Imaging, CA, USA
[7] Williamsville, NY, USA
[8] The University of Texas Medical Branch, TX, USA
[9] Three Palm Software, CA, USA
heidiz@threepalmsoft.com
[10] NDS Surgical Imaging, MA, USA

Abstract. The purpose of this study was to determine if displays that provide more grayscale levels (10bit vs. 8bit) can improve observer performance in breast cancer detection. The study was also designed to determine if 3MP (million-pixel) displays can achieve similar observer performance as compared to 5MP displays. The study was performed using the WorkstationOne mammography software on Dome® E5 and E3 high-resolution displays. Ten radiologists reviewed 33 digital mammography screening studies. On 5MP displays, the average Az value across all observers was 0.7912 using 8bit displays, and 0.8306 using 10bit displays. The difference between 8bit and 10bit displays is statistically significant (F=4.43, p=0.0157). The difference of the average Az value from 8bit 3MP displays is not statistically significant compared to reading with 5MP displays. The implications of the study are that, using appropriate software and hardware, 5MP 10bit displays may improve diagnostic accuracy and 3MP displays may not impact negatively diagnostic accuracy.

Keywords: Digital Mammography, Medical Display, Observer Study.

1 Introduction

Most evidence suggests that the human visual system is able to perceive around 1000 grayscale levels over the luminance range currently used in medical displays [1-4]. However an 8bit display can only provide 256 grayscale levels, whereas most digital mammography acquisition devices (such as FFDM or Mammography CR) produce images at higher bit depths, ranging from 10 to 16 bits. This means that all the acquired

* Corresponding author.

J. Martí et al. (Eds.): IWDM 2010, LNCS 6136, pp. 280–287, 2010.

grayscale levels cannot be displayed at once, or at least not displayed at the contrast sensitivity that human observers can utilize. This could result in a potential loss of information during the diagnostic interpretation, with the implication that the observer may require frequent changes to a localized contrast window and level in order to see all the information, which would impact the observer's efficiency.

This study assessed whether human observer perceptual performance would improve with 10bit 5MP (million-pixel) displays compared to 8bit 5MP displays. This study also assessed whether 3MP displays can achieve similar observer performance to 5MP displays. The reading time was also recorded in order to determine if the three different displays have an impact on the efficiency of the reading workflow.

2 Methods

The study used the WorkstationOne software (Three Palm Software, Los Gatos, CA) on the Dome E5 5MP and E3 3MP displays (NDS Surgical Imaging, San Jose, CA) driven by FX4600 graphic cards (NVIDIA, Santa Clara, CA). Dome E5 and E3 displays are FDA-cleared high-resolution grayscale monitors for displaying medical images (Dome E5 is also FDA-cleared for Mammography use). All displays were calibrated to the DICOM GSDF requirements. WorkstationOne has FDA clearance for use with digital mammography systems to interpret digital mammography images by radiologists. WorkstationOne was designed to maximize the radiologists' efficiency as well as accuracy in reading digital mammograms, specifically; the software is designed to support a mammography specific interpretation workflow. With this feature, the radiologist's reading performance is expected to be similar between 3MP displays and 5MP displays. By providing higher pixel grayscale levels, the radiologists' performance is expected to be improved with 10bit displays.

A set of 33 digital mammography cases were randomly selected from an existing database. Among the 33 cases, 16 cases contain mammography visible cancer lesions confirmed by biopsy reports. Each case consists of the standard four view mammogram images. The pixel depth of the images was 10bit or 12bit, and the pixel size of the images was 50µm, 70µm or 94µm. All patient identifiers were removed from the images.

The study recruited 10 observers, including 9 certified radiologists and 1 fellow. The average digital experience was 16.6 months with a range from 0 months to 4 years. The years reading mammograms were from 8 months to 23 years. All participates were voluntary and were from the teaching course "Multimodality Detection and Diagnosis of Breast Diseases". A written informed consent form was provided and signed by each subject before the study began. Participation in this study posed no risks. There is no foreseen benefit to the individuals in this study, and the subjects were not financially compensated.

The WorkstationOne software was downloaded to 6 computers, each with 2 Dome grayscale displays. Among the 6 computers, 2 were setup with 8bit 3MP displays, 2 with 8bit 5MP displays and 2 with 10bit 5MP displays. The study was conducted in a dark room associated with the Mammography Education course held January 21-24, 2009 at the Phoenix convention center.

Reading of the cases was divided into three sessions. In each session, each observer read all 33 cases using the WorkstationOne software. The observers were blinded to the reading results for all three sessions. For session A, the observer used 2 Dome E3 displays (8bit-3MP session). For session B, the observer used 2 Dome E5 8bit displays (8bit-5MP session). For session C, the observer used 2 Dome E5 10bit displays (10bit-5MP session). The order of these sessions was random for each observer. For each case, the observers marked up any lesion that they found, indicated the type of lesion, and ranked the lesion's suspicion (quasi-continuous) level from 0-100%. 10 cases (5 cancers and 5 normal cases) were provided for training before the study began. The similar cancer and normal case distribution for the 33 cases was known to the observers.

A reading methodology was followed on WorkstationOne, which includes the following viewing workflow steps:

1. Overview viewing to display the standard four views of a mammography case from both current and prior (if exists) cases;
2. Bilateral current viewing with same size fit to the display size;
3. Current and prior (if exists) comparison to enhance the detection of tissue density changes;
4. Systematic comparison of left and right breasts using masking [5] to enhance the detection of structural asymmetries;
5. All-pixel viewing of full resolution images using automatic tracking of the viewing path to ensure that there are no areas in the images that are not viewed;
6. Report of interpretation findings.

The study used a standard methodology for multi-reader multi-case (MRMC) receiver operating characteristic (ROC) observer studies [6-12] with a sequential reading model. The MRMC ROC analysis software (DBM MRMC v2.2) from the Kurt Rossmann Laboratories at the University of Chicago was used to calculate ROC curves and the area under the curve (Az value). The software also provided statistical analysis to compare the Az values between the 8bit 3MP group and the 8bit 5MP group; as well as between the 8bit 5MP group and the 10bit 5MP group. The sensitivity and specificity for each observer were also computed. The t-test was used to compare these three measures pooled over the observers for each pair of reading conditions.

3 Results

The study data pooled from all three display configurations were analyzed using DBM MRMC (v2.2) software (see previous section). The program employed jackknifing and ANOVA (Analysis of Variance) techniques [6-12]. The analysis was reported by treating both observers and cases as random samples, i.e., results apply to the observer and case populations. The null hypothesis of equal "treatments" (display configurations) is tested, and the treatments difference 95% confidence intervals are given.

3.1 ROC Area under the Curve (A_z) Analysis

The data collected from 3 "treatments" (display configurations), 10 readers (observers) and 33 cases (17 normal and 16 abnormal mammogram studies) are loaded into the program DBM MRMC for analysis. The curve fitting methodology was PROPROC [6-12]. A graph of the ROC areas under the curve (Az values) for each observer and for each display configuration is shown in the following figure.

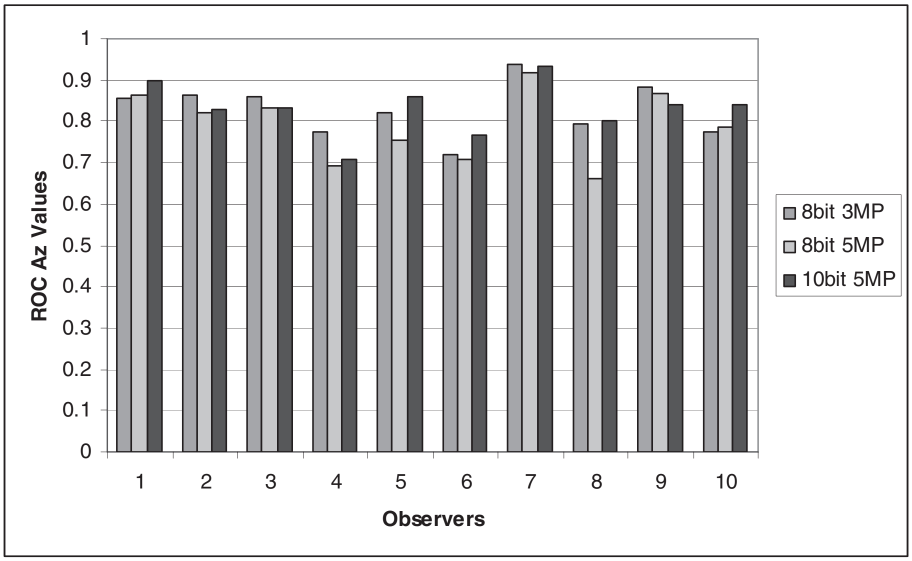

Fig. 1. The Az values for each observer and for each display configuration

The mean values (average across observers) are shown in the following table.

Table 1. The mean values across all observers for each display configuration

Display	Mean Az
8bit 3MP	0.82779296
8bit 5MP	0.79116570
10bit 5MP	0.83063642

The Az values of the three display configurations are not equal, $F(2,18) = 4.43$, $p = 0.0273$. The 95% confidence intervals for the difference between 8bit 3MP and 8bit 5MP is not significant, $p = 0.0235$ (> *0.0167*). The 95% confidence intervals for the difference between 8bit 3MP and 10bit 5MP is not significant, $p = 0.8498$ (> *0.0167*).

The 95% confidence intervals for the difference between 8bit 5MP and 10bit 5MP is significant, p = 0.0157 (< *0.0167*).

Since the study made comparisons among 3 configurations, the critical p-value (0.05) was adjusted by dividing by 3, so the critical p-value *0.0167* was used.

3.2 Sensitivity Analysis

The sensitivity analysis was also performed at specificity = 0.5 (considering around 50% of normal population of the study data). The mean values (average across observers) are shown in the following table.

Table 2. The mean sensitivities across all observers for each display configuration

Display	Average Sensitivity
8bit 3MP	0.88977372
8bit 5MP	0.83981452
10bit 5MP	0.89634391

The sensitivities of the three display configurations are not equal, $F(2,50) = 4.27$, $p = 0.0194$. The 95% confidence intervals for the difference between 8bit 3MP and 8bit 5MP is not significant ($p = 0.0222$). The 95% confidence intervals for the difference between 8bit 3MP and 10bit 5MP is not significant ($p = 0.7574$). The 95% confidence intervals for the difference between 8bit 5MP and 10bit 5MP is significant ($p = 0.0102$).

3.3 Sequential Reading Analysis

An ROC analysis was also performed for the sessions to determine whether 3 sequential readings in three days would change observer performance. The display configuration was randomly assigned to each session for each observer based on the equipment availability.

Table 3. The mean Az values across all observers for each reading session

Session	Average Az
first	0.83393918
second	0.82404307
third	0.81189700

The Az values of the three sessions are not statistically significantly different, $F(2,62) = 0.74$, $p = 0.4832$. So the session arrangement of 3 sequential readings in three days may not change observer performance.

The reading time was automatically recorded during the study, and the average reading time for the 33 cases is shown in the following table.

Table 4. Reading time for each configuration

Display	Average Time (minutes)
8bit 3MP	34.3
8bit 5MP	31.4
10bit 5MP	35.9

The reading time on 8bit 3MP and 10bit 5MP were longer than 8bit 5MP. But analysis was not performed to determine if the differences between these times are statistically significant.

4 Conclusions

Observer average performance was better on 10bit 5MP displays compared to 8bit 5MP displays, and the difference was statistically significant. Observers also performed at higher detection sensitivity (6% better) on 10bit displays.

Observer performance difference was not statistically significant on 3MP displays compared to 8bit 5MP displays.

The sequential sessions of repeat reading three times of same cases in three days did not change observer performance.

5 Clinical Relevance/Application

The study results suggested that 10bit displays may improve readers' performance for mammography interpretation. If the current recommended 5MP displays are not practical, 3MP displays may provide comparable performance with appropriate software.

6 Discussion

This study results showed that 10bit displays are better than 8bit displays. However another published study (Krupinski chest nodule study) showed no difference [2]. The unique (subtle) image characteristics and features on mammography images and the perceptual reading methodology [5] for mammography may be the factors that enable 10bit displays to improve readers' performance.

3MP displays performed slightly better than 5MP displays, which could be attributed to the perceptual reading methodology [5] provided on WorkstationOne which includes the Tabar masking viewing and all-pixel viewing techniques.

Readers spent slightly more time on 8bit 3MP displays than 8bit 5MP displays (and more time on 10bit 5MP displays than 8bit 3MP displays) which may have contributed to the slightly better performance. More displayed image pixel information (10bit vs. 8bit in depth) might have made the observers involuntarily spend more time to read each case; and more time was required for the observers to scan through all pixels on the lower resolution displays (3MP vs. 5MP), which might explain why observers' performance improved.

7 Limitations

This study used only 33 cases with a variety of image pixel bit depth, a small number which can be a potential bias or imprecision when generalizing the study result. A larger number of cases (>100 cases) with the detailed lesion information (such as lesion type, size and subtlety) and tissue density assessment (especially for normal cases) is planned to be used for a follow-on study.

Validity of whether the reading environments are consistent and whether the time interval between each reading session should be longer or shorter is debatable and needs to be validated.

The study tools (software and hardware) were limited to a small set of manufacturers. Thus it is unknown if the study results apply in general (i.e., to any software and hardware).

References

1. Barten, P.G.J.: Contrast sensitivity of the human eye and its effects on image quality. SPIE Press, Bellingham (1999)
2. Krupinski, E.A., Siddiqui, K., Siegel, E., et al.: Influence of 8-bit vs 11-bit digital displays on observer performance and visual search: a multi-center evaluation. Journal of the Society for Information Display 15, 335–429 (2007)
3. Kimpe, T., Tuytschaever, T.: Increasing the Number of Gray Shades in Medical Display Systems - How Much is Enough? Journal of Digital Imaging 20(4), 422–432 (2007)
4. Samei, E., Ranger, N.T., Delong, D.M.: A Comparative Contrast-Detail Study of Five Medical Displays. Med. Phys. 35(4), 1358–1364 (2008)
5. Tabar, L., et al.: Breast Cancer: The Art and Science of Early Detection with Mammography – Perception, Interpretation, Histopathologic Correlation. Georg Thieme Verlag, New York (2005)
6. Dorfman, D.D., Berbaum, K.S., Metz, C.E.: Receiver operating characteristic rating analysis: Generalization to the population of observers and patients with the jackknife method. Investigative Radiology 27, 723–731 (1992)
7. Dorfman, D.D., Berbaum, K.S., Lenth, R.V., Chen, Y.F., Donaghy, B.A.: Monte Carlo validation of a multiobserver method for receiver operating characteristic discrete rating data: Factorial experimental design. Academic Radiology 5, 591–602 (1998)
8. Hillis, S.L., Berbaum, K.S.: Power estimation for the Dorfman-Berbaum-Metz method. Academic Radiology 11, 1260–1273 (2004)
9. Hillis, S.L., Obuchowski, N.A., Schartz, K.M., Berbaum, K.S.: A comparison of the Dorfman-Berbaum-Metz and Obuchowski-Rockette methods for receiver operating characteristic (ROC) data. Statistics in Medicine 24, 1579–1607 (2005), doi:10.1002/sim.2024

10. Hillis, S.L.: Monte Carlo validation of the Dorfman-Berbaum-Metz method using normalized pseudovalues and less data-based model simplification. Academic Radiology 12, 1534–1541 (2005), doi:10.1016/j.acra.2005.07.012
11. Hillis, S.L.: A comparison of denominator degrees of freedom for multiple observer ROC analysis. Statistics in Medicine 26, 596–619 (2007), doi:10.1002/sim.2532
12. Hillis, S.L., Berbaum, K.S., Metz, C.E.: Recent developments in the Dorfman-Berbaum-Metz procedure for multireader ROC study analysis. Academic Radiology 15, 647–661 (2008)