Eur Radiol (2017) 27:5049–5055
DOI 10.1007/s00330-017-4923-8

COMPUTER APPLICATIONS

Comparison of medical-grade and calibrated consumer-grade displays for diagnosis of subtle bone fissures

Daniel Pinto dos Santos[1,2] **· Jonas Welter**[1] **· Tilman Emrich**[1] **· Florian Jungmann**[1] **· Evelyn Dappa**[1] **· Peter Mildenberger**[1] **· Roman Kloeckner**[1]

Received: 19 December 2016 / Revised: 26 May 2017 / Accepted: 1 June 2017 / Published online: 28 June 2017

Abstract
Objective To compare the diagnostic accuracy of medical-grade and calibrated consumer-grade digital displays for the detection of subtle bone fissures.
Methods Three experienced radiologists assessed 96 digital radiographs, 40 without and 56 with subtle bone fissures, for the presence or absence of fissures in various bones using one consumer-grade and two medical-grade displays calibrated according to the DICOM-Grayscale Standard Display Function. The reference standard was consensus reading. Subjective image quality was also assessed by the three readers. Statistical analysis was performed using receiver operating characteristic analysis and by calculating the sensitivity, specificity, and Youden's J for each combination of reader and display. Cohen's unweighted kappa was calculated to assess inter-rater agreement. Subjective image quality was compared using the Wilcoxon signed-rank test.
Results No significant differences were found for the assessment of subjective image quality. Diagnostic performance was similar across all readers and displays, with Youden's J ranging from 0.443 to 0.661. The differences were influenced more by the reader than by the display used for the assessment.
Conclusion No significant differences were found between medical-grade and calibrated consumer-grade displays with regard to their diagnostic performance in assessing subtle bone fissures. Calibrated consumer-grade displays may be sufficient for most radiological examinations.
Key points
• *Diagnostic performance of calibrated consumer-grade displays is comparable to medical-grade displays.*
• *There is no significant difference with regard to subjective image quality.*
• *Use of calibrated consumer-grade displays could cut display costs by 60-80%.*

Keywords Quality Assurance, Health Care · Radiography · Fractures, Bone · Diagnostic Imaging · Computer Terminals

Abbreviations and acronyms

CR	computed radiography
DICOM	Digital Imaging and Communication in Medicine
DR	direct radiography
GSDF	Grayscale Standard Display Function
LCD	liquid crystal displays
PACS	picture archiving and communication systems
ROC	receiver operating characteristic
LED	light-emitting diode
IPS	in-plane switching
px	Pixel
cd	Candela
AUC	area under the curve

Introduction

As radiology has moved from film to a fully digitalised workflow using picture archiving and communication systems

✉ Daniel Pinto dos Santos
daniel.pinto-dos-santos@uk-koeln.de

[1] Department of Diagnostic and Interventional Radiology, University Medical Centre of the Johannes Gutenberg-University Mainz, Mainz, Germany

[2] Department of Radiology, University Hospital Cologne, Köln, Germany

(PACS) and dedicated workstations, the display monitor quality has become a relevant concern. Display quality has dramatically improved over time with the advancement from cathode-ray tube monitors to modern liquid crystal displays (LCDs). In the past, cathode-ray tube monitors that were suitable for radiological image reading were expensive and highly specialised pieces of hardware designed to fulfil the specific requirements of medical use. Displays that are suitable for radiological image interpretation require high luminance values, contrast ratios, resolution and low veiling glare [1]. This ensures that the display has sufficient grey-level differentiation for subtle lesions to be visible to the radiologist. The Digital Imaging and Communication in Medicine (DICOM) Grayscale Standard Display Function (GSDF) was developed in order to standardise the appearance of medical images on displays [2].

In general, modern LCD displays should be suitable for reading images, although national legislation may prevent consumer-grade displays from being used routinely for radiological image reading. For example, the German Industry Standard (DIN) 6868-157 [3] and the FDA 510(k) standard [4] require very specific testing of viewing devices before they can be used for medical applications. Accordingly, the distributors of medical-grade displays base their business models on providing tested hardware that fulfils such requirements, even though consumer-grade displays may be of comparable quality. Medical-grade displays represent a major expense, being about five-fold more expensive than comparable consumer-grade displays. Various studies seem to show that this difference in price does not reflect important differences in diagnostic performance.

Recently, software packages have been developed that are able to calibrate and test consumer-grade displays, effectively turning them into medical-grade displays. One display/calibration software combination was recently approved by the FDA, opening the market to other similar combinations. However, some argue that image perception and diagnostic performance still depend on the display used for radiological image interpretation due to more subtle underlying technical differences. Therefore, the aim of this study was to evaluate and compare the diagnostic performance as well as the subjective image quality of two medical-grade displays and one calibrated consumer-grade display.

Materials and methods

Image acquisition and selection

All digital radiographs were acquired with various computed radiography (CR) or direct radiography (DR) systems according to the institution's standard operating procedures. The images were digitally archived with lossless compression in the departments' PACS and were retrieved at the time of display. Both the image archive and viewer were provided by the same vendor (SECTRA AB, Linköping, Sweden).

All images were taken as part of routine clinical practice, and no additional images were obtained specifically for the purposes of this study. The need for ethical approval was waived by the institutional review board. During the initial phase of the study, an experienced independent radiologist identified consecutive cases that were referred from the institution's trauma department in which subtle bone fissures were either deemed present or highly suspected. A similar number of cases from the same department that lacked visual abnormalities were included for comparison.

Displays

Three displays were used for the evaluation of the radiographs (Table 1). Two displays were medical-grade colour displays; one was a 4 megapixel (MP) EIZO RX440 display (list price 8,300€; EIZO, Hakusan, Japan), and another was a 6MP EIZO RX650 display (list price 10,290€; EIZO, Hakusan, Japan). The other display was a 4MP DELL U3014 consumer-grade colour display (list price 1,100€; DELL, Round Rock, TX, USA). To avoid bias, all monitors were the same size (30 inches/76 cm) and were covered with paper shields, with only the display visible to the readers (Fig. 1). All three monitors were connected to identical computers (DELL Precision T1700 with a professional graphics board, Nvidia Quadro K2200). Both medical-grade displays had built-in calibration according to the DICOM-GSDF, whereas the consumer-grade display had been calibrated to the DICOM-GSDF using a separate tool (PerfectLum, Qubyx, Wilmington, DE, USA; the list price for the PerfectLum Server is around 600€ plus an additional 400€ per monitor). All displays were brand new and had not been used prior to calibration. The display brightness was set to 260 cd/m^2 for all monitors using the respective calibration tools.

Image reading and consensus reading

Three experienced radiologists, each with 5–7 years of experience in musculoskeletal imaging, were recruited for this study. The radiologists were blinded to each patient's characteristics, conditions, medical history, and radiological findings as well as to the proportion of images with and without pathological findings. Image reading was performed under standardise conditions in dim light suitable for radiological image reading (ambient lighting <50 lx indicated by a dedicated photometer [3]). Reading was performed separately by each reader for each monitor. A different order of monitors was randomly assigned to each reader. To avoid recall bias, each reader evaluated the images in a randomized order on a single monitor per session with two weeks between each reading

Table 1 Technical specifications and approximate prices of the three displays used in this study

	EIZO RadiForce RX440	EIZO RadiForce RX650	DELL UltraSharp U3014
Technology	LED backlight / IPS	LED backlight / IPS	LED backlight / IPS
Diagonal size	76 cm	76 cm	76 cm
Resolution	2560 x 1600 px	3280 x 2048 px	2560 x 1600 px
Pixel pitch	0.25 mm	0.197 mm	0.25 mm
Colour depth	1.07 billion	1.07 billion	1.07 billion
Max. brightness	750 cd/m^2	800 cd/m^2	350 cd/m^2
Set brightness	260 cd/m^2	260 cd/m^2	260 cd/m^2
Contrast ratio	1000:1	1000:1	1000:1
DICOM-GSDF	Yes (built-in)	Yes (built-in)	Yes (with PerfectLum)
Price	Around 8,300€	Around 10,290€	Around 1,100€
			(also required: PerfectLum Server around 600€ plus around 400€ per monitor)

LED light-emitting diode, *IPS* in-plane switching, *DICOM-GSDF* Digital Imaging and Communication in Medicine-Grayscale Standard Display Function

session. Readers were allowed a maximum of 1 min per displayed image. They were asked to state whether or not a bone fissure was present and to give their impression of the image quality using a 5-point Likert scale (1 = excellent quality to 5 = insufficient quality) [5]. The radiologists' answers were recorded by independent student assistants.

After all blinded evaluation sessions were conducted, a separate consensus reading was performed where all three blinded readers as well as the independent radiologist that selected the cases took part. In this consensus reading session, the radiologists determined whether or not a bone fissure could be seen in the presented image (Fig. 2). Each image was discussed until a final unanimous verdict was reached, which then served as the reference standard for the following statistical analysis.

Statistical analysis

In preparation for this study, we consulted an independent statistician. The data were registered and stored using Excel 2016 (Microsoft, Redmond, WA, USA). Statistical analysis was carried out using SPSS version 22 (IBM, Armonk, NY, USA). Sensitivity and specificity were calculated separately for each reader/display combination. Youden's index was used to further compare diagnostic performance [6]. Receiver operating characteristic (ROC) analysis was performed to visualise potential differences in sensitivity and specificity. Cohen's unweighted kappa statistics were calculated for the comparison of the different readers and monitors to the consensus reference. Cohen's unweighted kappa statistics were interpreted as follows: 0.00–0.20 = slight agreement, 0.21–

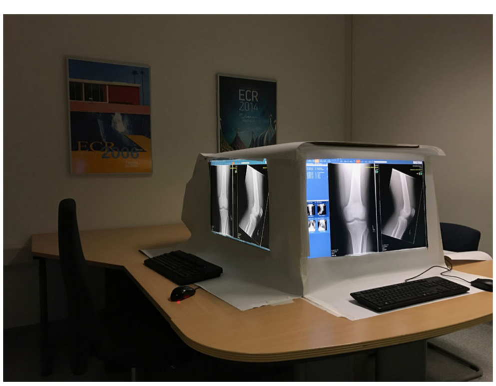

Fig. 1 Study setup in bright light. The monitors were covered with paper shields in order to allow for blinded evaluation. Ambient light was dimmed for image reading

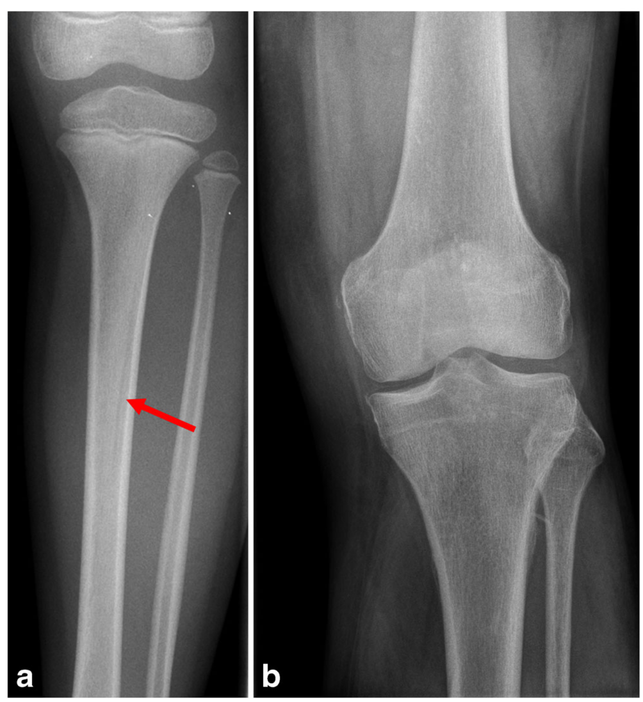

Fig. 2 Examples of radiographs (a) with and (b) without bone fissures

Table 2 Results of the consensus reading for all included images (n = 96 images)

	No fissure	Fissure present	Total
Head (nasal bone)	0	2	2
Arm	1	12	13
Hand / Wrist	11	13	24
Finger	9	9	18
Leg	7	6	13
Foot / Ankle	11	9	20
Toe	1	5	6

0.40 = fair agreement, 0.41–0.60 = moderate agreement, 0.61–0.80 = substantial agreement, and 0.81–0.99 = almost perfect agreement [7]. Differences in subjective image quality between the different displays were compared per reader using the Wilcoxon signed-rank test. For all tests, p-values <0.05 were considered statistically significant.

Results

Results were obtained from a total of 864 observations (96 images × 3 radiologists × 3 displays). Of the 100 images included in this study, 56 images were found to have subtle bone fissures in the final consensus reading, while four images were excluded from further analysis as a unanimous decision was not reached during the consensus reading. The remaining 40 images were found to have no perceivable abnormality. Table 2 lists additional details about the included images.

Diagnostic accuracy

Table 3 lists the sensitivities and specificities and the corresponding Youden's index (J) values. Diagnostic performance was comparable between all displays and readers (Fig. 3). Using Youden's index as an outcome measure, reader 1 performed most consistently across all three monitors (J = 0.564 vs. J = 0.579 vs. J = 0.564), whereas reader 2 performed slightly better on the medical-grade 6MP display (J = 0.443 vs. J = 0.582 and J = 0.475), which in turn was the one where reader 3 performed worst (J = 0.661 and J = 0.493 vs. J = 0.604). However, when comparing sensitivity and specificity, reader 1 showed consistently lower sensitivity than specificity (mean sensitivity 70.2%, mean specificity 86.6%), whereas reader 3 had consistently higher sensitivity than specificity (mean sensitivity 82.7%, mean specificity 75.3%).

Reflecting the difficult task of identifying subtle bone fissures, all three readers showed only fair to moderate agreement irrespective of the display used compared to the reference standard established in the consensus reading (κ ranging between 0.425 and 0.673; Table 4).

Subjective image quality

No statistically significant differences were found for the reported subjective image quality of the different displays, with the median being either 1 or 2 (excellent or good quality) on a 5-point Likert scale. Accordingly, all but one comparison using the Wilcoxon signed-rank test showed p-values between 0.106 and 1.000 (Table 5). The only exception was for reader 2, who scored the medical-grade 6MP display slightly better than the consumer-grade display (mean 2.10 vs. 2.01); however, this difference did not reach statistical significance (p = 0.06).

Discussion

The ongoing development of high-performance consumer-grade digital displays raises the question of whether specialised medical-grade displays are still needed to achieve the highest possible diagnostic accuracy. When radiological images were first displayed electronically, cathode-ray tube monitors were state-of-the-art. At that time, it was reasonable to set relatively high standards with regards to display performance, since the quality was markedly different between low-end

Table 3 Sensitivity, specificity, calculated Youden's index, and area under the curve for the ROC analysis for each reader/display combination

	Medical-grade 1 (EIZO RX440)	Medical-grade 2 (EIZO RX650)	Consumer-grade (DELL U3014)
Reader 1 Sensitivity / specificity	71.4% / 85.0%	67.9% / 90.0%	71.4% / 85.0%
Youden's J / AUC	J = 0.564 / AUC = 0.782	J = 0.579 / AUC = 0.789	J = 0.564 / AUC = 0.782
Reader 2 Sensitivity / specificity	64.3% / 80.0%	73.2% / 85.0%	62.5% / 85.0%
Youden's J / AUC	J = 0.443 / AUC = 0.721	J = 0.582 / AUC = 0.791	J = 0.475 / AUC = 0.738
Reader 3 Sensitivity / specificity	91.1% / 75.0%	76.8% / 72.5%	80.4% / 80.0%
Youden's J / AUC	J = 0.661 / AUC = 0.830	J = 0.493 / AUC = 0.746	J = 0.604 / AUC = 0.802

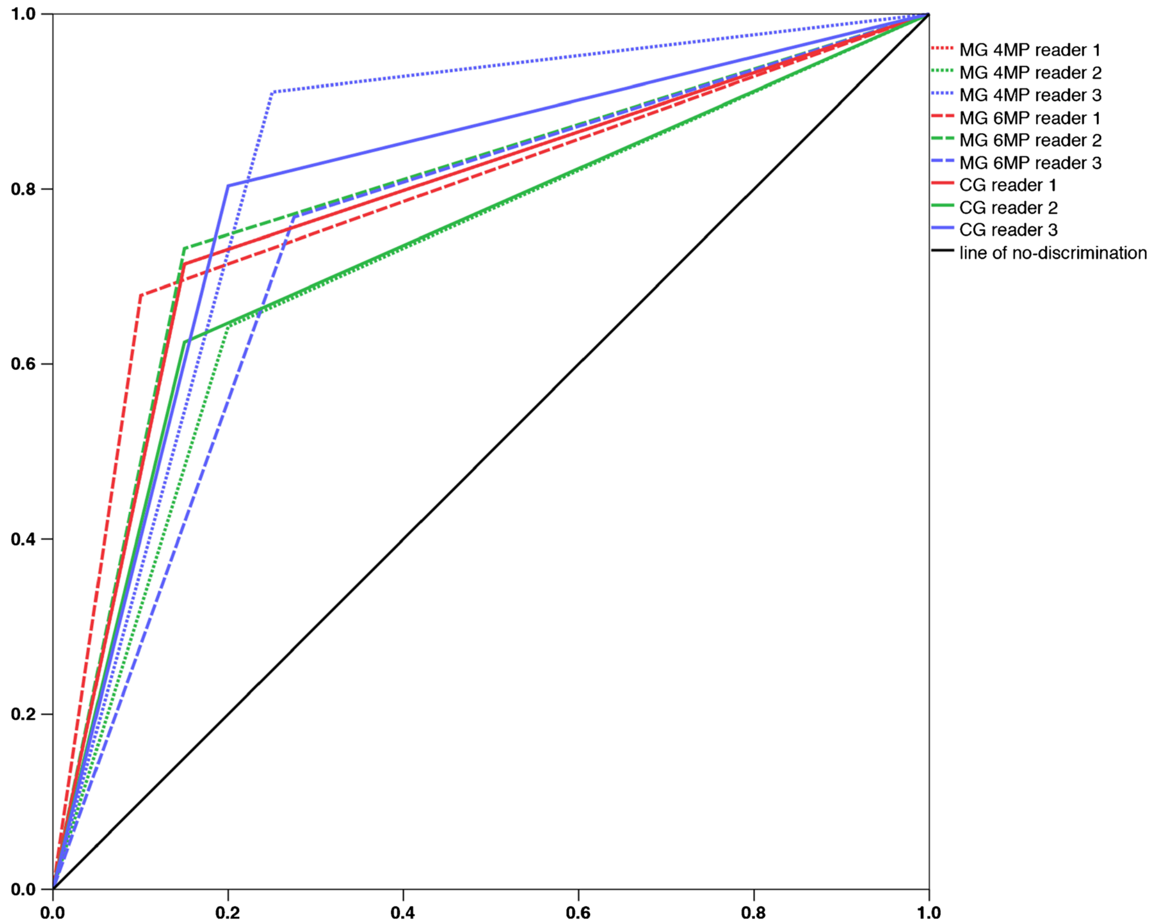

Fig. 3 Receiver operating characteristics for each reader/display combination. As reflected by the Youden's index (J) values and areas under the curve (AUC), all readers performed similarly well independent of the display used (also see Table 3)

consumer-grade monitors and the high-end monitors that met the criteria for displays used in a medical/radiological context. Currently, the technical specifications for consumer-grade displays are no longer very different from those for medical-grade monitors (Table 1).

Several studies have investigated the use of various display types, including mobile devices, for radiological applications [8–12]. Our study supports the results of these previous reports, most of which failed to demonstrate statistically significant differences between medical-grade displays and other display devices under standard reporting conditions.

Table 4 Unweighted Cohen's kappa for each reader/display combination compared to the reference standard set in the consensus reading

	Medical-grade 1 (EIZO RX440)	Medical-grade 2 (EIZO RX650)	Consumer-grade (DELL U3014)
Reader 1	κ = 0.545	κ = 0.551	κ = 0.545
Reader 2	κ = 0.425	κ = 0.564	κ = 0.451
Reader 3	κ = 0.673	κ = 0.489	κ = 0.597

However, one study suggested that aged consumer-grade displays may perform worse than comparably aged medical-grade displays [13].

Theoretical reasoning may point to favourable properties of medical-grade displays and higher performance, although even this is controversial. A study by Hiwasa et al. concluded that displays with 10-bit grayscale technology performed better with regards to discerning low-contrast objects [14]. In contrast, Bender et al. concluded that at higher resolutions, 8-bit displays performed better, with some radiologists noting that 10-bit displays lacked sharpness and contrast [15]. Interestingly, one study that used contrast phantoms and objective methods to assess display quality also failed to show a benefit for monochrome medical-grade displays over consumer-grade colour displays [16].

Most studies of display performance have focused on chest x-rays [9, 11]. Other studies focus more on the clinical setting, such as display performance in emergency radiology for pulmonary embolism, abdominal haemorrhage, or acute appendicitis. In these settings, the imaging findings are sometimes accompanied by collateral phenomena that might help make

Table 5 Comparison of subjective image quality for each reader/display combination using the Wilcoxon signed-rank test

	Medical-grade 1 vs. Medical-grade 2 (EIZO RX 440 vs. EIZO RX650)	Medical-grade 1 vs. Consumer-grade (EIZO RX 440 vs. DELL U3014)	Medical-grade 2 vs. Consumer-grade (EIZO RX 650 vs. DELL U3014)
Reader 1	$p = 0.758$	$p = 0.442$	$p = 0.296$
Reader 2	$p = 1.000$	$p = 0.106$	$p = 0.055$
Reader 3	$p = 0.841$	$p = 0.724$	$p = 0.590$

the correct diagnosis [12, 17–20]. To overcome this limitation, we investigated displays by evaluating the presence or absence of subtle bone fissures. This is an especially difficult clinical task, as these fissures sometimes appear as only a thin low-contrast dark line amidst other low-contrast but bright lines that represent trabecular bone, and they are not accompanied by collateral phenomena like soft tissue swelling or joint effusion. The difficulty of this task is clearly reflected by the relatively low inter-rater agreement. A similar study on focal liver lesions in unenhanced CT scans has shown comparable results [21].

Outcome evaluation depends on the selected outcome measure. In our study, we used Youden's index as an outcome measure as it reflects the aggregated sensitivity and specificity of a test. Although inter-rater agreement was low, with reader 3 showing a tendency to over-read and reader 1 a tendency to under-read, Youden's index was similar for all readers. This further suggests that the display type did not impact diagnostic accuracy in our study.

In conclusion, our results show that the use of consumer-grade displays does not reduce diagnostic performance even for a relatively difficult task as identifying subtle bone fissures. This has practical and economic implications. The use of adequately calibrated and tested consumer-grade displays in radiological practices could help reduce display costs by 60%–80%, especially in larger institutions where a large number of displays is needed. However, to avoid legal and insurance-related liability issues, radiologists should check for approval of combinations of consumer-grade monitors and calibration tools by their respective authorities.

Acknowledgements The authors thank Ms. Helen Toder and Ms. Franziska Inka Meyer for their support in conducting this study and for their help with data collection. The authors would also like to thank Dr. Veronika Weyer-Elberich for her help in planning the statistical analysis for this study.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Prof. Dr. Peter Mildenberger.

Conflict of interest The authors of this manuscript declare no relationships with any companies, that have products or services may be related to the subject matter of the article.

Funding The authors state that this work has not received any funding.

Statistics and biometry Ms. Dr. Veronika Weyer- Elberich kindly provided statistical advice for this manuscript.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology
•experimental
•performed at one institution

References

1. Kagadis GC, Walz-Flannigan A, Krupinski EA, Nagy PG, Katsanos K, Diamantopoulos A et al (2013) Medical imaging displays and their use in image interpretation. RadioGraphics 33:275–290
2. Fetterly KA, Blume HR, Flynn MJ, Samei E (2008) Introduction to grayscale calibration and related aspects of medical imaging grade liquid crystal displays. J Digit Imaging 21:193–207
3. German Institute for Standardization. Image quality assurance in diagnostic X-ray departments - Part 157: X-ray ordinance acceptance and constancy test of image display systems in their environment (2014). Available via https://www.beuth.de/de/norm/din-6868-157/216064441
4. Food and Drug Administration. Display Accessories for Full-Field Digital Mammography Systems-Premarket Notification (510(k)) (2008). Available via http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm107549.htm
5. Likert R (1932) A technique for the measurement of attitudes. Arch Psychol 22:55
6. Youden WJ (1950) Index for rating diagnostic tests. Cancer 3:32–35
7. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174
8. Abboud S, Weiss F, Siegel E, Jeudy J (2013) TB or Not TB: interreader and intrareader variability in screening diagnosis on an iPad versus a traditional display. J Am Coll Radiol 10:42–44
9. Salazar AJ, Aguirre DA, Ocampo J, Camacho JC, Díaz XA (2014) DICOM gray-scale standard display function: clinical diagnostic accuracy of chest radiography in medical-grade gray-scale and consumer-grade color displays. AJR Am J Roentgenol 202:1272–1280
10. Ekpo EU, McEntee MF (2016) An Evaluation of Performance Characteristics of Primary Display Devices. J Digit Imaging 29): 175–182

11. Liukkonen E, Jartti A, Haapea M, Oikarinen H, Ahvenjärvi L, Mattila S et al (2016) Effect of display type and room illuminance in chest radiographs. Eur Radiol 26:3171–3179
12. Tewes S, Rodt T, Marquardt S, Evangelidou E, Wacker FK, von Falck C (2013) Evaluation of the use of a tablet computer with a high-resolution display for interpreting emergency CT scans. Röfo 185:1063–1069
13. Krupinski EA (2009) Medical grade vs off-the-shelf color displays: influence on observer performance and visual search. J Digit Imaging 22:363–368
14. Hiwasa T, Morishita J, Hatanaka S, Ohki M, Toyofuku F, Higashida Y (2009) Need for liquid–crystal display monitors having the capability of rendering higher than 8 bits in display-bit depth. Radiol Phys Technol 2:104–111
15. Bender S, Lederle K, Weiß C, Schoenberg SO, Weisser G (2011) 8-bit or 11-bit monochrome displays–which image is preferred by the radiologist? Eur Radiol 21:1088–1096
16. Geijer H, Geijer M, Forsberg L, Kheddache S, Sund P (2007) Comparison of color LCD and medical-grade monochrome LCD displays in diagnostic radiology. J Digit Imaging 20: 114–121
17. Faggioni L, Neri E, Bargellini I, Scalise P, Calcagni F, Mantarro A et al (2015) iPad-based primary 2D reading of CT angiography examinations of patients with suspected acute gastrointestinal bleeding: preliminary experience. Br J Radiol 88:20140477
18. Kim C, Cha H, Kang BS, Choi HJ, Lim TH, Oh J (2015) A Feasibility Study of Smartphone-Based Telesonography for Evaluating Cardiac Dynamic Function and Diagnosing Acute Appendicitis with Control of the Image Quality of the Transmitted Videos. J Digit Imaging 29:347–356
19. Johnson PT, Zimmerman SL, Heath D, Eng J, Horton KM, Scott WW et al (2012) The iPad as a mobile device for CT display and interpretation: diagnostic accuracy for identification of pulmonary embolism. Emerg Radiol 19:323–327
20. Schlechtweg PM, Kammerer FJ, Seuss H, Uder M, Hammon M (2016) Mobile Image Interpretation: Diagnostic Performance of CT Exams Displayed on a Tablet Computer in Detecting Abdominopelvic Hemorrhage. J Digit Imaging 29:183–188
21. Kawasumi Y, Yamada T, Ota H, Tsuboi M, Takase K, Sato A et al (2008) High-resolution monochrome liquid crystal display versus efficient household colour liquid crystal display: comparison of their diagnostic performance with unenhanced CT images in focal liver lesions. Eur Radiol 18:2148–2154