J Digit Imaging (2015) 28:646–653
DOI 10.1007/s10278-015-9783-7

Comparison of diagnostic accuracy of plain film radiographs between original film and smartphone capture: a pilot study

Mindy Y. Licurse · Sung H. Kim · Woojin Kim · Alexander T. Ruutiainen · Tessa S. Cook

Published online: 4 April 2015

Abstract The use of mobile devices for medical image capture has become increasingly popular given the widespread use of smartphone cameras. Prior studies have generally compared mobile phone capture images to digitized images. However, many underserved and rural areas without picture archiving and communication systems (PACS) still depend greatly on the use of film radiographs. Additionally, there is a scarcity of specialty-trained or formally licensed radiologists in many of these regions. Subsequently, there is great potential for the use of smartphone capture of plain radiograph films which would allow for increased access to economical and efficient consultation from board-certified radiologists abroad. The present study addresses the ability to diagnose a subset of radiographic findings identified on both the original film radiograph and the captured camera phone image.

Keywords Teleradiology · Diagnostic image quality · Image quality analysis · Digital image processing · ROC-based analysis

Introduction

Teleradiology—the electronic transmission of images across long distances for consultation and review—has become increasingly recognized as an invaluable tool throughout the world. This is particularly true in disadvantaged, underserved regions that may not have easy access to specialty-trained radiologists. With the persistent acceleration of mobile technology and improved digitization of images, the possibilities for seamless image transfer to support diagnostic interpretation continue to grow. Despite the lack of high bandwidth networks and advanced equipment, underserved populations tend to have nearly universal access to smartphones. For example, in South Africa, there were 68,934,000 mobile cellular subscriptions in 2012 [1] for a population of 58,800,000 individuals, [2] taking into account that some users have multiple subscriptions. In all of Africa, there were 69.3 mobile cellular subscriptions in 2014 per 100 people [1]. Unlike some telemedicine and teleradiology endeavors such as videoconferencing and evolving cloud-based picture archiving and communication systems (PACS), which have been more widely accepted in higher income regions, smartphones are inexpensive, user-friendly, and readily available.

There have been many recent studies regarding the use of mobile devices for medical image capture and imaging from multiple disciplines. Prior to increasingly widespread use of smartphone cameras, studies in the early 2000s concentrated on technical aspects of using digital cameras for sharing of diagnostic images in early telemedicine models [3–6]. While there have been some studies focused on diagnostic imaging, many of these studies have come from specialties other than radiology [7–10]. For example, orthopedics [7, 9] and ophthalmology [8] have demonstrated an interest in smartphone photography for clinical management. Padmasekara et al. 2012 [7] exhibited strong inter-rater agreement amongst orthopedic surgeons when comparing smartphone iPhone 3GS images of distal radial head fractures interpreted on the iPhone itself via multimedia messaging (MMS) when compared to digitized PACS images. Additionally, Bullard et al. 2013 [9] from the emergency department demonstrated that mobile phone images of CT scans appeared to provide adequate imaging for triaging neurosurgical patients to a level 1 trauma center with strong inter-reader agreement that increased upon

M. Y. Licurse (✉) · S. H. Kim · W. Kim · A. T. Ruutiainen · T. S. Cook
Department of Radiology, Hospital of University of Pennsylvania, 1 Silverstein Building, 3400 Spruce Street, Philadelphia, PA 19104, USA
e-mail: mindylicurse@gmail.com

adding imaging to available clinical data. While these studies demonstrated the adequacy of mobile capture images as compared to digitized images, there are fewer studies demonstrating comparisons between mobile capture images and original film radiographs. Addressing the controversy of loss in image quality by digital camera capture, a dentistry study in 2009 [11] did not find loss of critical information with digital camera capture compared to original film radiographs, while a study in 2001 [12] did note that consumer-level digital cameras were understandably of much poorer quality compared to commercial-grade digital cameras. Noting that the resolution of mobile capture images is most certainly less than the original film radiographs, there are scenarios in which this trade-off may be acceptable. Many underserved or rural areas without specialists in place would benefit from the ability to quickly and easily share images of film radiographs with trained radiologists across the world, especially when diagnoses may be vital to altering patient management. Subsequently, there is a need for collecting data comparing mobile phone capture to original film radiographs, while prior data has generally compared mobile phone capture to digitized images. The ability to digitize film radiographs is often inapplicable and lacking from many underserved or rural areas without PACS or advanced hardware in place. Furthermore, there is a paucity of specialty-trained or in some cases, formally licensed radiologists, in many of these regions, which leads to reliance upon those who may not have sufficient training for diagnostic interpretation. Diagnoses made in rural clinical practices in developing nations sometimes require the consultation of a board-certified radiologist from abroad; smartphone capture could increase access to such radiologists.

The present study aims to determine if reproductions of film radiographs produced by using smartphone capture are of sufficient diagnostically quality compared to viewing the original film radiographs using a lightbox. The primary outcome variable is the degree of detectability of a subset of radiographic findings identified on both the original film radiographs and the smartphone photos.

Materials and Methods

Institutional Review Board approval was obtained prior to the initiation of this study. Pre-existing screen-film radiographs archived within the teaching files of the Department of Radiology of the University of Pennsylvania were selected for review in a non-random fashion by the principal investigator (MYL) with guidance from a board-certified radiologist (TSC). Films were selected to include representative diagnoses ranging in frequency from common (lobar pneumonia) to rare (Loeffler's syndrome). Forty-four total radiographs were selected, consisting of 16 chest radiographs and 28 musculoskeletal radiographs.

An 8-megapixel smartphone camera (Apple® iPhone 4S) was used for digital image capture. The camera uses a back-side illuminated complementary metal-oxide-semiconductor sensor (Sony IMX105 Exmor R) measuring 1/3.2″ along the diagonal (4.54 mm × 3.42 mm), with a resolution of 8 megapixels (3264 × 2448 pixels) that produces photos with 1.4 μm pixels, and a 4.28 mm f/2.4 five-element plastic lens (35 mm-film equivalent) with a 6.5 cm macro working distance. Images were acquired at 3 ft from a standard 8-film viewbox (GE Medical Systems, Waukesha, WI). These images were then arranged in a set random sequence and distributed to the participating interpreting radiologists on departmental computers with standard liquid crystal display monitors. Figure 1 is an example of a smartphone-captured image.

Both the smartphone-captured images and the original film radiographs were reviewed by two board-certified radiologists blinded to the diagnoses, with a 1-week washout period between review sessions. The smartphone-captured images were reviewed on departmental computers as described, while the original films were reviewed on standard viewboxes. No clinical history was provided. The readers rated their confidence in the presence or absence of an abnormality in five independent categories using a Likert scale: "lung," "mediastinal or abdominal," cardiac," "bone," and "soft tissue". The radiologists were allowed to use routine

Fig. 1 Example of smartphone-captured image of elbow radiograph

manipulations of the smartphone photographs including pan, zoom, and image rotation. The Likert scale of subjective confidence in the presence or absence of an abnormality ranged from "definitely present," "probably present," "unsure" (equivocal), "probably absent," to "definitely absent," which was transposed on a numerical ordinal scale from 1 to 5, respectively, in an effort to quantify the process. The radiologists were asked to formulate a final diagnosis (e.g., "sarcoidosis") and enter it into a free-text box at the end of the survey.

The gold standard answer for each case was determined based on consensus between the final diagnosis on the case, the principal investigator (MYL), and a board-certified radiologist (TSC). Their distribution is detailed in Table 1.

The Likert scale of reader confidence was used to generate receiver operating characteristic curves (ROCs) for both readers using either film or smartphone capture for each disease category using the MedCalc statistical package (MedCalc Software, Ostend, Belgium). These were compared to one another using the methodology of Delong et al. [13] using a cutoff for statistical significance of $p<0.01$ due to the large number of independent comparisons.

Results

The true diagnoses for each case (some cases had more than one major primary diagnosis) are shown in Table 2 along with the concordance between the diagnoses and the two readers during both the film and smartphone capture.

Answers were considered "partial" agreement if a similar diagnosis or only part of the full diagnosis was given, at the discretion of the board-certified radiologist and radiology resident after consulting the gold standard diagnoses given by the teaching file.

Pairwise comparison of the two readers using either film or smartphone capture for the disease categories of lung, mediastinal/abdominal, bone, and soft tissue demonstrated no significant difference in the areas under the ROC curves (each $p>0.01$) (Fig. 2). Only a single positive cardiac finding was included in the data sample, which prevented the generation and comparison of ROC curves.

Additionally, Fig. 3 demonstrates the Likert scale results for reader 1 and reader 2, with the cases grouped in order of Likert scale from cases with the abnormality definitely present to definitely absent. The gray-shaded portion of the graph overlies the cases with true abnormalities present according to the pre-determined gold standard. For example, in the first row of graphs, the cases on the left are almost universally agreed upon by reader 1 and reader 2 to have an abnormality within the lungs when interpreted on film radiographs, and those cases are shown by the shaded box to have an abnormality in the lungs by the gold standard.

Discussion

The importance of determining the diagnostic accuracy of smartphone capture of film radiographs is immeasurable. In remote or otherwise inaccessible locations of the world, the ability to add potentially management-altering diagnostic information in a clinical setting which would otherwise lack the adequacy to have such information has been long sought after. For example, acute and emergent cases such as aortic dissection or pneumomediastinum from an esophageal tear were among the cases accurately diagnosed by smartphone capture. With inadequate access to radiologic expertise, these cases could very well be missed with dire consequences.

While prior studies have examined the utility and limitations of digital cameras [3–6], smartphone capture specifically has limited data. Given the increasingly universal access to smartphones both inside and outside of healthcare, the necessity to assess diagnostic accuracy of smartphone capture has become clear.

This study focused on the statistical analysis of the effect of smartphone digitization on general abnormalities identified on plain film radiographs, the mainstay of radiology in underserved regions of the world. Additionally, this study allowed for identification of the limitations of smartphone capture and understanding which abnormalities may have the most difficult interpretation with reduced diagnostic accuracy.

The majority of diagnoses were made on film radiographs as well as smartphone capture, without difficulty in comparison to the gold standard. The more common entities, such as infiltrative pneumonia and pulmonary edema, were more accurately described than the entities less common in the USA at this present time, such as measles pneumonia, Loeffler's syndrome, and Maffuci's syndrome. This is an important factor to consider as there are characteristic disease patterns endemic to certain regions and portions of the world, which may not be as readily identified by radiologists who have trained and

Table 1 Subcategories of diagnostic abnormalities

	Lung abnormality	Mediastinal or abdominal abnormality	Bone abnormality	Cardiac abnormality	Soft tissue abnormality
Number of cases present	12	7	28	1	15
Number of cases absent	32	37	16	43	29

Table 2 Table with the true diagnoses listed along with whether each respective reader made the complete or partial diagnosis on film and smartphone images

True Diagnosis	Reader 1 Agreed on Film (Yes, No, Partial)	Reader 2 Agreed on Film (Yes, No, Partial)	Reader 1 Agreed on Smartphone (Yes, No, Partial)	Reader 2 Agreed on Smartphone (Yes, No, Partial)
Scleroderma, breast cancer	Partial	Yes	Partial	Partial
Achalasia	Yes	Yes	Yes	Yes
RML pneumonia, pregnancy	Partial	Yes	Partial	Yes
Pneumomediastinum and subcutaneous emphysema	Yes	Yes	Yes	Yes
Pleural plaques with new lung nodule	Yes	Yes	Partial	Yes
Lymphadenopathy	Yes	Yes	Yes	Yes
Retropharyngeal abscess	Yes	Yes	Yes	Yes
Loeffler's syndrome	No	No	No	No
Right pneumothorax	Yes	Yes	Yes	Yes
Measles pneumonia	Yes	No	Yes	Partial
Chronic eosinophilic pneumonia (pre- and post-treatment)	Yes	No	Partial	No
Rheumatoid lung	Partial	Yes	Partial	Yes
LUL carcinoma not seen on prior portable CXR	No	Yes	No	Yes
Seminoma	Yes	No	Yes	No
Sarcoid	Yes	Yes	Yes	Partial
Non-cardiogenic edema following molar pregnancy	Yes	Yes	Yes	Yes
Aortic dissection	Yes	Yes	Yes	Yes
Mafucci's syndrome	Yes	No	No	No
Scurvy	No	Partial	Partial	Partial
Paget's disease	Yes	Yes	No	Yes
Ewing's sarcoma	Partial	Yes	Partial	Yes
Hangman's fracture	Yes	Yes	Yes	Yes
Chance fracture	Yes	Yes	Yes	Yes
Osteoid osteoma	Yes	No	No	No
Gout	Yes	Yes	No	Yes
Pathologic fx through enchondroma	Yes	Yes	Partial	Yes
Sarcoidosis (Musculoskeletal)	No	Partial	No	No
Split depressed lateral tibial plateau fracture	Yes	Yes	Yes	Yes
LisFranc fracture	Yes	Yes	No	Yes
Patellar fracture	Yes	Yes	Yes	Yes
Radial buckle fracture	Yes	Yes	Yes	Yes
Anterior shoulder dislocation with greater tuberosity fracture	Yes	Yes	Yes	Yes

Table 2 (continued)

2 pelvic rami fracture on right	Yes	Yes	Yes	Yes
Pelvic diastasis	Yes	Yes	Yes	Yes
Pseudohypoparathyroidism	Yes	Yes	Yes	Yes
Chondrocalcinosis	Yes	Yes	Yes	Yes
Osteopetrosis	Yes	Yes	Yes	Yes
Congenital rubella syndrome	Yes	Partial	Partial	No
Pathologic fracture with metastatic lung cancer	Yes	Yes	Partial	Partial
Fibular fracture and soft tissue swelling	Yes	Yes	Yes	Yes
Scaphoid fracture	Yes	Yes	No	Yes
Radial head fracture with soft tissue swelling	Yes	Yes	Yes	Yes
Volar plate fracture	No	Yes	No	Yes
Radial buckle fracture	Yes	Yes	Yes	Yes

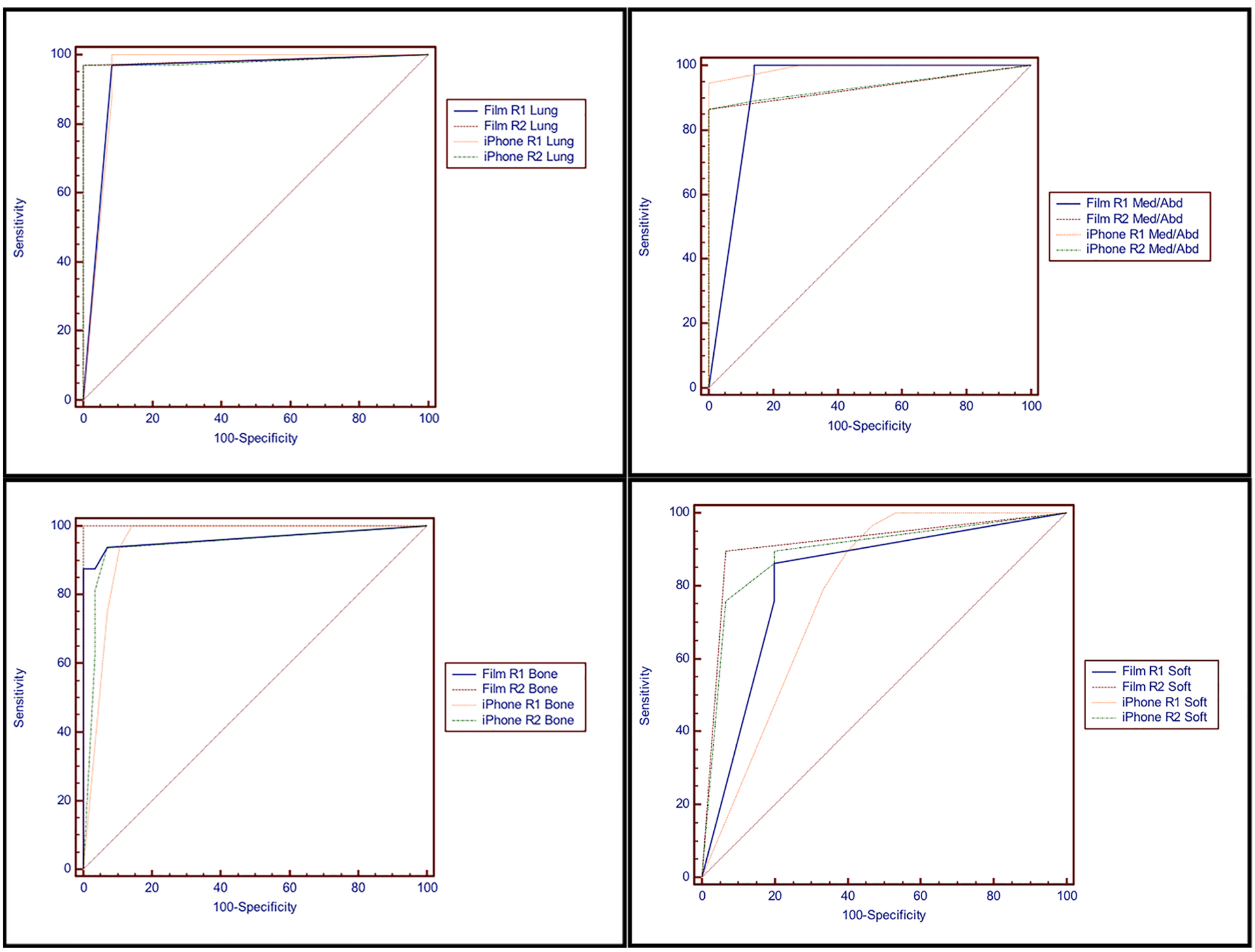

Fig. 2 Comparison of the two readers using film and iPhone capture in assessing four subcategories including "lung," "mediastinal or abdominal," "bone," and "soft tissue" abnormalities; there was no significant difference in the areas under the ROCs (p>0.01)

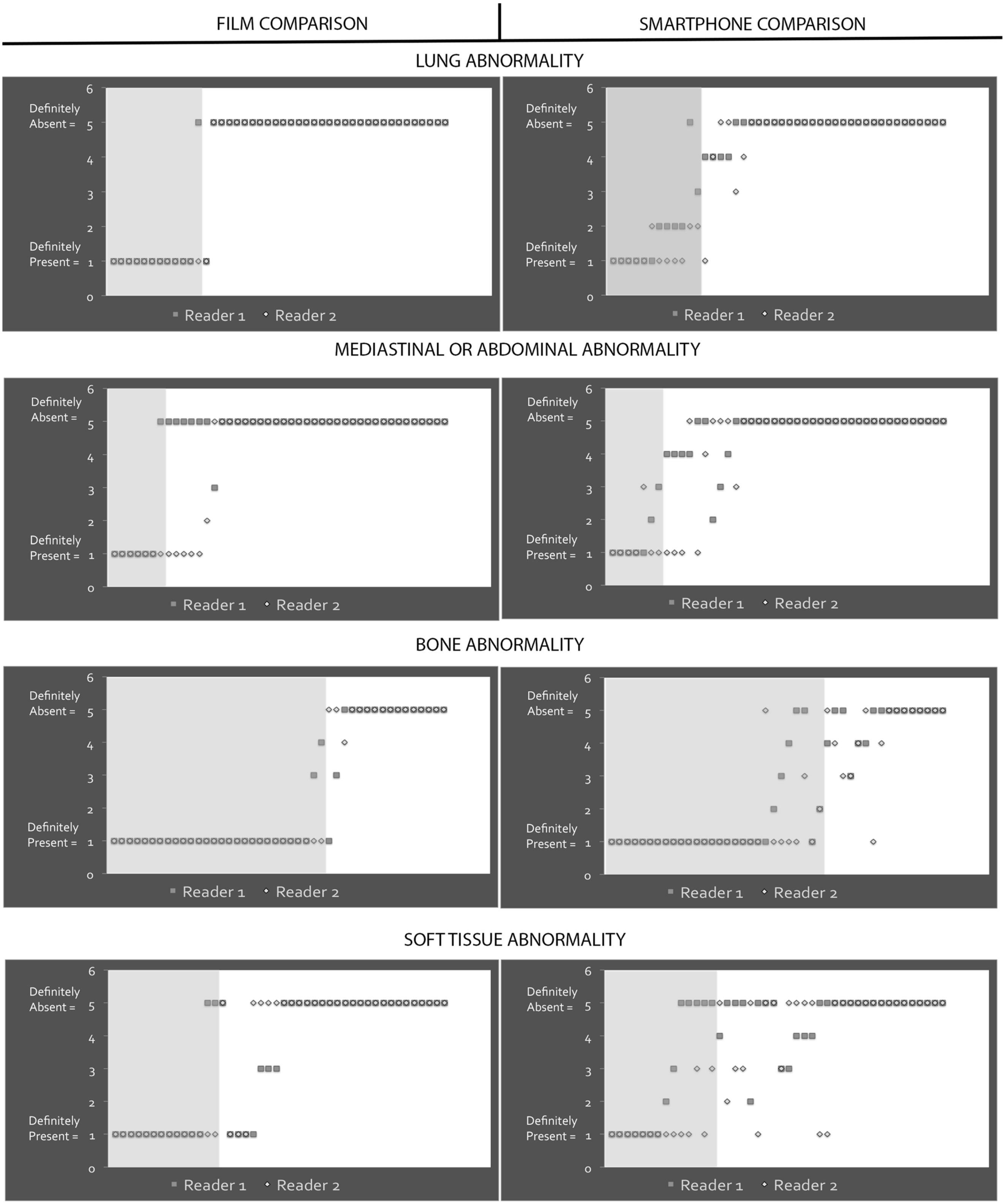

Fig. 3 Comparison of reader 1 and reader 2 Likert scale results for assessing the presence or absence of four subcategory abnormalities (lung, mediastinal or abdominal, bone, and soft tissue) grouped by Likert scale results with "abnormality definitely present" (Likert scale= 1) to "abnormality definitely absent" (Likert scale=5). The true positives ("abnormality definitely present") by gold standard are represented by the gray-shaded box. The left column delineates comparisons between reader 1 and reader 2 results for cases when viewed as original film radiographs on viewboxes. The right column delineates comparisons between reader 1 and reader 2 results for cases when viewed by smartphone capture images

practiced mainly in the USA. Given these limitations, further evaluation in differing regions of actual underserved or rural areas should be carried out to better understand diagnostic utility and management algorithms. Following these studies, guidelines for the utilization of smartphone capture in such areas may be more specifically designated.

No significant differences in diagnosing a subcategory (e.g., lung) abnormality were seen between film and smartphone capture, suggesting that identifying the presence or absence of an abnormality on smartphone captured images can be done with a high degree of confidence by a board-certified radiologist. Although not significantly different, only a few diagnostic abnormalities were misinterpreted on film, while a higher number of diagnostic abnormalities were missed on smartphone capture images with greater variance, as would be expected given the greater degree of technical limitations. In this scenario, consideration must also be given to the concept that certain diagnostic subcategories likely have inherently high inter-reader variability, which may increase the apparent effect of variance seen between film and smartphone capture diagnoses. An example is the specific diagnostic abnormality of cardiomegaly, for which the common degrees of "mild," "moderate," or "severe" likely vary greatly from reader to reader.

As Table 2 demonstrates, overall diagnostic accuracy by the two board-certified radiologists was high. Nonetheless, making the overall diagnosis was clearly more difficult on smartphone capture, again as would be expected given the greater degree of technical limitations. In particular, the diagnoses that proved most challenging generally appear to involve rare, more complex syndromes. These syndromes tend to require identification of additional subtle findings that may be especially hard to recognize with certainty on smartphone captured images. It is therefore prudent to recognize that a major disadvantage of smartphone capture is the decreased ability to detect complex syndromes that depend upon higher resolution and details that may not be recognizable on smartphone-captured images.

Additionally, many of the more problematic diagnoses tend to suffer the most from lack of clinical history, which was not provided to the reviewers. For example, understanding a patient that has blood eosinophilia or an atopic history is highly beneficial in leading a radiologist to suggest Loeffler's syndrome given the appropriate imaging characteristics, such as migratory peripheral consolidations in the mid to upper lung zones, in this specific case. Furthermore, differences in prioritizing diagnostic abnormality categories may differ greatly given clinical history. For example, given the complaint of "cough" versus "weight loss" may lead a radiologist down separate diagnostic paths and search patterns.

One of the limitations of this study was the recall bias as the same radiologists interpreted the same cases on film using a lightbox as well on the smartphone captured images. In an effort to minimize this bias, there was randomization of the smartphone captures. Additionally, there was a 1-week washout period of time instituted between interpretation of the smartphone captures versus the film studies.

A technical limitation of the study was also optimization of the smartphone capture technique. It was found that at a distance of less than 2.5–3 ft from the lightbox, an oscillating dark-band artifact would appear across the photo, obscuring the image, which limited the proximity at which the photograph could be taken. Assessing the technical components of smartphone capture of film hanging on a lightbox would be helpful in future optimization of this approach, especially given the need for stable reproducibility worldwide. As with any imaging modalities, smartphone capture certainly has additional artifacts unique to the modality, which must be explored further to ensure proper recognition if this modality is to be implemented on a larger basis. Additionally, other challenges in real-life applications must be recognized such as the inability to control ambient light conditions in many regions which smartphone capture may be applied.

An additional limitation includes the small sample size of readers, which clearly limits the power of the study. Further phases of this study may be expanded to include increased numbers of radiologists with varied experience, including trainees, as well as physicians (both radiologist and non-radiologist) from different subspecialties, to determine whether these conclusions hold true.

A final limitation is taking into account the differences in quality of smartphone cameras, which may be pertinent to the variations in smartphone quality secondary to economic variations in underserved populations throughout the world. However, the rate of smartphone camera quality has risen exponentially in recent years, and it is likely that most smartphones in the world at large are equivalent or better in resolution to the iPhone 4S 8-megapixel camera used in this study.

Conclusions

Overall, the findings suggest that identifying the presence or absence of major diagnostic abnormalities, in addition to identifying most well-known diagnoses, on smartphone-captured images utilizing at least an 8-megapixel camera, can be done with a high degree of confidence by a board-certified radiologist. While the majority of diagnoses commonly encountered in the USA were made, the more rare and complex disease processes such as Loeffler's syndrome were missed. Given the differences in certain disease patterns throughout the world, identifying the disease processes which are common or endemic to specific regions of the world would be helpful for understanding how best to prepare and optimize diagnostic approaches in other countries outside of the USA. Additionally, subanalyses of the specific features of each disease on film radiograph versus smartphone capture photographs may be of utility.

Other variables that may be further evaluated include the effect of image manipulation by the interpreter such as zooming, rotating, etc. A challenge that will eventually need to be addressed also involves the transfer of such images, which has been studied sparsely and in developed rather than developing countries [14].

The most appropriate next step would be for real-world testing outside of a controlled environment. This would allow us to evaluate impact on clinical care and infrastructure limitations that may not be evident at this time. Additionally, we would be able to propose more robust guidelines for the use of smartphone capture in teleradiology and address other challenges, such as the variability in ambient lighting in which film radiographs are interpreted.

Acknowledgments We gratefully acknowledge Judith Aronchick, MD and Michael Love, MD, for providing access to their film-based radiology teaching files.

References

1. International Telecommunication Union. Available at http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx; Accessed August 30, 2013
2. Population Reference Bureau. Available at http://www.prb.org/Publications/Datasheets/2012/world-population-data-sheet/world-map.aspx#/map/population; Accessed August 30, 2013
3. Brault B, Hoskinson J, Armbrust L, Milliken G: Comparison of Seven Digital Cameras for Digitizing Radiographs. Vet Radiol Ultrasound 45:298–304, 2004
4. Corr P, Couper I, Beningfield SJ, Mars M: A simple telemedicine system using a digital camera. J Telemed Telecare 6(4):233–236, 2000
5. Salazar AJ, Camacho JC, Aguirre DA: Comparison between different cost devices for digital capture of X-ray films: an image characteristics detection approach. J Digit Imaging 25(1):91–100, 2012
6. Szot A, Jacobson FL, Munn S, Jazayeri D, Nardell E, Harrison D, Fraser HSF: Diagnostic accuracy of chest X-rays acquired using a digital camera for low-cost teleradiology. Int J Med Inform 73(1):65–73, 2004
7. Padmasekara G, Nazarian N, Wall C: The reliability of mobile multimedia messaging (MMS) for decision making in distal radius fractures: an effective alternative. J Mob Technol Med 1:8–12, 2012
8. Lamirel C, Bruce BB, Wright DW, Newman NJ, Biousse V: Nonmydriatic digital ocular fundus photography on the iPhone 3G: the FOTO-ED study. Arch Ophthalmol 130:939–940, 2012
9. Bullard TB, Rosenberg MS, Ladde J, Razack N, Villalobos HJ, Papa L: Digital images taken with a mobile phone can assist in the triage of neurosurgical patients to a level 1 trauma centre. J Telemed Telecare 19(2):80–83, 2013
10. Tennant JN, Shankar V, Dirschl DR: Reliability and Validity of a Mobile Phone for Radiographic Assessment of Ankle Injuries: A Randomized Inter- and Intraobserver Agreement Study. Foot Ankle Int 34:228, 2013
11. Amir E, Kaffe I, Peretz B: Digital images obtained with a digital camera are no associated with a loss of critical information—a preliminary study. Br Dent J 206(5):268–9, 2009
12. Davidson HC, Johnston DJ, Christian ME, Harnsberger HR: Comparison of Radiographic Image Quality From Four Digitzation Devices as Viewed on Computer Monitors. J Digit Imaging 14(1): 24–29, 2001
13. Delong ER, Delong DM, Clarke-Pearson DL: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44:837–845, 1988
14. Yamada M, Watarai H, Andou T, Sakai N: Emergency image transfer system through a mobile telephone in Japan: technical note. Neurosurgery 52:986–990, 2003