Evaluation of the Use of a Tablet Computer with a High-Resolution Display for Interpreting Emergency CT Scans

Evaluation eines Tablet-Computers mit hochauflösendem Display zur Befundung von Notfall-CT-Untersuchungen

Abstract

Purpose: Evaluation of the potential usability of an iPad 3 with a high-resolution display in CT emergency diagnosis compared to a 3D PACS workstation.

Materials and Methods: 3 readers used a 5-point Likert scale to evaluate 40 CCT scans and 40 CTPA scans to determine the detectability of early signs of infarction in CCT or segmental and subsegmental pulmonary embolisms in CT angiography of the pulmonary arteries (CTPA) on the iPad 3 (Apple Inc., USA) using an application for image viewing (Visage Ease, Visage Imaging GmbH, Berlin) and on a 3D PACS workstation (Visage 7.1, Visage Imaging, Berlin) using a certified monitor for image viewing. The results were compared using the Wilcoxon rank sum test, Spearman’s correlation coefficient, and a kappa statistic.

Results: There was no significant difference in the median evaluations for the readings of both the CCT scans and the CTPA scans on the iPad 3 and on the workstation (p > 0.05) for all three readers. The mean Spearman’s correlation coefficient for CCT and CTPA was 0.46 (± 0.2) and 0.69 (± 0.16), respectively, for the comparison iPad/PACS, 0.41 (± 0.16) and 0.68 (± 0.06), respectively, for the interobserver agreement on the iPad, and 0.35 (± 0.05) and 0.68 (± 0.10), respectively, for the interobserver agreement on the PACS. Mean kappa values for CCT of 0.52 (± 0.17) for the comparison iPad/PACS and 0.33 (± 0.16) and 0.32 (± 0.16), respectively, for the interobserver agreement on the iPad and the PACS were achieved. For CTPA average kappa values of 0.67 (± 0.19) were calculated for the comparison iPad/PACS and 0.69 (± 0.08) and 0.60 (± 0.14), respectively, for the interobserver concordance on the iPad 3 and the PACS. All differences were not statistically significant (p > 0.05).

Conclusion: The variability of the interpretation of typical emergency scans on an iPad 3 with a high-resolution display and on a 3D PACS workstation does not differ from the interobserver variability.

Key Points:

- Variability of interpretation of emergency scans on an iPad 3 and on a 3D PACS workstation shows no difference in interobserver variability.
- New applications for the iPad in radiology include quick preliminary evaluation on duty and real-time consultation of radiological background services or expert opinions.
- Limitations for the application of tablet-computers in radiology are screen size and inconsistent and non-standardized environmental conditions when using a mobile image display device.

Citation Format:


Zusammenfassung

Ziel: Evaluation einer potentiellen Nutzbarkeit eines iPad 3 mit hochauflösendem Display im Rahmen der CT-Notfalldiagnostik im Vergleich zu einer 3D-PACS-Workstation.

Material und Methoden: Die Erkennbarkeit von Infarktfrühzeichen in der CCT bzw. von segmentalen und subsegmentalen Lungenembolien in der CT-Angiografie der Pulmonalarterien (CTPA) wurde durch 3 Leser anhand von jeweils 40 Un-
Introduction

Since the introduction of the Apple iPad in April 2010, the prevalence of mobile tablet computers has increased rapidly and they already comprise a major portion of the PC market. Numerous different versions and competing products are now available. Based on initial studies with smartphones, both users and manufacturers showed great interest in applications for viewing radiological image data soon after the market introduction of the iPad [1 – 4]. After significant initial regulatory concerns, the FDA finally approved the first mobile radiology application for the iPad for computed tomography. The approval makes it possible to use a tablet for the primary interpretation of image data. Prior to inclusion in the study, the data were verified by two of the authors in consensus. 20 additional native CCT scans with no abnormal acute ischemic findings were also selected. In total, 40 CCT scans from 40 patients (f = 23, m = 17, average age: 64, range: 22 – 91) were used. All CCT scans were acquired on 16-row multidetector spiral computed tomography units (MDCT) using a sequential scan technique and orbitomeatal angulation of the gantry. Two data sets with a slice thickness of 2.5 mm and 5.0 mm were reconstructed from the raw data using filtered back projection.

CCT data acquisition

To evaluate early signs of infarction, 20 native cerebral CT scans (CCT) acquired between 10/2010 and 5/2012 during the night or weekend shift and exhibiting one or more early signs of infarction (hyperdense middle cerebral artery sign, loss of insular cortex, inability to differentiate between gray and white matter, sulcus effacement) were selected retrospectively from the PACS. The selection was performed primarily on the basis of the radiological findings. Prior to inclusion in the study, the data were verified by two of the authors in consensus. 20 additional native CCT scans with no abnormal acute ischemic findings were also selected. In total, 40 CCT scans from 40 patients (f = 19, m = 21, average age: 55, range: 23 – 86) were used. All CCT scans were acquired with a 64-row MDCT unit (VCT, GE Healthcare, Chalfont St. Giles, Great Britain). The contrast agent (Imeron 350, Bracco, Constance, Germany), followed by 50 ml of a physiological saline solution, was injected via an antecubital indwelling venous cannula (18G) with a flow rate of 4 ml/s using a double-head injector. The examination region extended from the apex to the base of the lung. Data acquisition was started via automatic bolus detection 5 seconds after reaching a threshold value of 100 HU in the right ventricle. The image data were reconstructed with a slice thickness of 1.25 mm and a reconstruction interval of 1.0 mm via filtered back projection using the standard algorithm.

Reader study

The CT data sets were evaluated under controlled ambient conditions by three radiologists with 3, 4, and 6 years of experience evaluating emergency CT scans. The first reading was performed on a standard PACS workstation (HP Z800, Hewlett-Packard, Palo Alto, USA) with a calibrated monitor of category B (Eizo Radiforce RS210 (21.2 inches), Eizo, Hakusan, Japan) using the 3D thin-client solution Visage 7.1 (Visage Imaging, Berlin). The readers had access to all
software functions (including windowing, zooming, measuring, and multiplanar reformation). The second reading was performed at an interval of at least 3 weeks after renewed randomization of the sequence of the data sets on a 3rd generation iPad (iPad 3, Apple Inc., Cupertino, USA) using an application for displaying radiological image data (Visage Ease, Visage Imaging, Berlin). A screenshot of the application with additional explanations is provided in Fig. 1. The iPad 3, which was released 5/2012, achieves a pixel density of 264 dpi at a resolution of 2048x1536 pixels and a screen size of 9.4 inches. The indicated application (app) allows the viewer to perform basic post-processing functions (windowing, zooming, measuring). The image data were loaded on a per case basis from the Visage server via a secure (WPA2) WLAN connection on the hospital-wide intranet. The image data were evaluated with respect to the presence of early signs of infarction (see above) or a pulmonary artery embolism on the basis of a 5-point Likert confidence scale (definitely pathological – probably pathological – uncertain – probably not pathological – definitely not pathological). The readers did not know the clinical finding or the relative proportion of pathological scans. However, the readers were informed that the frequency distribution did not correspond to the clinical routine and that the entire range of the confidence scale was to be used for the evaluation. The readers were also asked to present their subjective impressions of the ease of use of the iPad and the application.

Statistics
The Likert scale evaluations were non-parametric on the ordinal scale level. The significance test of the median scores between the iPad and the workstation were performed for all three readers using the Wilcoxon rank sum test. The Spearman’s rank correlation coefficient (Spearman’s rho) was calculated for the concordance analysis between the different readers and the two interpretation situations. Moreover, the Cohen’s kappa statistic was calculated for the different observers as well as the two different interpretation situations according to the Likert scale (definitely pathological – probably pathological – uncertain – probably not pathological – definitely not pathological – not pathological). Differences in the mean correlation coefficients as well as the kappa values were checked for statistical significance via t-test. The null hypothesis was rejected for all tests with a p-value of less than 0.05. All statistical evaluations were performed using the SPSS software package (SPSS 20, IBM, Ehningen).

Test of the iPad 3 in accordance with the quality assurance guidelines
The image display properties of the iPad 3 were tested in accordance with the requirements of the quality assurance guidelines [6] and DIN V 6868 – 57 [7] using a calibrated device for measuring luminance (LX Plus, IBA Dosimetry GmbH, Schwarzenbruck) on the basis of the SMPTE test image.

Fig. 1
The figure shows a screenshot of the iPad app, Visage Ease, which was used to evaluate the image data. The scan is a CTPA scan with evidence of a pulmonary embolism. The application makes it possible to download and view image data on the Visage PACS server. In addition to being able to scroll through a series, the following basic post-processing functions can be operated via touchscreen: windowing, zooming, and panning.
when using the iPad 3 and 0.68 (±0.10) for the interobserver agreement when using the workstation (Fig. 2). The differences were also not statistically significant (p > 0.05). According to the Likert scale, the CCT scans resulted in mean kappa values of 0.52 (±0.17) for the comparison of the two interpretation situations and 0.33 (±0.16) and 0.32 (±0.16), respectively, for the interobserver concordance on the iPad 3 and the workstation. For CTPA average kappa values of 0.67 (±0.19) were calculated for the comparison of the two interpretation situations and 0.69 (±0.08) and 0.60 (±0.14), respectively, for the interobserver concordance on the iPad 3 and the workstation (Fig. 2). All differences were not statistically significant (p > 0.05). Based on Landis and Koch, the kappa values were evaluated as follows: k < 0: poor agreement, 0 < k < 0.20: minimal agreement, 0.21 < k < 0.40: sufficient agreement, 0.41 < k < 0.60: moderate agreement, 0.61 < k < 0.80: considerable agreement and k > 0.81: almost complete agreement.

As an example, typical findings of an early sign of infarction with (a) and without (b) a finding discrepancy between the iPad 3 and the workstation are illustrated in Fig. 3a, b. In the subjective evaluation of the usability of the iPad 3, all readers stated independently that windowing via touchscreen in particular was very unfamiliar and significantly less easy to use compared to the PACS workstation.

The test of the display quality of the iPad 3 according to the quality assurance guidelines yielded a veiling luminance L_s of 0.3 cd/m², a minimum luminance L_min of 0.4 cd/m² and a maximum luminance L_max of 410 cd/m². The deviations in luminance within the image were less than ±15% compared to the center of the image and the visual image quality criteria were fulfilled. In Table 1 the data are compared to the requirements of the quality assurance guidelines and the values of the monitor used for image viewing.

Discussion

In the present study, two different interpretation tools, the iPad 3 with a high-resolution display and a 3D workstation as the gold standard, were examined in a reader study with respect to the evaluation of CT scans. The CT scans were selected on the basis of two typical issues in radiological emergency diagnosis: The detection of early signs of infarction in CCT and the diagnosis of pulmonary artery embolisms [9, 10]. The results show that the agreement of the evaluation of the two types of scan on the iPad 3 and the 3D PACS workstation does not differ from the agreement between the different readers. The kappa values calculated by us are within the range of the data published in the radiology literature. Therefore, in a metaanalysis of 15 studies for detecting early signs of infarction using native CCT, Wardlaw et al. calculated kappa values of 0.14 to 0.78 with an average sensitivity of 66% [11]. With respect to CT diagnosis of pulmonary artery embolisms, the significantly higher kappa values of 0.60–0.67 calculated by us also...
correspond to the data published in the literature. Consequently, in a prospective multicenter study, Courtney et al. calculated a kappa value of 0.75 for the interobserver agreement [12]. Kelly et al. report kappa values of 0.69 in the case of embolism on the lobe level. A limitation of the analysis to the segment and subsegment level, as in our study, resulted in kappa values of 0.54 and 0.44 [13].

The previously published studies regarding the use of the iPad for radiological diagnosis use older versions of the device with a lower-resolution display (1024 × 768px, 132ppi). These studies yielded satisfactory to good agreement with the interpretation using PACS workstations. In a study regarding the detection of 50 pulmonary embolisms using the 1st generation iPad compared to the PACS, two readers interpreted 98% of the cases correctly in both interpretation situations [14]. Mclaughlin et al. were able to show agreement in the interpretation of CCT scans in 93 of 100 cases [15]. The misdiagnoses on the iPad related to very discrete findings. The same authors used a low-contrast phantom to evaluate the display quality and only achieved results on the iPad comparable to those of a monitor (2 megapixels) by using the zoom function. It must be taken into consideration that the display of the 3rd generation iPad used in our study has a significantly higher resolution than the iPad 1 used in the indicated study. In a study by John et al., 79 CT scans and 9 MRI scans were evaluated by three radiologists on the iPad and PACS with respect to typical emergency issues and major discrepancies were found in 3.4% of cases and minor discrepancies in 5.6% [16]. In the analysis of MRI scans in spinal emergency diagnosis, McNulty et al. did not find a difference in accuracy compared to interpretation using a “secondary class” monitor (comparable with category B of the quality assurance guidelines) [17].

In agreement with the mentioned studies, we see a potential role of the iPad 3 in radiology primarily as a supplement to established PACS workstations [14 – 17]. Mobile access to image data allows quick preliminary initial assessment as well as teleconsultation (e.g. of the radiological background service) under emergency conditions. Remarkably the FDA has already approved multiple iPad applications for primary interpretation of scans if a full workstation is (temporarily) not available [5]. Approval of the iPad for the interpretation of scans currently cannot be expected in Germany. Although the iPad 3 fulfills significant parts of the currently valid quality assurance guidelines (Table 1), it does not reach the required minimum diagonal screen size of 15” (LCD). The inconstant environmental conditions associated with mobile operation are also problematic even though similar concerns on the part of the FDA were able to be resolved by software solutions for ensuring appropriate contrast detectability. In principle, the iPad 3 can determine the environmental conditions via the integrated camera and adapt the display accordingly. The display of the iPad 3 can be calibrated by the software, e.g. according to DICOM – Display Function (GSDF) [18]. The possibilities for the clinical use of the iPad 3 are currently further limited in Germany since the current quality assurance guidelines recommend monitors of category B also for the “informative viewing of already diagnosed images and for diagnostic orientation” [6].

With the upcoming amendment of the quality assurance guidelines and the pending coming into effect of new standard DIN6868 – 157 as a replacement for prestandard DINV 6868 – 57, the requirements for approval and constancy testing of image display devices will increase further [19]. The introduction of room classes for the different application scenarios represents a significant change compared to prestandard DINV 6868 – 57. Depending on the activity primarily performed in the room from “evaluation” to “repetitive reproduction of a known and evaluated finding”, the maximum allowed room illumination increases in four stages from 50 lx, ≤100 lx and ≤500 lx to ≤1000 lx. The area of application of the standard now also clearly includes image viewing. Moreover, the entire image display system, i.e., the image display device including the software and the settings of the operating system, must be included in the test. Finally, limits for pixel errors were introduced and measurement methods and test images were adapted to DIN62 563 – 1. A significant new aspect in this study is the explicit requirement of a constancy test after “major changes” to the image display system including a change of the

Tab. 1  Technical data and measured values for the iPad 3 compared to the monitor and the current quality assurance guidelines [6].

<table>
<thead>
<tr>
<th>requirements according to quality assurance guidelines</th>
<th>iPad 3</th>
<th>monitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>diagonal screen size (according to manufacturer specifications)</td>
<td>≥ 15 inches (LCD)</td>
<td>9.7 inches</td>
</tr>
<tr>
<td>matrix size</td>
<td>≥ 1000 × ≥1000px</td>
<td>2048 × 1536px</td>
</tr>
<tr>
<td>max. luminance</td>
<td>&gt; 120 cd/m² (category B)</td>
<td>&gt; 200 cd/m² (category A)</td>
</tr>
<tr>
<td>max. contrast</td>
<td>&gt; 40 (category B)</td>
<td>&gt; 100 (category A)</td>
</tr>
<tr>
<td>functions (recommended)</td>
<td>windowing, zoom, MPR, double monitor</td>
<td>depending on the software</td>
</tr>
<tr>
<td>display characteristics/ calibration capability</td>
<td>DICOM/CIE, calibration capability</td>
<td>DICOM GSDF (in principle, able to be implemented and calibrated on the software side [18])</td>
</tr>
<tr>
<td>luminance deviation</td>
<td>≤ ± 20 % from the image center (cat. B)</td>
<td>≤ ± 15 % from the image center (cat. B)</td>
</tr>
</tbody>
</table>

¹ According to current constancy test or acceptance protocol

Tewes S et al. Evaluation of the... Fortschr Röntgenstr 2013; 185: 1063–1069
room class and location changes as regularly occur in the case of mobile applications.

In addition to these stricter normative framework conditions, additional limitations must be taken into consideration when discussing the use of the iPad 3 in radiology. The device is only available with a reflective display which further emphasizes the significance of an interpretation situation with appropriate background lighting. Moreover, data protection and access rights administration are particularly important in the case of mobile devices. Software-based measures must be taken to prevent unauthorized access to patient data in the case of public networks and theft or loss of the device. In addition, the operation of the iPad 3 via touchscreen requires a novel operating concept for image viewing and also results in practical disadvantages such as increasing contamination of the screen with potential effects on image quality and in hygiene problems particularly when used in a hospital [20, 21].

Possible application areas are currently teaching, demonstration of diagnosed images to colleagues or patients, or use as a medical information system and expert system [22 – 24].

The present study has different limitations. First, the number of cases and the number of readers was limited as in every reader study. For reasons of practicability and to simplify the statistical analysis, the evaluation of the scans was limited to two areas evaluated on the basis of confidence scales. A full diagnosis was not established so that differences in the interpretation scenario with respect to the detection of secondary findings could not be evaluated.

The potential limitations of the iPad 3 in relation to the detectability of low-contrast lesions were only examined in this study on the basis of early signs of cerebral infarction and not on the basis of abdominal scans or phantom data which might show more subtle differences between the two interpretation scenarios. Moreover, the selected CT scans have a relatively high interobserver variability which is known from the literature and was also confirmed in this study [11 – 13]. However, this limitation is to be qualified by the fact that the focus of this study was typical emergency cases in which the use of the device is of clinical interest. There is no advantage to mobile image display for routine diagnostics and this was therefore not the subject of this study.

Moreover, a comparison of the time needed for diagnosis was dispensed with since a comparable routine for both tools would only have been able to be realized with substantial iPad 3 training. Therefore, it is unclear whether the handling criticism of all readers is a fundamental problem of the operation via touchscreen or is the result of a learning curve and would be resolved with increasing experience on the part of the user [15]. In addition, data transmission is extremely important when using a mobile diagnostic tool. While data transmission via a WLAN with a corresponding network connection of the WLAN router is non-critical, the use of mobile networks can result in significant limitations of the transmission rate depending on the quality of the network, thus preventing real-time diagnosis via the mobile device.

In summary, we were able to show in the present study that the variability of the interpretation of typical emergency scans on the iPad 3 and a PACS workstation do not differ from the interobserver variability. From a clinical standpoint, there are potentially attractive application areas for the tested device primarily with respect to quick preliminary evaluation on duty or for real-time consultation of a radiological background service or experts. The results only apply to the iPad 3 used in this study. The inconsistent and difficult to standardize environmental conditions when using a mobile image display device represent further limitations as is expressed in the stricter normative framework conditions.

References
4 Savel RH, Munro CL. Scalpel, stethoscope, iPad: the future is now in the intensive care unit. Am J Crit Care 2011; 20: 275 – 277


24 Franko OI, Bhola S. iPad apps for orthopedic surgeons. Orthopedics 2011; 34: 978 – 981