

# User-friendly tools on hand-held devices for observer performance study

Takuya Matsumoto, Takeshi Hara, Junji Shiraishi\*, Daisuke Fukuoka\*\*  
Hiroyuki Abe\*\*\*, Masaki Matsusako\*\*\*\*, Akira Yamada\*\*\*\*\*, Xiangrong Zhou, and Hiroshi Fujita  
Department of Intelligent Image Information, Gifu University Graduate School of Medicine

1-1 Yanagido, Gifu, Gifu 501-1194, Japan

\* Department of Radiological Technology, Kumamoto University  
4-24-1 Kuhonji, Kumamoto, Kumamoto 862-0976, Japan

\*\* Department of Technology Education, Gifu University  
1-1 Yanagido, Gifu, Gifu 501-1193, Japan

\*\*\* Department of Radiology, the University of Chicago  
5841 S Maryland Ave., Chicago, IL 60637

\*\*\*\* Department of Radiology, St. Luke's International Hospital  
9-1 Akashi-cho, Chuo, Tokyo 104-8560, Japan

\*\*\*\*\* Department of Radiology, Shinshu University School of Medicine  
3-1-1 Asahi, Matsumoto, Nagano 390-8621, Japan

## ABSTRACT

ROC studies require complex procedures to select cases from many data samples, and to set confidence levels in each selected case to generate ROC curves. In some observer performance studies, researchers have to develop software with specific graphical user interface (GUI) to obtain confidence levels from readers. Because ROC studies could be designed for various clinical situations, it is difficult task for preparing software corresponding to every ROC studies. In this work, we have developed software for recording confidence levels during observer studies on tiny personal handheld devices such as iPhone, iPod touch, and iPad. To confirm the functions of our software, three radiologists performed observer studies to detect lung nodules by using public database of chest radiograms published by Japan Society of Radiological Technology. The output in text format conformed to the format for the famous ROC kit from the University of Chicago. Times required for the reading each case was recorded very precisely.

**Keywords:** observer performance study, ROC, iPad, handheld device, confidence level

## 1. INTRODUCTION

In clinical situations, many researchers try to use handy devices such as iPad or iPhone for treatments and managements because the handy devices have enough resolutions to display medical images and good connection with computer networks to retrieve image data and clinical information. Takao et al. proposed a support system for diagnosis of brain stroke by using mobile devices for emergency medical care [1]. Nam et al. also developed a decision support tool for stroke classification [2]. Results of those studies showed that the developed software on handy devices was successful to improve the flow of information among medical staff.

ROC studies take very important roles in medical fields to compare imaging devices, system performance, and readers response changes with/without any clinical or quantitative information [3], but the studies require complex procedures to select cases from many data samples, and to set confidence levels in each selected case to generate ROC curves. In some observer performance studies, researchers have to develop software with specific graphical user interface (GUI) to obtain confidence levels from readers. Because ROC studies could be designed for various clinical

---

\*Corresponding author information: E-mail: takeshi.hara@mac.com, Telephone: +81-58-230-6511

situations, it is difficult task for preparing software corresponding to every ROC studies. The purpose of this study was to develop a helper application running on handheld devices such as iPad or iPhone/iPod touch to improve the data flow during observer performance studies.

## 2. MATERIALS AND METHODS

We have developed software for obtaining confidence levels on tiny personal handheld devices such as iPhone, iPod touch, and iPad (Fig.1). This software records the confidence levels by sliding a finger tip on digital scale bar and the location(s) where readers suspected on the screen according to case lists. The number of locations was not limited to one because FROC analysis allows multiple locations to be pointed as suspect area. The display shows two thumbnail images in JPEG/TIFF/PNG/BMP format as indicated in a case list file that the order of the image were written in CSV format.

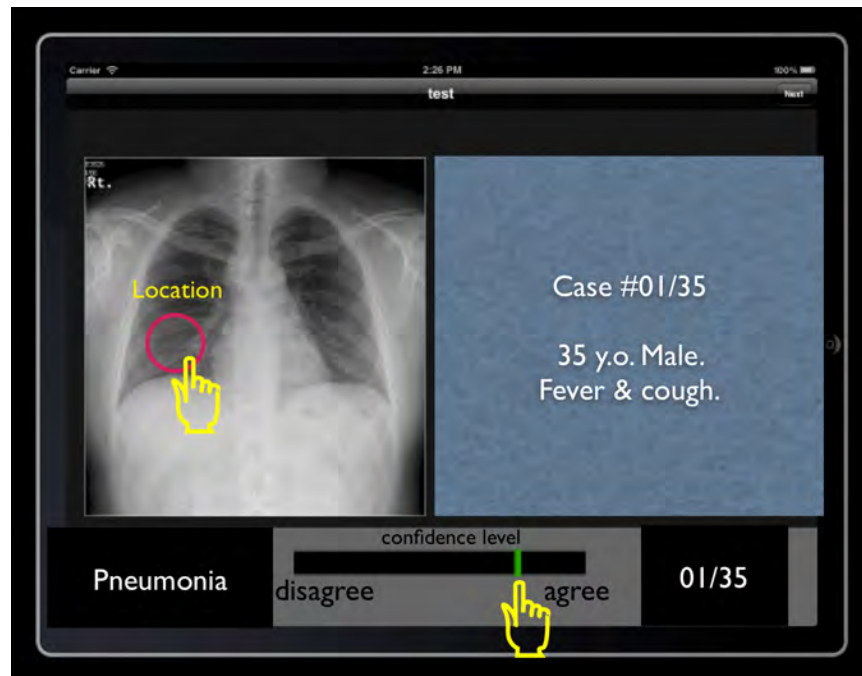


Fig.1 Example of the developed application on iPad. Reader can record the confidence level (below) and suspect location(s).

Mailadress	me@info.gifu-u.ac.jp						
Lfilenames	Rfilenames	Truth	Modality	Part	Number	memo1	memo2
LN001.jpg	readoriginal.png	P	test		1 a		
LN001.jpg	LN001cad.png	P	cad		1 a		
LN002.jpg	readoriginal.png	P	test		2 a		
LN002.jpg	LN002cad.png	P	cad		2 a		
LN201.jpg	readoriginal.png	N	test		3 a		
LN201.jpg	LN201cad.png	N	cad		3 a		
LN202.jpg	readoriginal.png	N	test		4 a		
LN202.jpg	LN201cad.png	N	cad		4 a		

Fig.2 Example of the caselist file to set up the image order on the device.

### 2.1 Caselist file

A case list file is required to show the thumbnail images on it, and is transferred from PC/Mac to the device via iTunes or URL (Uniformed Resource Locator) on configuration window on the device. Figure 2 shows an example of caselist file in CSV format. The content next to Mailadress indicate a email address to send the result of the study. The result file contains a log file to record the reading time and a list of confidence levels with ROC kit compatible type after the confidence levels were sorted based on the truth, the modality, the part, and the number indicated in the caselist file.

The column in Lfilenames and Rfilenames indicate the filenames that displayed on the left and right on the device, respectively. The image files must be transferred from PC/Mac via iTunes and be stored in the device. URLs are also allowed on the cells to obtain the image data from web servers, if the storage in local area of the device are not preferable. The column in Truth indicates the truth of the case in the row. P and N mean positive and negative cases, respectively. The column in the Modality column indicates the system difference such as CT/MR, XP/CP, or With CAD/Without CAD. The numbers in Part column identify the pair of case combinations to compare each other. The letters in Number column identify the experiment session because the system allows multiple session during one reading experiment. Memo1 and Memo2 have no influence for the analysis and allow any comments to help the analysis from the researchers.

## 2.2 Logfile

Precise information of reading time is hard to obtain during image interpretation. The developed software can record the time when the observers made their decisions as results of determination of their confidence levels for the case. Figure 3 (a) shows the logfile example. After the information of observer's agreement for the experiment, the filenames to be displayed on the device were recorded, and the time when the images were presented to the readers. After the reader determined the confidence level for the case, the confidence level and locations (if indicated) in the x-y coordinate on left image on the device were recorded. The logfile was stored on the device during the reading session, and it sent to the email address on the caselist file after the session was completed.

## 2.3 ROCfile

The Caselist file includes the information whether the case is positive or not, which modality the case is. Based on those characters in columns of Truth and Modality, the system automatically sorts the confidence levels to fit the ROC kit [4-6]. Figure 3 (b) shows an example from an experiment with continuous confidence levels for an observer study with/without CAD information. By using the information in Part column, the pair of confidence levels were aligned in the same case in the row. Asterisk (\*) separates the confidence level list into two groups of positive and negative based on the Truth column.

```
02/01/2012 10:24:22 The reader agrees
with the observer study.
case001.png
q001.jpg
Start:02/01/2012 10:35:22
Confidence level: 0.967
Location: 104, 588
End : 02/01/2012 10:35:49

case002.png
q002.jpg
Start:02/01/2012 10:35:49
Confidence level:0.967
End : 02/01/2012 10:36:12
```

(a) Logfile

```
The result of chest study
KIT
"Without CAD" "With CAD"
CLL      CLL
0.16     0.16
0.64     0.68
0.90     0.94
0.32     0.62
0.18     0.25
*
0.02     0.10
0.22     0.45
....
```

(b) ROCfile

Fig.3 Example of logfile with confidence levels and clocked reading time and ROCfile compatible for ROCKIT.

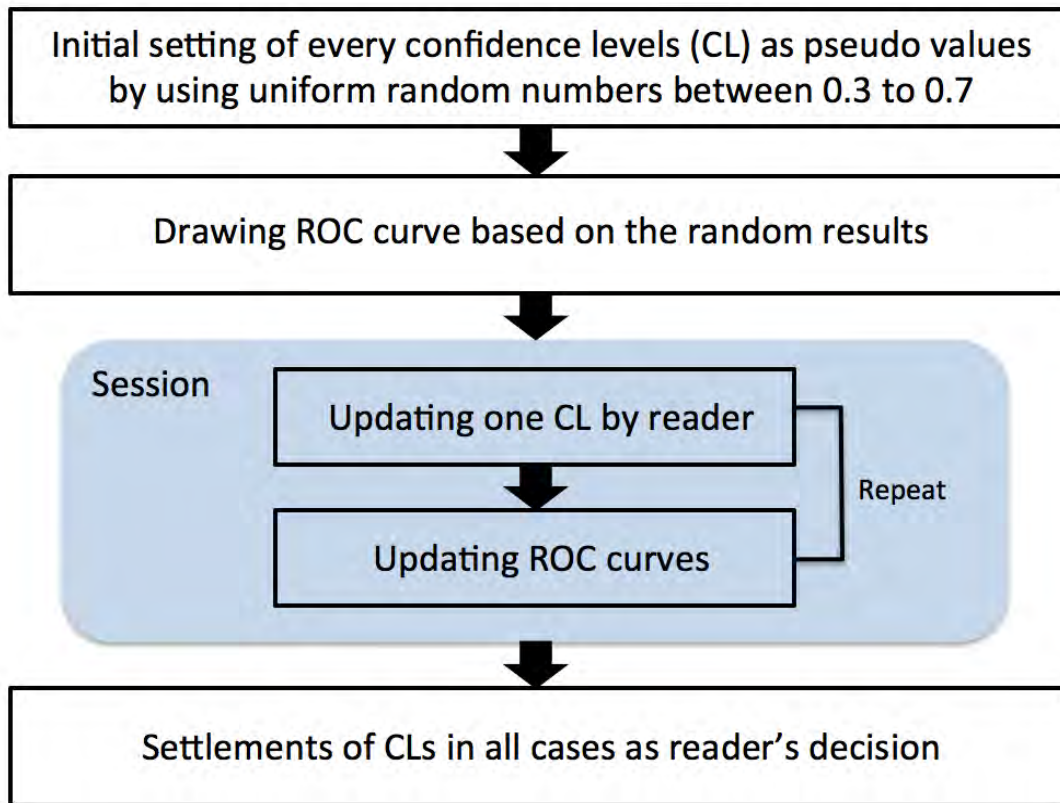


Fig.4 Overview of real-time generation of ROC curves during interpretation session.

#### 2.4 Real-time generations of ROC curves

Presenting ROC curves during observer performance studies will be a great bias to interfere with readers' decisions because the readers can understand the truth of the presented case by confirming the increment or the decrement of the curve, and they suppose whether the next case is positive or negative after the same truth of positive or negative was continued. However, the presentation of ROC curves, when reading of a case was concluded, will be a good motivation for readers to practice the interpretation skills since readers can estimate the truth of the presented case by checking the changes of ROC curves. The derived important value of AUC (area-under-the-curve) can be used to compare the current reading results with previous ones statistically if the case and the reading conditions were the same. Figure 4 shows the overview of the real-time curve generation. First, all of the value for confidence level in each case were set by using uniform random numbers between a range of 0.3 to 0.7 as the pseudo values to represent guessing results. In consequence of this initialization of confidence levels, the area-under-the-curve (AUC) of first ROC curve will be 0.5. The range at the initial setting is very important to reflect reader's confidence levels because readers tend to set their confidence levels as low or high score in the scale. If the range was wide enough, the curve will not changed very often because the reader's confidence levels sometimes equal to the initial random value. Figure 5 shows the concept of the distributions of confidence levels and the generated ROC curves from each distribution. The initial distribution of confidence level has a uniform shape, so that the ROC curve has an AUC value around 0.5. If one case was updated by reader correctly as shown on the figure, the confidence level of the positive case was increased, the AUC value also increased from 0.53 to 0.58. Readers can recognize the changes of curves and the AUC value as their decisions were correct because the AUC value was increased. The change will be a good motivation for interpretation training.

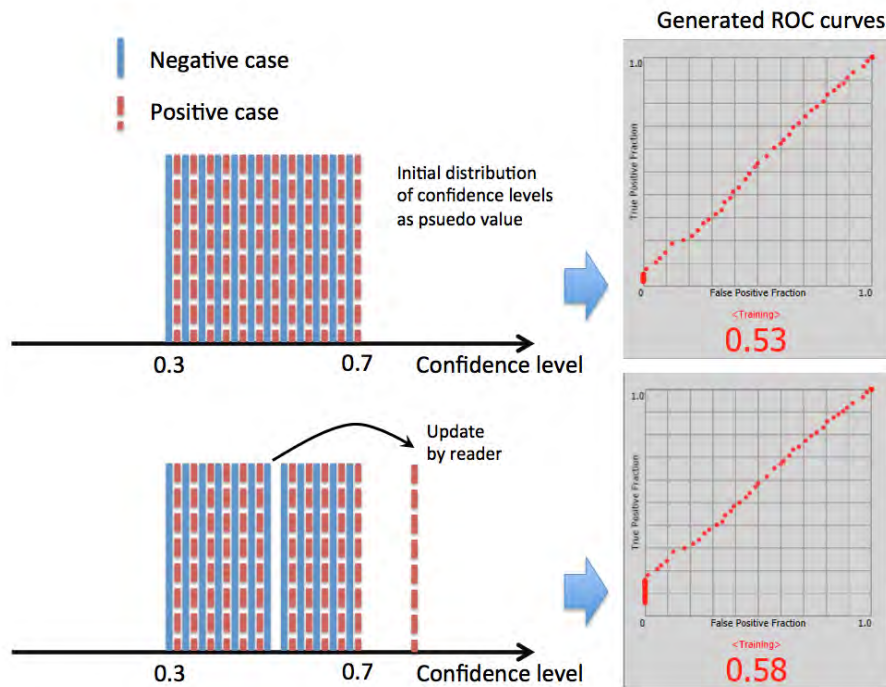


Fig.5 ROC curves and distributions (concept) of confidence levels at initial setting by using uniform random value.

### 3. RESULTS AND DISCUSSIONS

We carried out two evaluation tests to confirm the functions of our software. First, we set up an observer test for evaluating the usefulness of chest CAD system. In this study, three radiologists performed observer studies to detect lung nodules by using public database of chest radiograms published by Japan Society of Radiological Technology (JSRT). Figure 6 shows a situation of an observer test by using the developed tools. The software has no connection to the interpretation workstation in clinical situations. Readers have to change the case number according to the message from the device. It may be a hard task to select the indicated case in the list, but if the experiment design has its own case list on DICOM server, we suppose reader can find desired cases in very short time by comparing with the message and thumbnail images on the device. The output in text format conformed to the format for the famous ROC kit from the University of Chicago. The outputs of confidence level from multi-readers could be also employed to determine the difficulty of the case selection before observer studies were performed. The image database from JSRT includes the subtlety information of obvious, relatively obvious, subtle, very subtle, and extremely subtle in every chest image based on 20 radiologists interpretation results. Comparing with current procedure to record confidence levels of readers on paper by using marker pencils, the proposed method by using the device represents an advance not only in easy handling of the data conversion but also in the time clocking during interpretation. By analyzing the interpretation time in our logfile after the interpretation experiment, table 1 shows that the more subtle nodule radiologists read, the longer time they require. The time difference depending on the case subtlety was not obtained easily before the developed tool recorded the accurate interpretation time.

Another evaluation test was performed by setting up an educational system for chest image reading. The system includes presentations of a thumbnail image and a question on the device as shown on Fig. 7. The real-time generation of ROC curves was also included. Readers interpret the presented image on clinical workstation as the tool indicated the case number with the thumbnail image. The question and clinical information for the image was shown on the right on the device. The readers made their decisions as a confidence level on the device. After the determination of readers' decisions, the correct interpretation with other modality images, the ROC curves, and the AUC value were presented for the educational purpose.

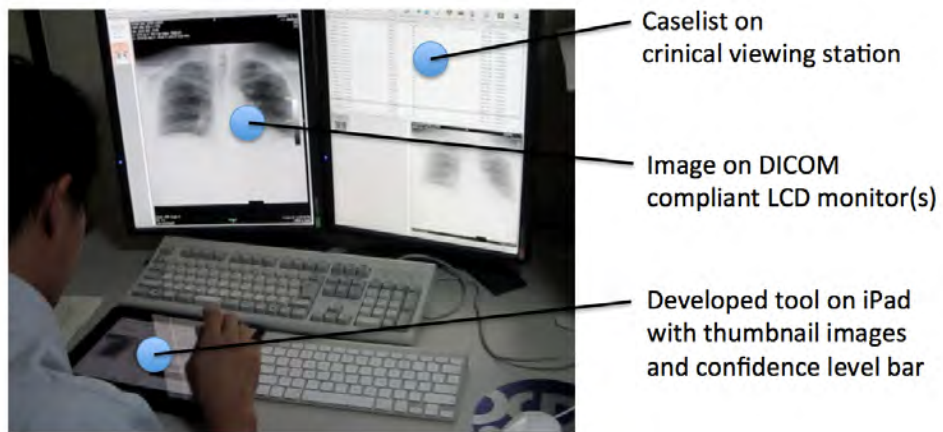


Fig.6 Overview of observer performance study by using developed tools in front of a clinical viewing station.

Table 1 Time differences depending on subtlety of lung nodules on chest radiograms from JSRT database.

	Without CAD	With CAD
Extremely subtle	30 sec / image	+9 sec / image
Very subtle	26 sec / image	+8 sec / image
Subtle	24 sec / image	+7 sec / image
Relatively obvious	18 sec / image	+4 sec / image
Obvious	15 sec / image	+5 sec / image

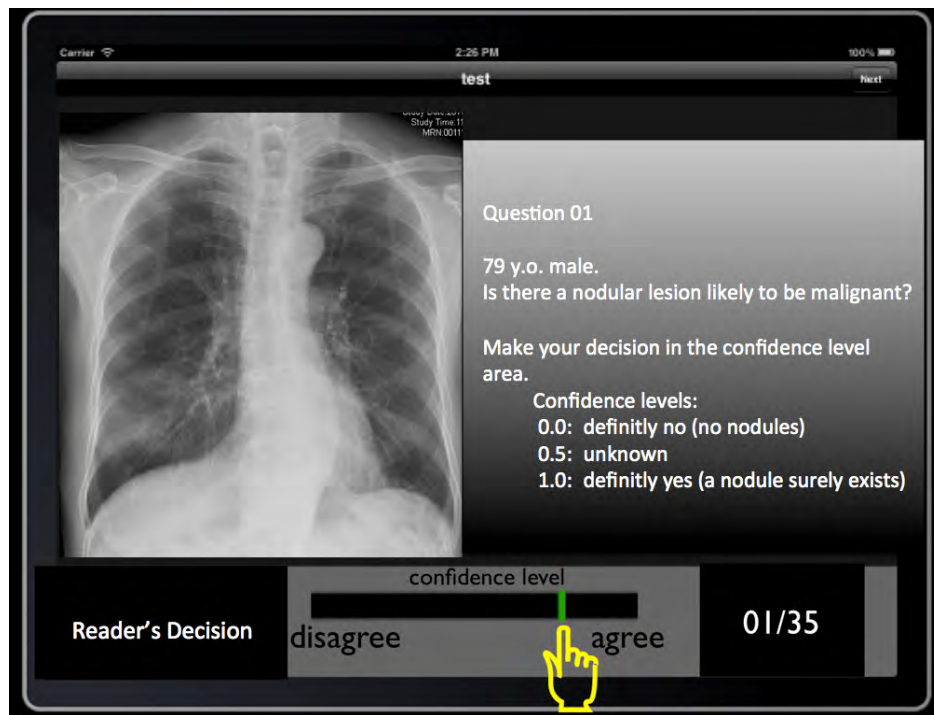


Fig.7 Overview of educational system for chest reading by using developed tools.

#### 4. CONCLUSION

We have developed measuring software of confidence levels on tiny personal handheld devices such as iPhone, iPod touch. The fundamental functions for ROC analysis were verified. As a result, this application will be a useful tool for ROC studies. This application may be used in observer performance studies of all modalities and be a helpful tool of studies by using clinical interpretation workstations. The procedure of real-time generation of ROC curves and presentation the AUC values were effective not only to confirm the change of the curves before the formal observer study, but also to encourage readers to maintain their motivations to obtain interpretation skills.

#### ACKNOWLEDGMENT

This work is partly supported by a research grant from Japanese Society of Radiological Technology (JSRT), a research grant of Grant-in-Aid for Scientific Research on Priority Areas, and a Grant-in-Aid for Scientific Research on Innovative Areas, which is awarded by MEXT, Japanese. The software has been developed by the working group belong to JSRT.

#### REFERENCES

- [1] H. Takao, Y. Murayama, T. Ishibashi et al., "A new support system using a mobile device (smartphone) for diagnostic image display and treatment of stroke," *Stroke*, 43(1), 236-9 (2012).
- [2] H. S. Nam, M. J. Cha, Y. D. Kim et al., "Use of a handheld, computerized device as a decision support tool for stroke classification," *Eur J Neurol*, (2011).
- [3] J. Shiraishi, L. L. Pesce, C. E. Metz et al., "Experimental design and data analysis in receiver operating characteristic studies: lessons learned from reports in radiology from 1997 to 2006," *Radiology*, 253(3), 822-30 (2009).
- [4] C. E. Metz, "Basic principles of ROC analysis," *Semin Nucl Med*, 8(4), 283-98 (1978).
- [5] C. E. Metz, "ROC methodology in radiologic imaging," *Invest Radiol*, 21(9), 720-33 (1986).
- [6] L. L. Pesce, and C. E. Metz, "Reliable and computationally efficient maximum-likelihood estimation of "proper" binormal ROC curves," *Academic Radiology*, 14(7), 814-829 (2007).
- [7] J. Shiraishi, S. Katsuragawa, J. Ikezoe et al., "Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," *AJR Am J Roentgenol*, 174(1), 71-4 (2000).