

ROCView: prototype software for data collection in jackknife alternative free-response receiver operating characteristic analysis

Abstract

ROCView has been developed as an image display and response capture (IDRC) solution to image display and consistent recording of reader responses in relation to the free-response receiver operating characteristic paradigm. A web-based solution to IDRC for observer response studies allows observations to be completed from any location, assuming that display performance and viewing conditions are consistent with the study being completed. The simplistic functionality of the software allows observations to be completed without supervision. ROCView can display images from multiple modalities, in a randomised order if required. Following registration, observers are prompted to begin their image evaluation. All data are recorded via mouse clicks, one to localise (mark) and one to score confidence (rate) using either an ordinal or continuous rating scale. Up to nine “mark-rating” pairs can be made per image. Unmarked images are given a default score of zero. Upon completion of the study, both true-positive and false-positive reports can be downloaded and adapted for analysis. ROCView has the potential to be a useful tool in the assessment of modality performance difference for a range of imaging methods.

Manuscript

The success of an imaging technique should not be judged by physical measures, such as signal-to-noise ratio or contrast resolution, alone. Human observations must be considered because the reader is an integral part of the diagnostic process. For the latter, receiver operating characteristic (ROC) methods successfully quantify the combined performance of imaging techniques and readers [1].

Traditional ROC analysis simply required readers to state whether they thought a case was normal or abnormal, using a rating scale to indicate decision confidence [2]. A notable development in ROC methodology has been the free-response ROC (FROC) method. Unlike ROC methods, FROC uses location information to resolve ambiguities in detection (lesion or lesion mimic) to prevent false identification of pathology, resulting in a true-positive result [3].

Jackknife alternative free-response ROC (JAFROC) methods [4] are the latest evolution of FROC methods in which multiple readers and modalities are compared with greater sensitivity to differences in performance from other ROC methods [5]. Table 1 lists the terms associated with JAFROC analysis. JAFROC methods demand a precise response from the reader, in which location and confidence information must be supplied for each lesion within a case [6]. This identification and scoring is typically referred to as a “mark-rating” pair [7]. Those setting up JAFROC studies are encouraged to allow mouse clicks to record reader responses and an acceptance radius around each area of pathology to classify each response as either a lesion localisation/true positive (LL/TP) or a non-lesion localisation/false positive (NL/FP) [8].

JAFROC methods can be valuable in the evaluation of different techniques in radiology. Like all observer performance studies, achieving optimal statistical power can require a large number of readers and cases [9, 10], and a 50:50 ratio of normal and abnormal cases [11]. However, JAFROC methods achieve greater statistical power than ROC for the same combination of readers and cases [7]. This in turn requires a reliable method of image display and response capture (IDRC). All ROC methods require accurate data recording; ROCView is proposed as a computer-based solution to accurate IDRC in JAFROC analysis—the current end point of the free-response paradigm.

Why develop ROCView?

A recent study of dose optimisation in CT [12] prompted investigation into a reliable solution for IDRC suited to JAFROC methodology. Although other software exists [13, 14], the authors took the opportunity to develop ROCView to suit the needs of their current research. As a web-based service ROCView would have excellent availability with no software download or installation required. A requirement of ROCView would be to produce data suitable for analysis via JAFROC [15] and Dorfman–Berbaum–Metz Multi-reader Multi-case (DBM-MRMC) software [16] for direct comparison of the performance of multiple readers and modalities.

Gathering suitable readers in a single centre for an observer study also presented itself as a boundary to conducting JAFROC research. Consequently, a geographically independent method of IDRC was required to allow readers to complete the observation—this prompted the development of a web-based solution. ROCView meets the demands of geographical independence and IDRC with the option to store resultant data in a form suitable for DBM-MRMC analysis. The nature of the software solution required a simplistic method to maintain the integrity of the data, as there would be no on-site support. Figure 1 describes the process of perception that readers will use when assessing studies on ROCView.

Design and development of ROCView

ROCView is implemented as a web application, with reader evaluations performed in a front-end display system (browser/client), with a web server supplying image data and recording reader responses into a database. A second web interface provides administration facilities: user management, modality and case management, creation of the “truth” and report generation. Guidance provided by Chakraborty [17] in relation to conducting a JAFROC study aided the design of ROCView. The development of a successful IDRC system required an understanding of the following key principles:

- image display
- comparing modalities
- setting the “truth”
- acceptance radius criterion N scoring confidence
- data recording.

The following description and the flow diagram shown in Figure 2 explain the process for setting up and running a JAFROC study on ROCView.

Image display

Image display is fairly straightforward for a JAFROC study. Cases from multiple modalities are displayed, with quick progression between them. ROCView automatically randomises the images from all modalities in order to reduce case memory. As a web-based application it is important to ensure that the viewing conditions are consistent and to a certain standard. A variety of monitor test patterns can be used to ensure monitor performance [18], in which it is also important to ensure a consistent display response to allow a consistent response by the observer. Methods to assess the adequacy of display performance in accordance with the digital imaging and communication in medicine (DICOM) greyscale standard display function (GSDF) are available [19].

Comparing modalities

To allow direct comparison of two modalities, such as a variation in image acquisition parameters, ROCView allows the upload of a large number of image data sets via file transfer protocol (FTP) [20]. All uploaded modalities should be named or numbered appropriately and all cases should be numbered and case-matched across all modalities. When assigned to the appropriate study, cases are automatically randomised. From this point the study is ready for the “truth” to be set.

Setting the “truth”

A suitably trained person should access the study as an administrator and create the “truth”, which is stored in a database on ROCView to act as a standard with which all further reader responses are compared, via the implementation of a pre-defined acceptance radius arising from the pixel that is marked on the image. When performing this task the administrator is able to view the modality and case identifier alongside the image, together with the localisation marker showing the acceptance radius, in order to minimise error in lesion localisation. This is particularly useful for phantom studies in which the lesion sites are known. All readers are blinded to modality and case identifiers. The administrator can make up to nine lesion localisations per case.

Acceptance radius criterion

An acceptance radius is a region surrounding a lesion that allows slight error in localisation. ROCView allows adjustment of the acceptance radius in the source code. The acceptance radius for current studies, determined by lesion size, is nine pixels' radial distance from the centre of the localisation made by the administrator creating the "truth". All marks within this radius are classified as LLs. All other marks are NLs, penalising the reader for inaccuracies and false localisations. This is, however, dependent on the resolution of the monitor and the image size. Currently, the largest size of image used has been 5126512 pixels, which allows for a 1.8% error with an acceptance of 9 pixels in all directions from the point of localisation. If monitor resolution does not allow the maximum image size then it will be scaled, with a 100% viewing window, and centred over the cursor, available in the top right of the screen. Figure 3 explains the implication of acceptance radius size; a clinical example would equate to a vessel (lesion mimic) next to a lesion in the thorax being incorrectly localised but being classified as LL owing to it falling within the acceptance radius of the lesion. Different values of radii clearly affect reader performance [21], where a relaxed criterion (20–40 pixels) improved apparent reader performance. Localisation accuracy is an important distinction between conventional ROC and free-response methods; consequently, this criterion, at the discretion of the researcher, must be controlled to ensure that JAFROC maintains precision. When a reader clicks on the image in ROCView a 19-pixel-diameter marker appears to aid the reader's judgement of accuracy (Figure 4). Existing phantom-based ROCView studies have not used simulated lesions that are .19 pixels in diameter. Therefore, localisation of the centre of the lesions would always result in a TP/LL result.

Scoring confidence

Many variations of confidence scales are described [22–24] and one must be mindful that one scale may not be suited to all investigations. To address this, ROCView offers two types of scale: continuous and discrete (Figure 5). When creating a study the administrator can choose which type of scale will be presented to the reader following a mark made on an image. The discrete scale appears as a pop-up box showing a series of ranked radio buttons (1–5, low to high confidence) from which readers can make their selection. The continuous scale presents as a slider bar that can function as a 101-point scale similar to those in use [24]. The data can be resampled to an 11-point discrete rating scale if there is not a good distribution of responses—a problem consistent with this type of scale arising from the inability to estimate within a few percentage points [23].

Data recording

Once confidence has been scored, mark-rating pairs are stored in a database (TP or FP) in the format shown in Tables 2 and 3, which can then be downloaded and analysed via JAFROC methods.

For each successful localisation five integer values are recorded in the TP sheet. Four integers are recorded for each incorrect localisation in the FP sheet. The following data need to be recorded as identifiable integers:

- reader ID
- case ID
- modality ID
- lesion ID (1, 2, 3 etc.)—not used in the FP worksheet N confidence scale rating (TP or FP).

When no mark-rating pairs are made, a default score of zero is stored for each case and lesion. These results are not required for JAFROC analysis but have been included in the reports to validate the results recorded by ROCView.

Currently the reader ID is recorded as the login details (email address) of the reader. Once the data have been downloaded, this can easily be converted to an integer in Microsoft Excel (Microsoft Excel, Redmond, WA). Once downloaded, TP and FP results need correlating with a “truth” sheet in a single Excel file as described by Chakraborty and Yoon [25]. This data file can then be analysed using JAFROC v. 4.0 [15] to transform the FROC data to inferred ROC data (*.lrc file), which can be analysed using DBM-MRMC software [16].

Software development

When the application starts, it makes an asynchronous request to the web server to retrieve the modalities to be evaluated by the user. This is dictated by the activation key used when registering as a user, since ROCView runs multiple studies concurrently. Cases are displayed in turn, and, as the user creates mark-rating pairs, each response is sent asynchronously to the server and is stored immediately in the appropriate database (TP or FP). The software has been made functional for general purpose personal computers, in terms of the processing power and memory required, as a result of the tools used to develop the software (Table 4). As with all FROC studies, monitor performance and viewing conditions remain a concern and these should meet the required standards consistent with the study objectives.

The client has been designed to use the native HTML and JavaScript capabilities of browsers. Consequently, browser plugins are not necessary and the hardware requirements of the client remain low. The native solution is at risk of manipulation by the user; for example, by manipulation of the page uniform resource locators or by reloading the evaluation page in unsupervised studies. Therefore, a function is applied to the evaluation data to maintain integrity between server requests. This prevents the user from reloading the evaluation page and returning to evaluate previous images while ensuring that evaluations resume at the last stored point if they are not completed in one attempt (due to loss of internet connection or computer hardware failure).

Future application, flexibility and development

ROCView has the potential to deal with any images acquired in a manner suitable for FROC analysis. With the capability to display images from a wide range of acquisition methods (CT, ultrasound, MRI, nuclear medicine, mammography and computed/digital radio- graphy) there is a relatively unlimited application for lesion detection-based studies. A phantom-based lesion detection study of CT acquisition parameters highlighted the type of research study for which this software is suitable [12]. Producing case-matched images is a significant challenge for MRMC studies, but if this can be achieved ROCView can be a powerful tool for the display of images and the accurate recording of reader responses.

A significant development to the software would see the creation of an acceptance volume, to allow classification of LLs and NLs, for lesions that cross multiple cross-sectional images. Software to allow this is available [14] and has been used in a FROC study of mammography techniques [34]. A further development would allow hybrid images to be displayed in their component parts (emission, transmission and fused) to allow novel FROC studies of diagnostic performance in single photon emission CT and positron emission tomography/CT. There is also scope for the development of a region-of-interest style acceptance criterion to enable a form of JAFROC analysis suited to fracture detection in which the over-riding

outcome would be a study in a similar vein to mammography PERFORMS [35] to quantify trainee performance in fracture detection. ROCView, as a web-based service, requires no download of software. ROCView is currently being used in a research programme but may become available commercially to other researchers in the future.

References

1. Gur D, Rockette H, Bandos A. 'Binary' and 'non-binary' detection tasks: are current performance measures optimal? *Acad Radiol* 2007;14:871–6.
2. Manning D. Evaluation of diagnostic performance in radiography. *Radiography* 1998;4:49–60.
3. devchakraborty.com [homepage on the internet]. Philadelphia: Dev Chakraborty; 2011 [updated 2011; cited 4 November 2011]. Available from: <http://www.devchakraborty.com/Historical%20Background.html>
4. Chakraborty D. Analysis of location specific observer performance data: validated extensions of the jackknife free-response (JAFROC) method. *Acad Radiol* 2006;13:1187–93.
5. devchakraborty.com [homepage on the internet]. Philadelphia: Dev Chakraborty and Hong-Jun Yoon; 2011. JAFROC 4.0 user manual [updated 2010; cited 10 August 2011]. Available from: <http://www.devchakraborty.com/Receiver%20operating%20characteristic.pdf>
6. Zarb F, Rainford L, McEntree M. Image quality assessment tools for optimisation of CT images. *Radiography* 2010;16: 147–53.
7. Chakraborty D. Validation and statistical power comparison of methods for analysing free-response observer performance studies. *Acad Radiol* 2008;15:1554–66.
8. Chakraborty D, Hong-Jun Y, Mello-Thomas C. Spatial localization accuracy of radiologists in free-response studies: inferring perceptual FROC curves from mark-rating data. *Acad Radiol* 2007;14:4–18.
9. Obuchowski N. Reducing the number of reader interpretations in MRMC studies. *Acad Radiol* 2009;16:209–17.
10. Chakraborty D. How many readers and cases does one need to conduct an ROC study? *Acad Radiol* 2011;18:127–8.
11. Manning D, Leach J, Bunting S. A comparison of expert and novice performance in the detection of simulated pulmonary nodules. *Radiography* 2000;6:111–16.
12. Thompson J, Hogg P, Szczepura K, Tootell A, Sil J, Manning D. Determination of optimal CT exposure factors for lung lesions using an anthropomorphic chest phantom for SPECT-CT. *Eur J Nucl Med Mol Imaging* 2010;37(Suppl. 2):S494.
13. Jacobs F, Zanca R, Oyen H, Bosmans J. Sara, an advanced software platform for human observer performance experiments. In: Proceedings of the XIVth Medical Image Perception Conference; 9–12 August 2011; Dublin, Ireland. Available from: <http://home.comcast.net/~eakmips/update.htm> (software available from: <http://www.kuleuven.be/radiology/lucmfr/sara/index.html>)
14. Hakansson M, Svensson S, Zachrisson S, Svalkvist A, Bath M, Mansson L. ViewDEX: an efficient and easy-to-use software for observer performance studies. *Radiat Prot Dosim* 2010;139:42–51. Software available from: http://helios.ifss.gu.se/viewdex/index_v3_slide0001.htm
15. devchakraborty.com [homepage on the internet]. Philadelphia: Dev Chakraborty; 2011. JAFROC 4.0.1 analysis software [updated 2011; cited 10 August 2011]. Available from <http://www.devchakraborty.com/downloads.html>

16. perception.radiology.uiowa.edu [homepage on the internet]. Iowa City, IA: The Medical Image Perception Laboratory; 2011. DBM-MRMC v. 2.3 [updated 2011; cited 10 August 2011]. Available from: <http://perception.radiology.uiowa.edu/Software/ReceiverOperatingCharacteristicROC/DBMMRMC/tabid/116/Default.aspx>
17. devchakraborty.com [homepage on the internet]. Philadelphia: Dev Chakraborty; 2011. How to conduct a free-response study [updated 2011; cited 10 August 2011]. Available from: <http://www.devchakraborty.com/HowToConduct.html>
18. Deckard.mc.duke.edu [homepage on the internet]. Durham, NC: American Association of Physicists in Medicine; 2011. Online report no. 3: assessment of display performance for medical imaging systems [updated 2005; cited 10 August 2011]. Available from: http://deckard.mc.duke.edu/_samei/tg18_files/tg18.pdf
19. Fetterly K, Blume H, Flynn M, Samei E. Introduction to grayscale calibration and related aspects of medical imaging grade liquid crystal displays. *J Digit Imaging* 2008;21:193–207.
20. filezilla-project.org [homepage on the internet]. Aachen, Germany. Filezilla; 2011. Filezilla 3.5.0: The free FTP solution [updated 2011 May 22; cited 10 August 2011]. Available from: <http://filezilla-project.org/>
21. Gur D, Bandos A, Klym A, Cohen C, Hakim C, Hardesty L, et al. Agreement of the order of overall performance levels under different reading paradigms. *Acad Radiol* 2008;15:1567–73.
22. Park S, Goo J, Jo C-H. Receiver operator characteristic (ROC) curve: practical review for radiologists. *Korean J Radiol* 2004;5:11–18.
23. Hadjiiski L, Chan H-P, Sahiner B, Helvie M, Roubidoux M. Quasi-continuous and discrete confidence rating scales for observer performance studies: effects on ROC analysis. *Acad Radiol* 2007;14:38–48.
24. Rockette H, Gur D. Selection of a rating scale in receiver operating characteristic studies: some remaining issues. *Acad Radiol* 2008;15:245–8.
25. devchakraborty.com [homepage on the internet]. Philadelphia: Dev Chakraborty; 2011. JAFROC 4.0 user manual [updated 2010; cited 10 August 2011]. Available from: <http://www.devchakraborty.com/downloads.html>
26. apache.org [homepage on the internet]. Los Angeles, CA: Apache Software Foundation; 2011. Apache HTTP server [updated 22 May 2011; cited 10 August 2011]. Available from: <http://httpd.apache.org/>
27. php.net [homepage on the internet]. Denmark: PHP; 2011. PHP Scripting Language [updated 28 June 2011; cited 10 August 2011]. Available from: <http://www.php.net/>
28. mysql.com [homepage on the internet]. Redwood Shores, CA: MySQL; 2011. Open source database [updated 25 July 2011; cited 10 August 2011]. Available from: <http://www.mysql.com/>
29. adobe.com [homepage on the internet]. San Jose, CA: Adobe; 2011. Flex 4 SDK [updated 20 June 2011; cited 10 August 2011]. Available from: <http://opensource.adobe.com/wiki/display/flexsdk/Flex+SDK>
30. jquery.com [homepage on the internet]. Boston, MA: jQuery; 2011. JavaScript Library [updated 23 March 2010; cited 10 August 2011]. Available from: <http://jquery.com/>
31. ant.apache.org [homepage on the internet]. Los Angeles, CA: Apache; 2011. Apache Ant Java Library [updated 27 December 2010, cited 10 August 2011]. Available from: <http://ant.apache.org/>
32. jslint.com [homepage on the internet]. USA: JSLint; 2011. JavaScript Code Quality Tool [updated 26 August 2011, cited 10 August 2011]. Available from: <http://www.jshint.com/>
33. crockford.com [homepage on the internet]. USA: JSMIn; 2011. JavaScript Minifier [updated 2003 Dec 04, cited 10 August 2011]. Available from: <http://www.crockford.com/javascript/jsmin.html>

34. Svahn T, Andersson I, Chakraborty D, Svensson S, Ikeda D, Förlin D, et al. The diagnostic accuracy of dual-view digital mammography, single-view breast tomosynthesis and a dual-view combination of breast tomosynthesis and digital mammography in a free-response observer performance study. *Radiat Prot Dosim* 2010;139:113–17.
35. Scott H, Gale A. Breast screening: PERFORMS identifies key mammographic training needs. *Br J Radiol* 2006;79:S127–33.