Region Marking Software Tool for Medical Images

Dinu Dragan and Dragan Ivetic
Computing and Control Department
Faculty of Technical Sciences, University of Novi Sad
Novi Sad, Serbia
e-mail: {dinud, ivetic}@uns.ac.rs

***Abstract*—Subjective quality evaluation of medical images is mostly achieved using some version of ROC (Receiver Operating Characteristic) evaluation. In the majority of studies ROC analysis is the focus of research. The software tool used to mark medical images and to gather diagnostic data is either not described or described only briefly. During our research on quality evaluation system for medical images used in PACS system (Picture Archiving and Communication System), we decided to implement such a software tool. We decided to base our solution on web technologies. One of our goals was to enable in a single application on site evaluation of medical images in a local hospital and remote evaluation of medical images. In the paper, we report our goals, design decisions, and we describe the current version of the region marking software tool. The presented software tool is more than just an additional tool for subjective quality evaluation of medical images because it can be used as a software platform for education in diagnostic imaging.**

Keywords-eHealth application; medical image quality; ROC evaluation; medical image marking

I. INTRODUCTION

The quality of a digital medical image (henceforth, medical image) is an important issue in medical imaging because it directly influences the procedures based on medical images. Medical image quality can be degraded due to several sources, from the acquisition process to image compression, noisy channels and so on [1]. Therefore, it is useful to evaluate the quality of medical images.

In medical imaging, quality is evaluated using domain specific subjective evaluations [2]. Medical image accuracy in a diagnostic task is measured and quantified. Medical image influence on diagnostic accuracy is evaluated by measuring whether the same results are achieved compared to some form of gold standard [3]. The same diagnostic task is conducted using an original reference image and the evaluated medical image. The results are usually quantified using ROC analysis or some of its variations [4]. ROC evaluation is the most frequently used subjective evaluation of medical image quality [5][6]. It is conducted in two phases:

1) Data gathering – observers performe a diagnostic task on medical images by markin regions of interest.
2) Data analysis – gathered data are processed using ROC analysis**.**

In the *data gathering* phase of the ROC evaluation, depending on the variation of the ROC analysis used, it is necessary to use additional software for performing observers' diagnostic task in the manner suitable for ROC evaluation [7]. What surprised us during our evaluation of technical literature is the lack of description for this kind of software [4][8]. Also, we did not manage to find any free software for the *data gathering* phase of the ROC evaluation. Therefore, during our research on a quality evaluation system for medical images in PACS systems [9][10] we were forced to implement a region marking software tool for ROC evaluation of medical images. We had several goals in mind:

- It should resemble software tools used in everyday clinical tasks.
- It should be accessible from many places.
- It should be usable in education.
- It should be free.

Therefore, we decided to implement it using web technologies. In the paper we describe our ideas and present the current version of the software tool.

The organization of the paper is as follows. The background is described in Section 2. Software design and tools used in implementation are presented in Section 3. The current version of the region marking software tool for medical images is described in Section 4. Section 5 concludes the paper.

II. BACKGROUND

There are two approaches to evaluate image quality in general [5][8]:

- Objective evaluation – based on a mathematical or a statistical model, which is easy to compute, rate, and implement on a computer.
- Subjective evaluation – based on a subjective evaluation of restored images by single or multiple observers.

However appealing, objective evaluation did not replace subjective quality evaluation in medical imaging [2][11]. Even today, the quality of medical images is tested using subjective quality evaluation. Generally, subjective quality evaluations can be broadly categorized into two types [12]:

- Fidelity subjective evaluation. Quality of the image is defined in comparison to another, referenced image, and difference between the images is measured.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-179-3

- Domain specific subjective evaluation. Quality of an image is defined by the image usability in some domain specific task.

ROC evaluation is a representative of domain specific subjective evaluations and is one most often employed in medical imaging [3]. Qualified observers evaluate medical images as they would in a clinical task. For every medical image they have to provide a decision if an abnormality is present or not and to quantitatively describe their degree of certainty. This is usually a number from 1 to 5, where 1 means definitely negative confidence and 5 means definitely positive confidence [13]. The resulting diagnostic accuracy is compared with original image or to the gold standard which defines the truth. Therefore, observers can either correctly identify the anomaly (*true positive*), or miss it (*false negative*). For each of anomaly detected by the observer, either it agrees with the gold standard (*true positive*) or not (*false positive*). Also, the observer can correctly identify the absence of the anomaly (*true negative*). A subjective confidence rating of the diagnoses is then used as if it were a threshold to adjust for detection accuracy [2][4]. This threshold is used for plotting ROC curves which describe detection accuracy, Fig. 1. The plot is a summary of the trade off between true positives and false positives. The area under the curve can be used to summarize the overall quality or the efficiency of the detection process [13]. The size of the area under the curve directly corresponds to the image quality.

ROC analysis is used in medicine frequently, wherever there is a need to describe a binary decision (detection) problem. However, conventional ROC evaluation has its limits and drawbacks when used for medical image evaluation [13]:

- It can be used only to describe binary decisions. It cannot be used for evaluating an image with more than one anomaly.
- It is not location specific. It is possible for the observer to miss the anomaly in a correct spot, but to mistakenly identify it in another. This would be scored as a *true positive* when it should be scored as a *false positive*.
- Observers are forced, unnaturally, to indicate their confidence. When making decisions, clinicians usually use qualitative ways for describing their confidence rather than numerical rankings.

Several variations of ROC evaluation have been proposed to overcome the limits of the conventional ROC form. They address localization and binary decision issues because the confidence rating is not easy to overcome. The most popular variations of ROC medical image evaluation are [14][15]:

- Localization ROC (LROC). Observers are required to specify the single location, if any, at which an anomaly is judged to be present.
- Free-Response Operating Characteristic (FROC). In a sense, this is a generalization of LROC because the observer can specify more than a single anomaly and locations. This approach is limited because the results depend on the number of the locations allowed by the data analyst.
- Alternative FROC (AFROC). It is similar to FROC but enables a mandatory number of anomalies detected.
- Jack-knife analysis of FROC (JAFROC). A re-sampling method that does not assume independence of responses within the same study applied to FROC. Very stable method but it requires multiple readers and a substantial number of images for good results.
- Differential ROC (DROC). Determines the differences between modalities.

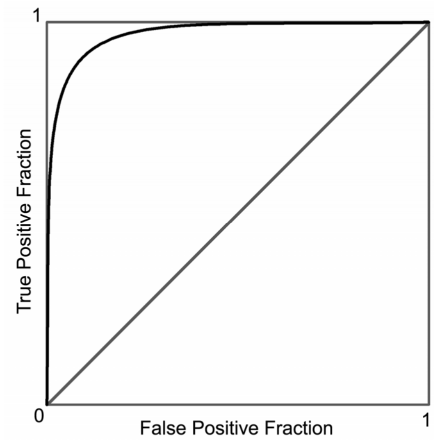

Figure 1. Example of ROC curve

A detailed description of ROC evaluations can be found in the aforementioned literature.

There is no standard for the *data gathering* phase of the ROC evaluation [7]. But it is possible to distinguish two approaches:

1) Based on *conventional software* used in everyday diagnostic tasks.
2) Based on *additional software* for performing observers' diagnostic task in the manner suitable for ROC evaluation.

The *Conventional software* approach enables the observers to work in familiar surroundings and in a familiar way, because they work with software tools used in everyday clinical practice. It is the cheapest way as it is not necessary to develop new software. However, this approach relies heavily on persons conducting the test. They follow observer performance and note the answers (detected anomalies and confidence rates). This approach is manageable for conventional ROC evaluation because it is only necessary to note the presence of an anomaly and the confidence rating. However, it is not suitable for ROC variations because the observers are forced to verbally describe the location of anomalies detected and there is a larger margin for error. The evaluator has to estimate if the answer is a *true* or *false positive*. This approach has been used in [16][17][18].

The *Additional software* approach is expensive and the clinical practice is only simulated. But it is easier to conduct ROC evaluation and gather data this way, especial in the case of the ROC variations. Observers mark the anomaly directly on the image evaluated, pinning its location. This data is immediately compared against the gold standard and memorized as a *true/false positive* or a *true/false negative* for latter analysis. This approach has been used in [19][20], but the software is only mentioned (it is not described) and it is not publicly available.

III. Design Considerations and Software Tools

As we described it in the Introduction, we had several goals in mind. The first goal, resemblance to software tools used in everyday clinical tasks, is very important for the success of ROC evaluation [13]. This means that GUI (Graphical User Interface) should resemble software tools used in everyday practice. The region marking software tool should be based on technologies supported in examination rooms because, the software is meant to be used in everyday clinical tasks.

But our decision to choose the underlying technologies that we did is additionally influenced by another fact. Clinicians are the primary persons undertaking the role of the observers in evaluation. Their schedule has few openings. It is hard for them to devote their time to ROC evaluation (at least it was like this in our case). It would be a good decision to allow the observers to evaluate medical images outside of examination rooms, in fact from any (even from their home if this is necessary and they are willing), at the time which is convenient to them. Although this does not correspond perfectly with clinical practice, it would enable far more clinicians to participate in the evaluation and it would help gather more statistical data. It is possible for several observers to evaluate images at the same time, thus greatly increasing the efficiency and, at the same time, reducing the cost of the ROC evaluation.

The region marking software tool should be accessible from many places by supporting remote access based on web technologies. This is the most common way for achieving access from several remote locations. The ultimate decision, whether the ROC evaluation will be conventional (constrained to examination room only), or it will support remote evaluation, is up to the persons conducting the test. It is only a matter of whether remote access is allowed or not which is set in the configuration files.

Our third goal, that it should be usable in education, dovetails well with our second goal. With little modification the software tool is suitable for remote education. The same principle used in ROC evaluation can be used in education. The difference lies in fact that it is not the image quality that is tested. Students will evaluate images remotely. They will mark anomalies they think are present in the image and they will describe their degree of certainty. Educational use differs from ROC evaluation in a way that the values of the gold standard will be shown at the end of the evaluation. Detailed explanations of clinical decisions will be presented for each anomaly and they will be compared to the answers of the students.

To achieve our last goal, a free software tool, we had to base our solution on technologies and software tools freely available. Fortunately for us, many of the web technologies belong to this category.

In line with previous explanations, we decided to use web technologies and we implemented the region marking software tool for medical images using HTML (Hyper Text Markup Language), PHP [21], and MySQL [22]. Web technology is supported almost everywhere and a web based solution can be used anywhere from an examination room to a handheld device. PHP is a general-purpose server-side scripting language used for Web development and for producing dynamic Web pages. In our solution PHP is used for generating parts of the GUI and for communicating with the MySQL database. MySQL is a relational database management system. We used MySQL for storing observers' answers, ROC evaluation data and image info. A good part of the GUI relies on JavaScript [23]. It is a weakly typed, prototype scripting language primarily executed in a Web browser. Multi-browser development can be an issue when pure JavaScript is used. As it is not possible to predict which Web browser will be used in evaluation, we decided to build the JavaScript part of the code through the JQuery library [24]. It is a cross-browser JavaScript library specifically designed to simplify client-side scripting of HTML and to enable, as much as possible, the same look for different Web browsers.

The application flow diagram of the region marking software tool for medical images is described in Fig. 2. There are two flow branches:

- ROC evaluation (right branch of Fig. 2).
- Educational use (left branch of Fig. 2).

Which branch in application flow diagram will be active depends on the user type. When the user logs into the application, hers/his type is determined (first step in Fig. 2) and appropriate branch is executed. User type is defined in user's profile stored in MySQL database.

Next step is the same in both cases. Users evaluate the medical image. They can insert more than one mark, modify previously defined marks, or completely delete a mark. Results of the observer's actions are memorized in MySQL database. The AJAX (Asynchronous JavaScript and XML) methods of JQuery library are used for data updating. This means that Web page is not reloaded every time observers

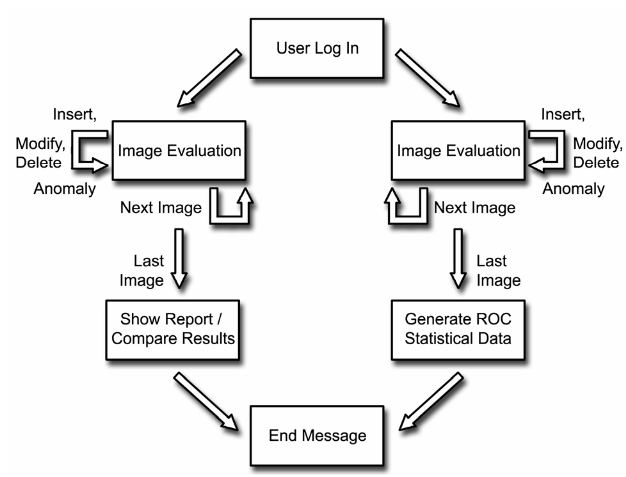

Figure 2. The application flow diagram of the region marking software tool for medical images

make some change. At any time, the observer can progress to the next medical image in the series. This effectively ends all actions on previous medical image meaning that it is not possible to modify observations made on the previous image.

It should be noted that observer can stop the evaluation at any given time. Next time, she/he logs to the application, medical image evaluation will continue from the last evaluated image.

When the observer reaches the end of the ROC evaluation, hers/his answers are converted to ROC statistical data, *positives* and *negatives*. The results are not reported to the observer. At the very end, user is presented with a message which designates the end of evaluation.

The end of evaluation in the educational version of the region marking software tool for medical images differs from the end of ROC evaluation. The observer's results are compared to the gold standard, a report is generated, and the results are presented to the observer. It is possible to view each of the images evaluated with observer's marks compared to the gold standard. At the end, observer is presented with the same message as in the ROC evaluation.

IV. The Region Marking Software Tool

At the moment of writing the ROC evaluation version of the software tool has been finished and prepared for clinical testing as it was needed for our quality evaluation system of medical images in PACS systems. This software version follows the right branch of application flow diagram described in Fig. 2.

The primary version of the software is written in the Serbian langue. However all language configurations are located in a PHP configuration file and are easily replaced with another language. We did this for the sake of presentation as all the images of the software tool contain English user interface labels.

The first step in the application is user identification through login page, Fig. 3. Only observers registered for ROC evaluation may log into the system.

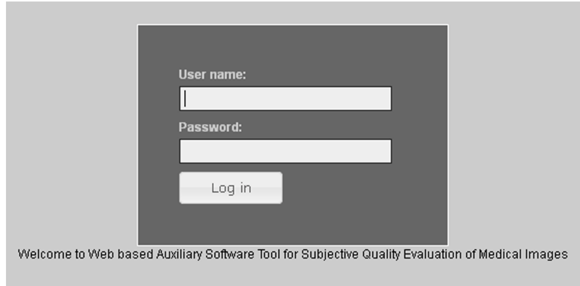

Figure 3. AST log in page

After login, observer starts/continues medical image quality evaluation by marking the places of anomalies on the image. Observer should left click on the approximate center of the anomaly. This action brings up a dialog for defining new anomaly, Fig. 4. Observer sets the size of anomaly and hers/his degree of certainty. Because of the technology limitations it is not possible to define custom sized anomaly in the current version of the application. Instead it is necessary to choose some of the predefined sizes: 50, 75, 100, 150, 200, and 300 pixels. This permits only basic overlap between the detections and the gold standard. We intend to improve on this in further software versions.

We decided to use the same scale for defining the degree of certainty as described in [13]. Instead of numbers, the observer chooses one of the answers which corresponds the best to hers/his degree of certainty. A different color is assigned to each of the answers. When the size of anomaly and degree of certainty are chosen, a circle is drawn around the place where observer left clicked, Fig. 5. The circle is drawn in the color corresponding to the chosen degree of certainty. Circles are drawn as an image overly using AJAX methods. Web page is not redrawn completely, but only part

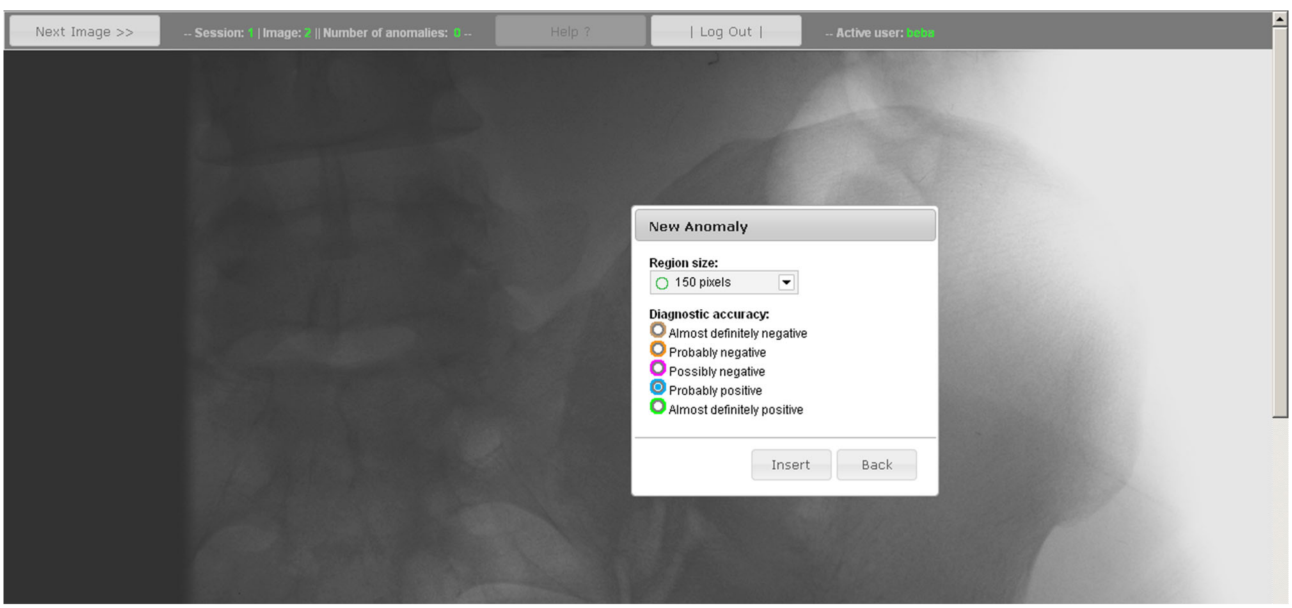

Figure 4. Dialog for defining new anomaly in the region marking software tool for medical images

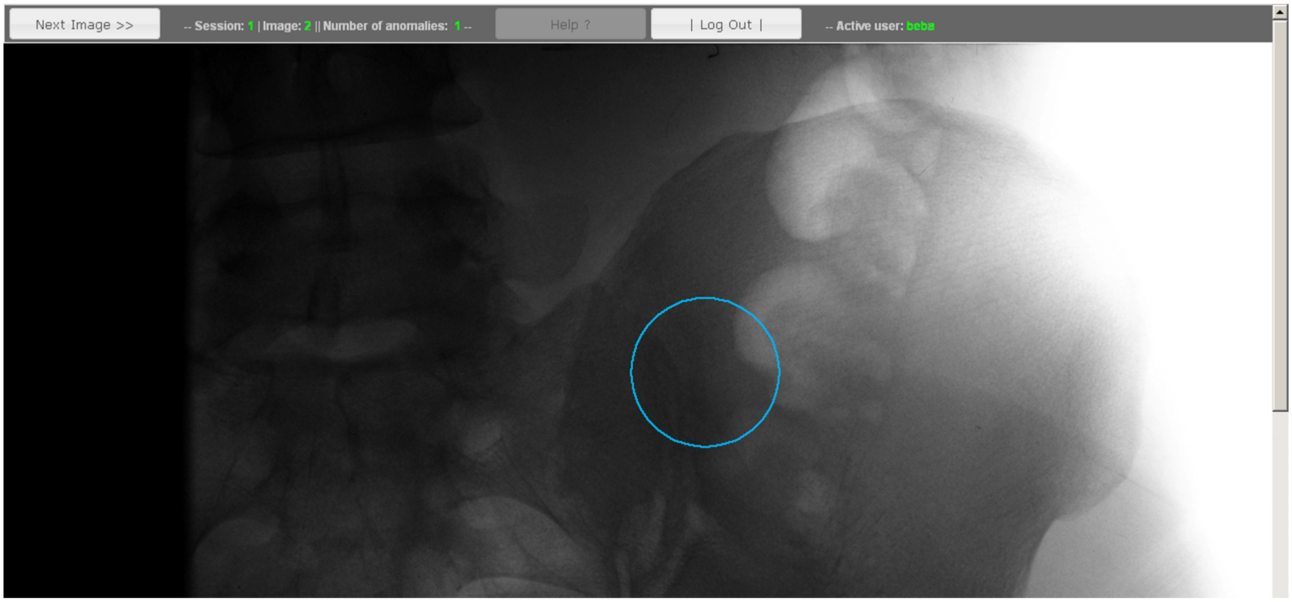

Figure 5. Example of Drawn circle around an anomaly in in the region marking software tool for medical images

of the image containing the circle. Also, AJAX methods are used in background to send updated data to MySQL database. These operations are fast and not noticeable by observers even on low end configurations.

It is possible to mark multiple anomalies on the image evaluated. But it is not possible to mark them on the same spot. If observer clicks on the already marked anomaly a dialog similar to the one described in Fig. 4 will display, Fig. 6. This dialog is used to modify the chosen anomaly or to completely remove it.

The observer may leave the evaluation by clicking the *Log Out* button. By clicking the *Next Image* button, observer will proceed to evaluate the next image. She/he cannot return to previous images evaluated. Absence of anomalies means that observer did not find any anomaly in the medical image evaluated. After the last medical image of the series is evaluated, the observer is presented with the ending message in which she/he is informed of the end of the evaluation and thanked for participating in the ROC evaluation.

After the last medical image is evaluated the background process is triggered. Observer's markings are converted into ROC statistical data used in ROC evaluation. Marked anomalies are compared to the gold standard. The center and size of the marked anomaly are compared to the data of the gold standard. If there is a match, a *true positive* is scored, if not a *false positive* is scored. Data gathered with the region marking software tool are statistically analyzed using some ROC analysis software tool.

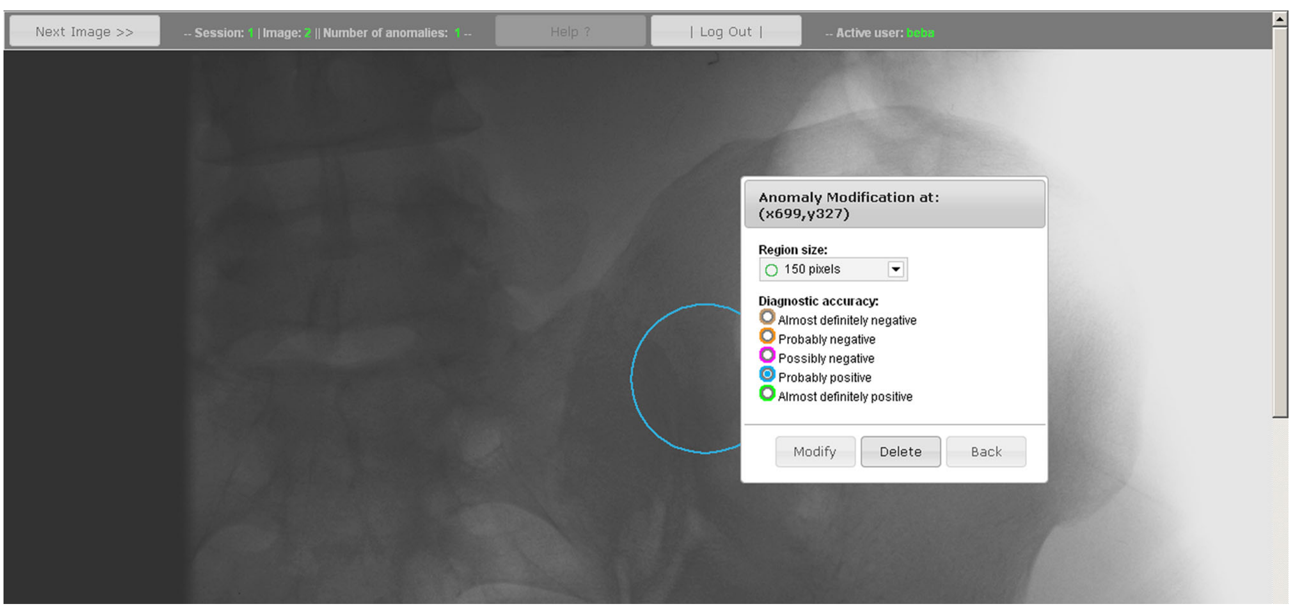

Figure 6. Dialog for modifying/deleting marked anomaly in the region marking software tool for medical images

V. Conclusion and Future Work

In the paper we described the region marking software tool for ROC evaluation of medical images. It is designed to be used with all versions of ROC evaluation. Therefore, an arbitrary number of anomalies and their location can be marked using this software tool. Also, the observers' degree of certainty is stored for every anomaly marked. It is designed for the on site evaluation (in accordance with everyday clinical tasks) and for the remote evaluation of medical images as well. The region marking software tool is designed to resemble similar software used in every day clinical practice, but it has one addition: remote accessibility. It is developed using web technologies such as HTML, PHP, JQuery for JavaScript, and MySQL. Internet access and a Web browser are the only requirements.

The region marking software tool for medical images is developed with an eye towards educational use. Training in diagnostic imaging does not differ much from ROC evaluation. Students need to mark the places they think anomalies exist and need to describe their degree of certainty. The only difference exists at the end. ROC observers are not presented with the results while students are. It is easy to extend the region marking software tool to support educational use as it is described in the paper. It is our intention to expand the software with this feature.

At the moment all the data gathered and processed are stored in a MySQL database. However, handling a MySQL database requires some computer skills. We intend to expend data management so that XML (Extensible Markup Language) files are used as an alternative to a MySQL database. Textual form of XML is easier to handle for inexperienced users. XML will contain description of the software interface. In this way it is possible to change the description of diagnostic accuracy and anomaly size.

As it is based on free technologies, the region marking software tool for medical images is practically free. It is our intention to make it freely available for the general public, after extensive clinical trials.

Acknowledgment

This research is financially supported by Ministry of Science and Technological Development, Republic of Serbia; under the project number TR32044 "The development of software tools for business process analysis and improvement", 2011-2014.

References

[1] S. Aja-Fernández, R. S. Estépar, C. Alberola-López, and C. F. Westin, "Image quality assessment based on local variance," Conference Proceedings of the IEEE Engineering in Medicine and Biology Socciety 2006, Vol.1, pp. 4815-4818, 2006.

[2] D. Smutek, "Quality measurement of lossy compression in medical imaging," Prague Medical Reports, Vol. 106, No. 1, pp. 5-26, 2005.

[3] B.J. Erickson, "Irreversible compression of medical images," Journal of Digital Imaging, Vol. 15, No. 1, pp. 5-14, 2002.

[4] D. Dragan and D. Ivetic, "Quality Evaluation of Medical Image Compression: What to Measure?," Proceeding of the 2010 IEEE 8th International Symposium on Intelligent Systems and Informatics, pp. 37-42, 2010.

[5] B.J. Erickson, "Image Compression," PACS: A Guide to the Digital Revolution, K.J. Dreyer, D. S. Hirschorn, J. H. Thrall, and A. Mehta, (Eds.), Springer-Verlag New York Inc., pp. 229-247, 2006.

[6] R. K. W. Schulze, et al, "|Diagnostic yield of ink-jet prints from digital radiographs for the assessment of approximal carious lesions: ROC-analysis," European Journal of Radiology, Vol. 79, No. 2, pp. 277-282, 2011.

[7] D. P. Chakraborty, "Recent advances in observer performance methodology: jackknife free-response ROC (JAFROC)," Radiation Protection Dosimetry, Vol. 114, No. 1-3, pp. 26-31, 2005.

[8] D. Dragan and D. Ivetic, "A Comprehensive Quality Evaluation System for PACS," Ubiquitous Computing and Communication Journal, Special Issue on ICIT 2009 Conference - Bioinformatics and Image, Vol. 4, No. 3, pp. 642-650, 2009.

[9] D. Ivetic and D. Dragan, "Medical Image on the Go!," Journal of Medical Systems, Vol. 35, No. 4, pp. 499-516, 2011.

[10] D. Dragan and D. Ivetic, "Request redirection paradigm in medical image archive implementation," Computer Methods and Programs in Biomedicine, In Press, doi: 10.1016/j.cmpb.2011.06.001, 2011.

[11] S. Winkler, "On the properties of subjective ratings in video quality experiments," Proceedings of the International Workshop on Quality of Multimedia Experience (QoMEX), doi: 10.1109/QOMEX.2009.5246961, 2009.

[12] A. Pommert and K. H. Hohne, "Evaluation of Image Quality in Medical Volume Visualization: The State of the Art," Proceeding of Medical Image Computing and Computer-Assisted Intervention (MICCAI 2002), Part II, T. Dohi and R. Kikinis (Eds.), Lecture Notes in Computer Science, Vol. 2489, pp.598-605, 2002.

[13] P. Cosman, R. Gray, and R. Olshen, "Chapter 49: Quality Evaluation for Compressed Medical Images: Fundamentals," Handbook of Medical Imaging, Processing and Analysis, Isaac N. Bankman (Ed.). Academic Press Inc., pp.803-819, 2000.

[14] C. Metz, "Receiver Operating Characteristic Analysis: A Tool for the Quantitative Evaluation of Observer Performance and Imaging Systems," Journal of the American College of Radiology, Vol. 3, No. 6, pp. 413-422, 2006.

[15] F. Zarb, L. Rainford, and M. F. McEntee, "Image quality assessment tools for optimization of CT images," Radiography, Vol. 16, No. 2, pp. 147-153, 2010.

[16] F. Li, et al, "Computer-Aided Detection of Peripheral Lung Cancers Missed at CT: ROC Analysis without and with Localization," Radiology, Vol. 237, No. 2, pp. 684-690, 2005.

[17] L. Zhigang, L.I. Kuncheng, Z. Jinghong, and L Shuliang, "The study of diagnostic accuracy of chest nodules by using different compression methods," European Journal of Radiology, Vol. 55, No. 2, pp. 255-257, 2005.

[18] D.H. Kim, et al, "Comparison and Evaluation of JPEG and JPEG2000 in Medical Images for CR (Computed Radiography)," Journal of the Korean Physical Society, Vol. 56, No. 3, pp. 856-862, 2010.

[19] M. Kallergi, et al, "Improved interpretation of digitized mammography with wavelet processing: a localization response operating characteristic study," American Journal of Roentgenology, Vol. 182, No. 3, pp.697-703, 2004.

[20] M. Kallergi, et al, "High-Performance Wavelet Compression for Mammography: Localization Response Operating Characteristic Evaluation," Radiology, Vol. 238, No. 1, pp. 62-73, 2006.

[21] PHP, general-purpose scripting language. [Online] Available at: http://www.php.net/, 11/26/2011.

[22] MySQL, open source database. [Online] Available at: http://www.mysql.com/, 11/26/2011.

[23] S. Suehring, "JavaScript(TM) Step by Step," Microsoft Press, p.432, 2008.

[24] JQueery, cross-browser JavaScript library. [Online] Available at: http://jquery.com/, 11/26/2011.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-179-3