

Improved Mammogram Interpretation with an Ontology-driven Editor and MammoViewer – Preliminary Results

**TERESA PODSIADŁY-MARCZYKOWSKA^{1,*}, ANNA WRÓBLEWSKA²,
ARTUR PRZELASKOWSKI²**

¹ *Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Science, Warsaw, Poland*

² *Institute of Radioelectronics, Warsaw, Poland*

Although mammography is a standard of reference for detection of early breast cancer, as many as 25% of breast cancers may be missed. To reduce the possibility of missing a cancer, the following methods and tools have been proposed: continuing education and training, prospective double reading, retrospective evaluation of missed cases, and use of computer-aided detection (CAD). The purpose of the reported work was to evaluate the usefulness and the potential of our aiding tools: an ontology driven editor for mammographic lesion description (MammoEdit) and a CAD-tool (MammoViewer) to enhance radiologist's diagnostic performance. To this end test sample of mammograms was analyzed twice, without and with aiding tools. The obtained data were analyzed using (ROC) analysis and Kappa statistics. Statistical analysis of the test data demonstrated potential of both tools to enhance radiologist's diagnostic performance.

K e y w o r d s: radiological interpretation, ontology, diagnostic accuracy, computer-aided diagnosis

1. Introduction

Mammography is currently the most effective tool for early detection of breast cancer. It is still the only technique that can detect breast cancer in a pre-invasive stage. Early detection via mammography increases breast cancer treatment options and the survival rate [1]. However, the diagnostic value of mammography is limited by significant and high, up to 25% ratio of missed breast cancers [2, 3].

* Correspondence to: Teresa Podsiadły-Marczykowska, Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, ul. Ks. Trojdena 4, 02-109 Warsaw, Poland, e-mail: tpodsiadly@ibib.waw.pl

Received 19 September 2007; Accepted 07 May 2008

Main, human-oriented causes lowering this ratio are radiologists' perception and interpretation errors. In [4] authors state that in a practice where radiologists are well trained, experienced, and where a medical audit indicates satisfactory performance, the frequency of missed malignant lesions indicates that such cases only rarely are effect of negligence. Therefore, strategies, means, and tools to reduce radiologists' errors are essential to improve diagnostic effectiveness in mammography.

Methods described in medical literature and commonly used to decrease the number of missed cancers include: continuing education and training, prospective double reading, retrospective evaluation of missed cases, and use of computer-aided detection (CAD) [4]. Computer-aided detection (CAD) and ontology have been discussed among others in [5, 6] as aiding tools for the interpretation of the mammograms.

The purpose of the study is the verification of the impact of aiding tools on radiologist's diagnostic performance. To this end the set of test mammograms was interpreted two times, without and with support of aiding tools, an ontology-based editor for mammograms description and MammoViewer, a CAD tool. Radiologist diagnostic performance was evaluated using ROC analysis and Kappa statistics. For MammoEdit observer performance of the initial interpretation without aid and second supported interpretation yielded the following areas under the curve: $AUC^I = 0.504$ and $AUC^{II} = 0.663$. The difference between areas under (ROC) curves was statistically significant ($p = 0.023$). Additionally, use of MammoEdit resulted in increased interobserver agreement from poor ($kappa^I = 0.136$) to fair ($kappa^{II} = 0.305$). Use of MammoViewer resulted in increase in diagnostic performance from $AUC^I = 0.617$ for originals mammograms to $AUC^{II} = 0.789$ for processed ones.

The paper is organized as follows: second section presents results of broad range review of medical literature search to determine what are the causes of the most commonly missed cancers, their types and features. The aim of the analysis was the selection of test sample composed of cases that would be probably false-negative when interpreting the mammograms. This section shortly describes our tools proposed to reduce the false-negative rate. Third part of the paper describes organization of tests, test samples and statistical analysis. The fourth presents the qualitative results of radiologist's of interpretation with the use of the editor, section four presents results of tests, section five concludes the paper and discusses future works.

2. Radiologist's Errors, Tools to Reduce Them

Our intention was to test aiding tools aimed at pathologies that are most likely missed. To identify such lesion we reviewed certain medical literature to determine why and what types of cancers are missed at mammographic screening and analyzed types and causes of radiologist's errors. According to [7], main causes of missed cancers can be related to radiologist's errors and poor technical conditions (the last goes beyond

the scope of this paper). Radiologist's errors can be classified as interpretation and perception errors [7, 8].

The perception error occurs when the lesion is included in the field of view and visible but is not recognized by the radiologist. The lesion may or may not have subtle features of malignancy. Possible causes for the perception errors include dense parenchyma obscuring a lesion, and "happy eye syndrome" (cessation of looking for additional lesions when one abnormality is seen).

The second cause of the missed breast cancers related to radiologist's error is incorrect lesion interpretation, which occurs when an abnormality with suspect features is observed and reported, but is misinterpreted as being definitely or at least probably benign.

According to [4, 7, 8, 9] the types of the most frequently misinterpreted or overlooked lesions are:

1. Cluster of microcalcifications (small or hardly visible in dense breasts – perception and interpretation errors)
2. Lesions with benign appearance – masses and clusters of microcalcifications (interpretation error)
3. Small masses (perception error)
4. Asymmetric density (perception and interpretation errors)

Cases for the test were selected according lesion types described in above list.

2.1. Ontology-driven Editor – MammoEdit

Our assumption was that using a reliable domain knowledge representation to design and control mammographic data entry has a potential to improve semantics (i.e. meaning and significance) and completeness of these data. To this end mammographic ontology [10] has been used as a partial set of design assumptions in developing a graphical editor for description of the mammograms. Radiologists are familiar with images, so the choice of graphical, iconic data presentation model for the mammography report editor has been obvious. Concepts necessary for mammographic lesion description have been represented by graphical primitives – icons. Data about lesions are captured and displayed using this iconic representation. Complete lesion definitions based on an ontological model of the domain and graphical data representation result in improved data completeness and in turn in better interpretation of the lesion features and diagnosis. The editor has been described in more details in [6].

2.2. MammoViewer – a Cad Tool

MammoViewer is a computer-aided diagnosis application [11, 12]. It utilizes effective methods of presentation, processing, analysis and interpretation of images. It can be used as an advanced viewer dedicated to medical images – particularly mammograms

– with widespread imaging options like measuring structures, operating in regions of interests, defining parameters of a sliding window, setting the range of the shown pixel values etc. On the other hand, as a scientific software package – MammoViewer includes a variety of methods to process medical images; processing in multiresolution wavelet domain appeared particularly useful in mammography.

Mammogram preprocessing refers to enhanced perception of pathologies in order to provide more conspicuous pathology signs for radiologists, who interpret mammogram exams. This was achieved by:

- denoising – removing or weakening information insignificant for diagnosis,
- local contrast enhancement – amplifying contours of the lesions (that allows radiologist better determining shapes of abnormalities and evaluating abnormality margins in more details) and emphasizing other diagnostically important lesion features like texture inside pathologies.

3. Aiding Tools' Tests – Material, Organization, and Statistical Analysis

The aim of the tests was to compare the effectiveness of mammogram interpretation made in routine way without aiding tools with the effectiveness of such interpretation made with our aiding tools. Tests were performed on mammograms from DDSM [13]. Each test case includes two projections of each breast with associated patient information among others ACR breast density and lesion subtlety rating (from 1 for lesion very difficult to perceive to 5 for obvious ones), also image information (scanner, spatial resolution). Test mammograms were digitized at a pixel size of 43.5 microns and 50 microns with a 12-bit grayscale, lesions were shown in two projections.

3.1. Test Sample for MammoEdit Verification

Test sample for editor test consisted of 72 mammograms, 20 negatives and 52 pathology cases. As it was stated above, to test the editor there were chosen the types of lesions that may have an impact on interpretation errors: 26 clusters of microcalcifications (according to [8] radiologists evaluate clusters worse than masses), and 26 (some with benign appearance [4]). Characteristics of the test sample in terms of lesions subtlety are presented in Fig. 1.

3.2. Test sample for MammoViewer Verification

For the perception improvement test the following types of the cases were chosen: clusters of small microcalcifications in dense breast tissue and small masses. The test sample for MammoViewer test comprised of 51 mammograms, 18 negatives, and

33 pathology cases – 9 masses and 24 microcalcification clusters. Characteristics of the test sample in terms of lesions subtlety are presented in Fig. 2.

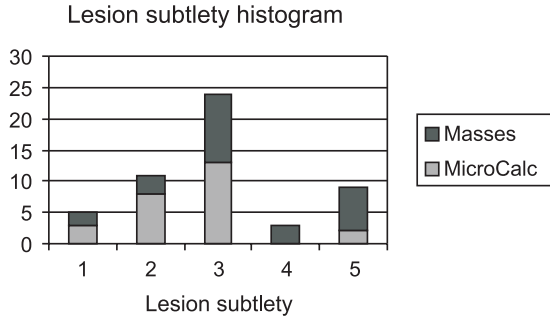


Fig. 1. Lesion subtlety histogram for the MammoEdit test sample. Average subtlety for all lesions – 3, for microcalcification clusters – 2.6 and 3.38 for masses

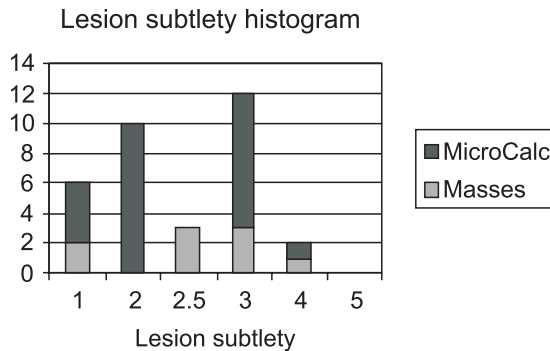


Fig. 2. Lesion subtlety histogram for the MammoViewer test sample. Average subtlety for all lesions – 2.35, for microcalcification clusters – 2.29 and 2.5 for masses

3.3. Tests Organization

To assess radiologist's diagnostic performance without and with aiding tools a set of test mammograms was interpreted by the radiologist two times, without and with support of aiding tools, at a time interval of approximately 3 months.

During the first reading of the mammograms – the first test stage – the cases were interpreted using MammoViewer as a medical image viewer (without scientific processing methods) the reports were dictated, lesion diagnosis was assigned using BI-RADS scale [14]. This first procedure was the same for both Mammoedit and MammoViewer tests.

During the MammoEdit second test stage both the tools MammoViewer and the editor were used for lesion description and assessment of the cases, lesion final diagnosis was assigned using BI-RADS scale [14].

During the MammoEdit second test stage both the tools MammoViewer and the editor were used for lesion description and assessment of the cases.

During the MammoViewer second test part the radiologist compared original mammograms with the corresponding processed images giving his opinion on quality of lesion diagnostic features using subjective, comparative measures of diagnostic symptoms quality (see Table 1). Moreover, interpreting the cases the radiologist read additionally processed images. After that, the radiologist verified CAD marked potential microcalcification clusters (CMMC) and extend final diagnosis in case of any additionally observed clusters.

Table 1. Subjective, comparative mark scale that was used in the perception improvement tests. All important image features were scored consequently and conditions of true diagnosis were assessed

Mark scale	Description of diagnostic image quality
+3	definitely better
+2	better
+1	slightly better
0	comparable with the original
-1	slightly worse
-2	worse
-3	definitely worse

3.4. Statistical Analysis

The obtained data were assessed using receiver operating characteristic (ROC) curves analysis [15]. Observer performance was measured by calculating the areas under the ROC curves (AUC). The observer performance at the initial interpretations and while using one of the aiding tools were evaluated by comparing the resulting areas under the ROC curves. The statistical significance was calculated using Dorman-Berbaum-Metz algorithm [16]. Statistical software used was JROCFIT [17] and MedCalc [18].

The final diagnosis assessment without and with MammoEdit compared to it's corresponding diagnosis in DDSM treated as "gold standard" was used to evaluate whether the use of MammoEdit can reduce variability in the interpretation of mammographic lesions. Agreement between the observers was determined using weighted kappa statistics [18].

4. Test Results

4.1. Improvement of Radiologist's Pefromance Using Mammoedit

For all test samples, observer performance of the initial interpretation without aid and the supported interpretation yielded the following areas under the curve:

$AUC^I = 0.504$ for the first test stage and $AUC^{II} = 0.663$ for the second stage with MammoEdit. Graphs showing ROC curves for test stages without and with MammoEdit are presented in Fig. 3. The difference between areas under the curves was statistically significant ($p = 0.023$). The detailed values of ROC analysis parameters showing increase in diagnostic performance of the radiologist using MammoEdit are presented in Tables 2 and 3.

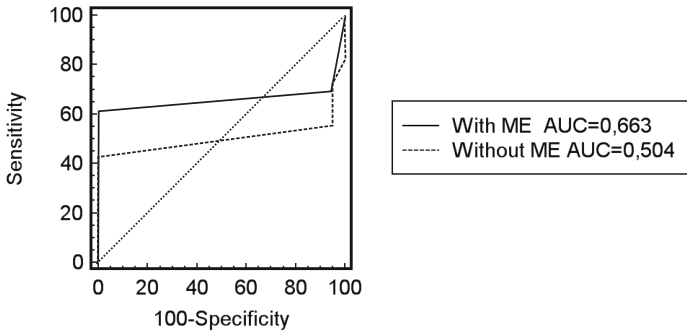


Fig. 3. Use of MammoEdit increase radiologist diagnostic performance in mammograms interpretation. ROC curves comparing radiologist’s diagnostic performance without and with MammoEdit aiding tool in interpreting a test set of 72 mammograms where sensitivity – $SE = TP/(TP+FN)$, specificity $SP = FP/(TP+FN)$, TP – true positive cases, FP – false positive cases, FN – false negative cases

Table 2. Parameters for ROC analysis curves showing increase in radiologists’ diagnostic performance with the aiding tool – MammoEdit

	AUC	Standard. Error	95% CI
Radiologist without ME	0.504	0.0766	0.384 – 0.624
Radiologist with ME	0.663	0.0678	0.542 – 0.771

Table 3. Difference between areas under (ROC) curves corresponding to radiologist’s diagnostic performance in two test stages was statistically significant $p = 0.023$

Pair wise comparison of ROC curves	
Difference between areas	0.159
Standard error	0.0701
95% CI	0.0218-0.296
Significance level	0.023

4.2. Interobserver Agreement

Use of MammoEdit for lesion interpretation resulted also in increase of diagnosis consistency. Strength of agreement in the final diagnosis assessment measured with Kappa statistics was evaluated for two pairs of results. Radiologist interpreting

mammograms without MammoEdit and DDSM lesion evaluation form the pair number 1, radiologist interpreting mammograms with MammoEdit and DDSM opinion on lesion diagnosis pair number 2. Differences in agreement between those two pairs of observers are presented in Table 4. Interpretation of the kappa values was based on the method described in [19].

Table 4. Comparison of agreement between radiologist interpreting mammograms without and with MammoEdit. Use of MammoEdit results in increase of diagnosis consistency

Pairs of observers	Weighted Kappa value	Interpretation
Pair no. 1	0.136	Poor agreement
Pair no. 2	0.305	Fair agreement

4.3. Improvement of Radiologist's Performance and Perception Using Mammoviewer

The results of the original and processed images quality comparison is shown in Table 5.

Table 5. Mean mark of the radiologist taking part in the tests

Types of lesions	All types	Negatives	Masses	MicroCalc
Mean radiologist's mark	+1.71	+1.39	+1.67	+1.96

For all cases, observer performance of the initial interpretation without aid and the supported interpretation yielded the areas under the ROC curves (AUC), accuracy (ACC), sensitivity (SE), and specificity (SPEC) shown in Table 6. The difference in diagnosis using only the original images and that with the processed images is not statistically significant (p -value is 0.1787). However, the difference between the diagnosis using the original images and the processed images with CMMC is statistically significant (p -value is 0.0226).

Table 6. Observer performance of the initial interpretation without aid, and the supported interpretation with the processed images and then CMMC

The test stage	Original images	Original images + the processed ones	Original images + the processed ones + CMMC
AUC (area under the ROC curve)	0.617	0.743	0.789
ACC (accuracy)	52.9%	58.8%	64.7%
SE (sensitivity)	60.6%	75.8%	84.8%
SPEC (specificity)	38.9%	27.8%	27.8%

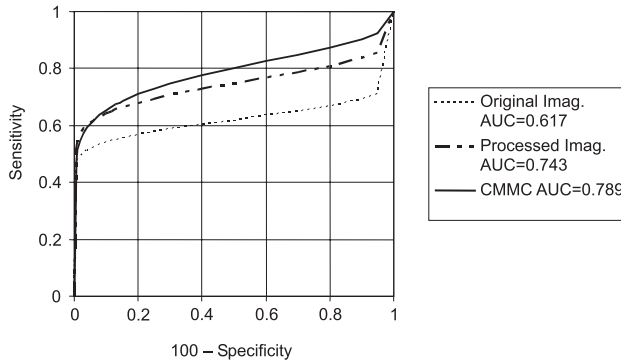


Fig. 4. ROC curves comparing the radiologist's diagnostic performance without and with the CAD tool MammoViewer (with an aid of the processed images and then CAD marked potential microcalcification clusters) in interpreting the test set of 51 mammograms

5. Conclusions

Initial verification of both the aiding tools demonstrated their potential to enhance radiologist's diagnostic performance. The use of MammoEdit to assist radiologist in interpretation of lesion features resulted in increment of area under ROC curve from $AUC^I = 0.504$ for routine interpretation to $AUC^{II} = 0.663$ for the interpretation with the interpretation supporting tool – MammoEdit. Additionally use of MammoEdit enhanced diagnosis consistency. Radiologist's agreement with the “gold standard” diagnosis – DDSM increased from slight ($\kappa = 0.136$) in first test stage to moderate ($\kappa = 0.305$) with the aiding tool.

Computerized image processing method makes the signs of pathologies slightly more conspicuous and results in improvement of lesion perception and finally in increment of radiologist's diagnostic performance. The problem was that using processed images the radiologist saw more precise structure that sometimes led to over diagnosing (sensitivity increased, but specificity decreased). Use of MammoViewer resulted in increase in diagnostic performance measured with AUC from $AUC^I = 0.617$ for originals mammograms to $AUC^{II} = 0.789$ for processed ones with CMMC.

To assess the tools in multi-user and multi-cases setting we plan further tests with a greater number of cases and cooperation of 3 radiologists at different levels of expertise.

References

1. Lee C.H.: Screening mammography: proven benefit, continued controversy. Radiologic Clinics of North America, 2002, 40, 395–407.

2. Sickles E., Wolverton D., Dee K.: Performance parameters for screening and diagnostic mammography: specialist and general radiologists. *Radiology*, 2002, 224, 861–869.
3. Elmore J., et al.: Screening mammograms by community radiologists: variability in false-positive rates. *Journal of the National Cancer Institute*, 2002, 94, 1373–1380.
4. Bird R.E., Wallace T.W., Yankaskas B.C.: Analysis of cancers missed at screening mammography. *Radiology*, 1992, 184, 613–617.
5. Przelaskowski A., Podsiadły-Marczykowska T., Wróblewska A., Boniński P., Bargieł P.: Computer aided interpretation of medical images: limitations and challenges. *Machine Graphics* 2007, 16, 3, 2007.
6. Podsiadły-Marczykowska T., Przelaskowski A., Wróblewska A., Boniński P.: A framework for integrated, diagnosis supporting interface for mammograms description – advantages and pitfalls, in *Proc. of the second Int. Conference on Human-Computer Interaction*, March 14–16, 2007, 80–86.
7. Majid A.S., De Parades E.S., Doherty R.D., Sharma N.R., Salvador X.: Missed breast carcinoma: pitfalls and pearls. *RadioGraphics*, 2003, 23, 881–895.
8. Markey M. K., Lo J.Y., Floyd C.E.: Masses vs calcifications: differences between computer-aided diagnosis of breast masses and that of calcifications. *Radiology*, 2002, 223, 489–493.
9. Birdwell R.L., Ikeda D.M., O’Shaughnessy K.F., Sickles E.A.: Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection. *Radiology*, 2001, 219, 192–202.
10. Podsiadły-Marczykowska T., Guzik A.: Mammographic ontology, model structure, definitions and concept instances. *Bio-Algorithms and Medical Systems*, 2005, 1, 1, 247–252.
11. Wróblewska A., Przelaskowski A., Bargieł P., Boniński P.: MammoViewer – CAD application based on effective multiscale image analysis. *Polish Journal of Medical Physics & Engineering*, 2005, 11, 1, 17–28.
12. <http://www.ire.pw.edu.pl/MammoViewer>
13. DDSM <http://marathon.csee.usf.edu/Mammography/Database.html>
14. American College of Radiology (ACR). Breast imaging reporting and data system (BI-RADS). Reston, 3 edition, 1998.
15. Metz CE.: Basic principles of ROC analysis. *Semin Nucl Med* 1978, 8, 283–298.
16. ROCKIT and DBM MRMC software. http://www.radiology.uchicago.edu/krl/KRL_ROC/software/index6.html.
17. JROCFIT <http://www.rad.jhmi.edu/jeng/javarad/roc/JROCFIT.html>
18. MedCalc software <http://www.medcalc.be>
19. Landis J.R., Koch G.G.: The measurement of observer agreement for categorical data. *Biometrics* 1977, 33, 159–174.