Computers in Biology and Medicine 96 (2018) 294–298

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Breast lesion shape and margin evaluation: BI-RADS based metrics understate radiologists' actual levels of agreement

Mohammad Rawashdeh [a,*], Sarah Lewis [b], Maha Zaitoun [a], Patrick Brennan [b]

[a] Faculty of Applied Medical Sciences, Jordan University of Science and Technology, Irbid 22110, Jordan
[b] Medical Image Optimisation and Perception Group (MIOPeG), Faculty of Health Sciences, The University of Sydney, Sydney, New South Wales, Australia

ARTICLE INFO

Keywords:
Lesion delineation
Radiotherapy
Breast lesion
Mammography
Electronic stylus

ABSTRACT

Background: While there is much literature describing the radiologic detection of breast cancer, there are limited data available on the agreement between experts when delineating and classifying breast lesions. The aim of this work is to measure the level of agreement between expert radiologists when delineating and classifying breast lesions as demonstrated through Breast Imaging Reporting and Data System (BI-RADS) and quantitative shape metrics.
Methods: Forty mammographic images, each containing a single lesion, were presented to nine expert breast radiologists using a high specification interactive digital drawing tablet with stylus. Each reader was asked to manually delineate the breast masses using the tablet and stylus and then visually classify the lesion according to the American College of Radiology (ACR) BI-RADS lexicon. The delineated lesion compactness and elongation were computed using Matlab software. Intraclass Correlation Coefficient (ICC) and Cohen's kappa were used to assess inter-observer agreement for delineation and classification outcomes, respectively.
Results: Inter-observer agreement was fair for BI-RADS shape (kappa = 0.37) and moderate for margin (kappa = 0.58) assessments. Agreement for quantitative shape metrics was good for lesion elongation (ICC = 0.82) and excellent for compactness (ICC = 0.93).
Conclusions: Fair to moderate levels of agreement was shown by radiologists for shape and margin classifications of cancers using the BI-RADS lexicon. When quantitative shape metrics were used to evaluate radiologists' delineation of lesions, good to excellent inter-observer agreement was found. The results suggest that qualitative descriptors such as BI-RADS lesion shape and margin understate the actual level of expert radiologist agreement.

1. Introduction

There is much literature describing the radiologic detection of breast cancer [1–6]; however, limited evidence is available on the agreement between radiology experts when delineating (segmenting) and classifying lesions. This information is of both academic and clinical relevance for improving the technical success of biopsies and contouring target volumes for radiotherapy [7]. Also, breast lesion descriptions, such as their shape and margin, are indicators of the local aggressiveness of invasive breast cancer [8–10]. The ability of radiologists to delineate cancers or provide realistic ground truth using currently available interactive tools is not yet fully understood [1–13].

The classification of breast masses based on perimeter details and parenchymal distortion is known to impact upon initial diagnosis and subsequent patient management. Such classification is not an easy task, and the error rates presented in the literature [13] highlight the number of potentially avoidable breast biopsies and the associated patient mental and physical anxiety and health service costs when an incorrect diagnosis and delineation are made. The American College of Radiology developed the Breast Imaging Reporting and Data System (BI-RADS) lexicon [14] to optimise the description of breast cancers according to breast density, lesion features and lesion shape, with the overall aim of improving accuracy in mammographic reporting. For this to be effective, consistency between and within observers must be achieved; however, previous research designed to validate the BI-RADS lexicon has demonstrated wide inter- and intraobserver variation [15]. It remains unclear whether this lack of agreement truly reflects radiologists' decisions, or rather an inability of the BI-RADS lexicon to accurately represent radiologists' opinions.

The aim of the current work was to establish the level of agreement

* Corresponding author. Dept. of Allied Medical Sciences, Faculty of Applied Medical Sciences, Jordan University of Science and Technology, Jordan.
E-mail address: marawashdeh@just.edu.jo (M. Rawashdeh).

https://doi.org/10.1016/j.compbiomed.2018.04.005
Received 27 February 2018; Received in revised form 8 April 2018; Accepted 9 April 2018

among expert radiologists in describing the shape and margin of breast lesions when using BI-RADS descriptors compared with quantitative shape metrics arising from their lesion delineations.

2. Materials and methods

2.1. Ethical approval

This study received ethics review board approval, and informed consent was obtained from each reader. Ethical approval was also provided by Breast Screen New South Wales for the use of de-identified patient images with patient consent waived.

2.2. Participants

Nine expert breast radiologists participated in the study. Information on the radiologists' professional experience and familiarity with mammography was collected using a questionnaire.

2.3. Selection and description of images

A test set of 40 images, consisting of 23 craniocaudal (CC) and 17 mediolateral oblique (MLO) digitally acquired projections of the breast, was collated by a senior breast radiologist with over 20 years experience. All images contained a single biopsy-proven malignancy (invasive cancer or ductal carcinoma in situ) which reflected the most common lesion types found in the general population, with the exception of calcifications which were excluded as the focus of this study was mass delineation. Examples of discrete masses ($n = 20$), asymmetric density ($n = 12$) and architectural distortions ($n = 8$) were included.

To produce a descriptive summary of the images, the nine radiologists were asked to classify each image according to breast density, lesion shape and margin. A summary of lesion and image characteristics, as described by the most common classification feature assigned by the nine radiologists, is shown in Table 1. For example, if five radiologists assessed a case and assigned it a score of BI-RADS 4 while the other four radiologists gave the same case a score of BI-RADS 3, we considered the case BI-RADS 4. A dot at the centre of each lesion was used to direct the participating radiologists toward the lesions in order to eliminate search errors.

2.4. Image evaluation

Each radiologist performed two tasks for each image. First, each radiologist outlined the external margin of each lesion, including the central mass and all spicules. No time limit was imposed, and the reader was able to go back and adjust the lesion outline until they were completely satisfied. Following this delineation task, each radiologist was asked to classify each lesion according to the BI-RADS fifth edition lexicon, firstly for shape [round (1), oval (2), lobular (3) and irregular (4)] and secondly, for the margin of the mass [circumscribed (1), indistinct (2), spiculated (3), microlobulated (4) and obscured (5)] [16]. The numbers shown in parentheses above represent the BI-RADS scores for the shape and margin classifications. Each reader completed the delineation and classification tasks for all 40 cases within 1 h. A practice set of three images was assessed by each reader before the study. The data generated from these were not included in the analysis.

Table 1
A summary of image and lesion features in terms of shape and margin.

Feature	Number of images (%)*
Lesion shape	
Round	5 (13)
Oval	6 (15)
Lobular	8 (20)
Irregular	21 (52)
Lesion margin	
Circumscribed	8 (20)
Indistinct	15 (38)
Spiculated	17 (42)

From the outlines, Matlab software (version 7.13, by MathWorks) was used to calculate the elongation, which is the ratio of the major axis length to minor axis length, and the compactness, a dimensionless quantity which is inversely related to the similarity of an object to a circle [17], defined by the following equation:

$$\text{compactness} = 1\text{-}\ (4\ \pi\ \text{area}/\ (\text{perimeter})^2)$$

A circle is the most compact shape with a compactness value of 0, and this value increases with the complexity of the contour to a maximum value of 1. Elongation and compactness descriptors can quantitatively represent the roughness and complexity of the shape, which helps in radiologic and automated classification of malignant and benign masses. A malignant tumour with a number of concavities or spicules has a higher compactness value than a smooth and round benign mass [17].

2.5. Image display

Lesion demarcation was performed in a reading room with an area of 60 m^2 and ambient lighting of 12–20 lux at the position of the reader, as measured using a calibrated photometer (model CL-200; Konica Minolta, Ramsey, NJ, USA). The walls of the room were painted with a matte finish in order to provide minimum specular reflection. The radiologists outlined each lesion using an interactive digital drawing tablet (56.1×42.1 cm; Cintiq 21UX; WACOM, Vancouver, WA, USA) connected to a standard laptop running ImageJ software (Fig. 1). Previous studies have shown that the use of a stylus can facilitate more precise lesion outlining compared to mouse clicking [18,19]. The orientation and angle of the tablet was adjustable so that it could be arranged in the most ergonomically comfortable position for each radiologist [18,19]. Each image was magnified 4× for the outlining process so that the lesion occupied at least half of the display area. The location of the lesion in each image was highlighted with a mark at the centre of the lesion. The tablet display had a resolution of 1600×1200 pixels and a contrast ratio of 550:1, and was calibrated to the Digital Imaging and Communication in Medicine Greyscale Standard Display Function (DICOM GSDF) using Verilum software and luminance pod (IMAGE Smiths Inc., Germantown, MD, USA) and the calibrated photometer described above.

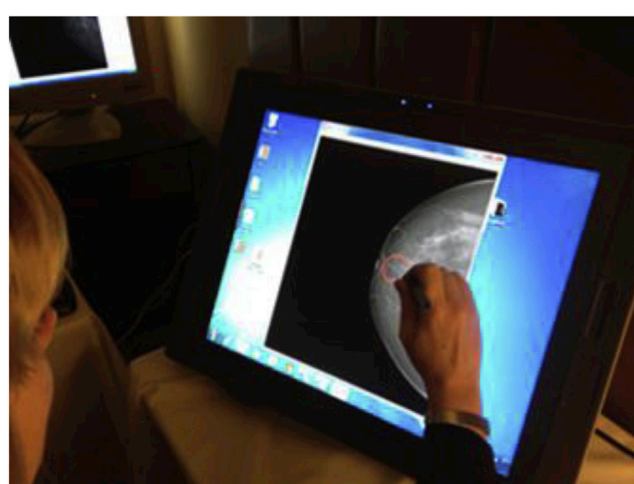

Fig. 1. Interactive digital drawing tablet used in the study.