Expert Systems With Applications 56 (2016) 1–8

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Predicting false negative errors in digital breast tomosynthesis among radiology trainees using a computer vision-based approach

Mengyu Wang[a,b,*], Jing Zhang[a,c], Lars J. Grimm[a], Sujata V. Ghate[a], Ruth Walsh[a], Karen S. Johnson[a], Joseph Y. Lo[a], Maciej A. Mazurowski[a]

[a] Department of Radiology, Duke University School of Medicine, Durham, NC, United States
[b] Schepens Eye Research Institute, Harvard Medical School, Boston, MA, United States
[c] Department of Computer Science, Lamar University, Beaumont, TX, United States

ARTICLE INFO

Article history:
Received 29 November 2015
Revised 22 January 2016
Accepted 25 January 2016
Available online 12 February 2016

Keywords:
Radiology education
False negative error prediction
Training protocol optimization
Tomosynthesis

ABSTRACT

Purpose: Digital breast tomosynthesis (DBT) can improve lesion visibility in comparison to mammography by eliminating breast tissue superimposition. While the benefits of DBT in breast cancer screening rely on well trained radiologists, the optimal training regimen in DBT is unknown. We propose a computer-aided educational system that individually selects the optimal training cases for each trainee. The first step towards this goal is to capture the individual weaknesses of each trainee. In this study, we present and evaluate a computer algorithm for this purpose with particular focus on false negative errors.
Methods: We developed an algorithm (a user model) that predicted the likelihood of a trainee missing an abnormal location. An individual model is applied for each trainee. The algorithm consists of three steps. First, the lesions on DBT images are segmented by a 3D active contour method with a level set algorithm. Then, 16 features are extracted automatically for the segmented lesions. Finally a multivariate logistic regression classifier predicts the likelihood of error based on the extracted features. The classifier is trained using the previous interpretation data of the trainee. We evaluated the individual predictive algorithms experimentally using data from a reader study in which 29 trainees and 3 expert breast radiologists read 60 DBT cases. Receiver operating characteristic (ROC) analysis, along with a repeated holdout approach, was used to evaluate the predictive performance of our algorithm.
Results: The average area under the ROC curve (AUC) of the algorithms which predicted which lesions will be detected and which will be missed by a specific trainee was 0.627 (95% CI: 0.579–0.675). The average performance was statistically significantly better than chance ($p<0.001$). Under the status quo, training involves no specific strategy for case presentation, and this random behavior corresponds to AUC of 0.5. Therefore, the proposed algorithm may provide a significant improvement in distinguishing abnormal locations that will be detected by a trainee from those that will be missed.
Conclusions: Our algorithm was able to distinguish abnormal locations that will be detected by a trainee from those that will be missed. This could be used to enrich the training set with cases that are likely to prompt error for the individual trainee while still maintaining a range of cases necessary for comprehensive education.

1. Introduction

Mammography is the most widely adopted screening modality with a proven capacity for early cancer detection. However, the presence of overlapping breast tissue in mammography can cause abnormalities to be obscured and also lead to unnecessary recalls (Baker & Lo, 2011; Sechopoulos, 2013). Digital breast tomosynthesis (DBT) is a new breast cancer screening technique which is designated to reduce the appearance of overlapping breast tissue, leading to reduced screening recall rates and increased invasive cancer detection rates compared to mammography (Ciatto et al., 2013; Rafferty et al., 2013; Skaane et al., 2013). The benefit of improved cancer detection by eliminating tissue superimposition has been shown for all patients and is not limited to individuals with denser breast tissue (Rafferty et al., 2014).

* Corresponding author at: Department of Radiology, Duke University School of Medicine, Durham, NC, United States. Tel.: +1 4129966190.
E-mail addresses: mengyu_wang@meei.harvard.edu, wmy32552316wmy@gmail.com (M. Wang), jing.zhang@lamar.edu (J. Zhang), lars.grimm@duke.edu (L.J. Grimm), sujata.ghate@duke.edu (S.V. Ghate), ruth.walsh@duke.edu (R. Walsh), karen.johnson2@duke.edu (K.S. Johnson), joseph.lo@duke.edu (J.Y. Lo), maciej.mazurowski@duke.edu (M.A. Mazurowski).

http://dx.doi.org/10.1016/j.eswa.2016.01.053

The rapid adoption of DBT into clinical practice requires a rapid adoption of educational plans to train the next generations of radiologists. The American College of Radiology and the Society of Breast Imaging has recommended that all residents be familiar with DBT (Monticciolo et al., 2013). But it is uncertain whether trainee mammography interpreting skills translate into DBT interpretation skills (Zhang et al., 2015). As there is very little systematic research on DBT education, there is a pressing need to investigate the best educational practices and techniques. In particular, computer aids may be used to enhance training.

Our group has demonstrated that computer-aided personalized education shows promise in mammography (Mazurowski, Baker, Barnhart, & Tourassi, 2010; Mazurowski, Barnhart, Baker, & Tourassi, 2012; Zhang et al., 2014). This topic has also gained interest with other groups (Lin, Yang, & Wang, 2014; Sun, Taylor, Wilkinson, & Khoo, 2008; Voisin, Pinto, Morin-Ducote, Hudson, & Tourassi, 2013). However, computer-aided education in DBT is unexplored. In personalized computer-aided education systems, a user's weaknesses are captured using computer models and then a customized set of training cases is selected for training. In this study, we propose to extend our approach from mammography to DBT. We will focus on the first and crucial component of a computer-aided training system: capturing trainees' error making patterns.

Specifically, in this study we propose an algorithm that analyzes abnormal locations in DBT images and predicts, individually for each trainee, whether the trainee is going to miss them. This is important in an adaptive educational system because it will allow identification of future troublesome cases for each individual trainee. A core component of our prediction approach is the use of automatic computer-extracted image features.

2. Methods

The proposed methodology is described in detail including: (1) the reader study, (2) the definition of error used in this study, (3) the algorithm used to predict false negative error including algorithm for image segmentation, feature extraction and selection, as well as predictive modeling, and (4) the evaluation of the predictive model.

2.1. Reader study

A reader study was conducted in which 3 fellowship-trained breast radiologists and 29 trainees interpreted the same set of 60 DBT cases consisting of craniocaudal (CC) and medio-lateral oblique (MLO) views of a single breast collected at Duke University Medical Center. All images were obtained based on a Siemens prototype MAMMOMAT Novation TOMO and a Siemens MAMMOMAT Inspiration breast tomosynthesis system (Siemens Healthcare, Erlangen, Germany) with a W/Rh anode/filter and 45° total angular span. Images were reconstructed to $85\,\mu m \times 85\,\mu m \times 1$ mm slices by the system using filtered back-projection.

Among the 29 trainees, 2 of them were medical students, 20 of them were radiology residents, 4 of them were non-breast imaging fellows and 3 of them were breast imaging fellows. The 29 trainee readers had varying degrees of breast imaging experience; however, they did not have any previous experience with DBT. Institutional Review Board approval was secured for this study. The readers were instructed to mark suspicious abnormal locations. The abnormal locations were identified by a mouse click. A minor error related to our graphical user interface occurred for 8 readers within 8 views (i.e. only 0.41% of all interpretations of an image view). This error was corrected manually after the study.

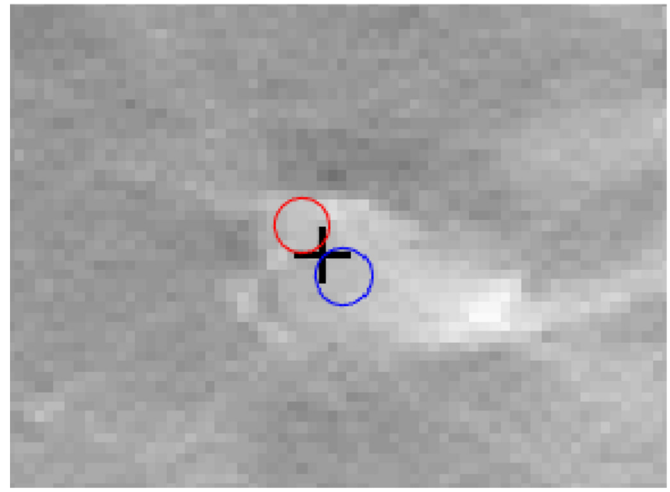

Fig. 1. Representation of the lesion location determined by the agreement of at least 2 expert radiologists: the circles refer to the marks by the radiologists and the cross is the average mark location (i.e. the ground truth).

2.2. Definition of error

Lesion locations were determined by the three expert radiologists. If at least 2 experts marked a lesion within a distance of 9 mm from each other, the centroid location of the expert marks was considered as the location of the lesion. A distance of 9 mm was chosen because it was the average radius of breast lesions seen on a previously reported series (Timp, Karssemeijer, & Hendriks, 2003). Note that the distance is calculated in 3 dimensions between the expert mark locations and not on a single slice only. The lesion locations found by the experts were deemed as the ground truth and later used to determine if the trainee detected a lesion. Fig. 1 shows a representative example of the lesion location marks by the experts and the final average lesion location (i.e., ground truth). Although all types of abnormalities including masses, microcalcifications, asymmetries and architectural distortions were present in the study cases, we excluded microcalcifications from the analysis to focus on mass-like abnormalities only. Error making and image analysis for microcalcifications are distinctly different from mass-like abnormalities.

We defined error as follows: if a trainee did not put a mark within a distance of 9 mm to the ground truth lesion location, then the lesion was considered undetected by the trainee. Again the distance was calculated in 3 dimensions between the lesion location and the trainee mark locations. Please note that while some studies on reader performance use biopsy results as the ground truth, we believe that our approach of using the expert consensus as the ground truth allows us to test for interpretation errors due to insufficient training, and exclude the limitations of the imaging modality that do not allow for correct interpretation even by expert readers. These limitations could include insufficient resolution, low contrast, or imaging artifacts.

2.3. Algorithm for predicting error

The proposed algorithm for predicting the likelihood of error in DBT was as follows. First, the lesion regions in the DBT images were segmented using image processing algorithms. Then, features were extracted to describe the properties of each lesion within the background context. Lastly, a classifier was applied to predict the likelihood of the lesion being missed by each trainee using the extracted image features. The overall structure of the proposed algorithm is shown in Fig. 2. Note that in our modeling the positive