Journal of Biomedical Informatics 54 (2015) 50–57

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Modeling false positive error making patterns in radiology trainees for improved mammography education

Jing Zhang [a,e,*], James I. Silber [b], Maciej A. Mazurowski [a,c,d]

[a] Department of Radiology, Duke University School of Medicine, Durham, NC, United States
[b] Department of Biomedical Engineering, Duke University Pratt School of Engineering, Durham, NC, United States
[c] Duke Cancer Institute, United States
[d] Duke Medical Physics Program, United States
[e] Computer Science Department, Lamar University, Beaumont, TX, United States

ARTICLE INFO

Article history:
Received 17 March 2014
Accepted 19 January 2015
Available online 30 January 2015

Keywords:
Breast cancer
Radiology education
Computer vision
Machine learning

ABSTRACT

Introduction: While mammography notably contributes to earlier detection of breast cancer, it has its limitations, including a large number of false positive exams. Improved radiology education could potentially contribute to alleviating this issue. Toward this goal, in this paper we propose an algorithm for modeling of false positive error making among radiology trainees. Identifying troublesome locations for the trainees could focus their training and in turn improve their performance.
Methods: The algorithm proposed in this paper predicts locations that are likely to result in a false positive error for each trainee based on the previous annotations made by the trainee. The algorithm consists of three steps. First, the suspicious false positive locations are identified in mammograms by Difference of Gaussian filter and suspicious regions are segmented by computer vision-based segmentation algorithms. Second, 133 features are extracted for each suspicious region to describe its distinctive characteristics. Third, a random forest classifier is applied to predict the likelihood of the trainee making a false positive error using the extracted features. The random forest classifier is trained using previous annotations made by the trainee. We evaluated the algorithm using data from a reader study in which 3 experts and 10 trainees interpreted 100 mammographic cases.
Results: The algorithm was able to identify locations where the trainee will commit a false positive error with accuracy higher than an algorithm that selects such locations randomly. Specifically, our algorithm found false positive locations with 40% accuracy when only 1 location was selected for all cases for each trainee and 12% accuracy when 10 locations were selected. The accuracies for randomly identified locations were both 0% for these two scenarios.
Conclusions: In this first study on the topic, we were able to build computer models that were able to find locations for which a trainee will make a false positive error in images that were not previously seen by the trainee. Presenting the trainees with such locations rather than randomly selected ones may improve their educational outcomes.

1. Introduction

Mammography is the most widely used screening technique for breast cancer early detection, which plays an important role in reducing the mortality of breast cancer. However, interpretation of mammograms is a very challenging task due to overlapping tissue that might both obscure signs of cancer (false negative errors) as well as create patterns that resemble true abnormalities and unnecessarily alert a radiologist (false positive errors) [3].

Our group has been working on the development of an adaptive computer-aided education system for mammography education. Specifically, in [12], we proposed a general framework for such a system and demonstrated that image features can be used to predict errors made by a trainee. In [13], we presented models for prediction of errors in assignment of BI-RADS features of masses and images. In [14], we investigated the use of collaborative filtering algorithms to model resident errors in mammography. Other work on the adaptive mammography education is limited; however, some related studies are available. Sun et al. [18,19] presented

* Corresponding author at: 2424 Erwin Road, Suite 302, Durham, NC 27707, United States.

E-mail address: jing.zhang2@duke.edu (J. Zhang).

http://dx.doi.org/10.1016/j.jbi.2015.01.007
1532-0464/

initial studies on developing an ontology related educational training system based on differences between radiologists. The studies by Mello-Thoms et al. [15], Tourassi et al. [22], Voisin et al. [23] investigate visual attention and spatial frequency representations, human perception and cognition, and eye gaze tracking to study error making in mammography. Some work in computer-aided detection is also relevant to our study in terms of the computer vision methods used, such as the studies presented by Masotti et al. [11], Wei et al. [24], and Mudigonda et al. [16].

In this paper we focus on a topic largely unexplored in the context of radiology education: false positive error making. Specifically, the task that we approach is to automatically find locations that will cause a trainee to make a false positive error. For this purpose, we propose an algorithm that identifies challenging locations using computer vision algorithms and machine learning models. The models are constructed individually for each trainee based on their prior interpretations to capture their individual error making patterns.

To our knowledge, this is the first study in which future false positive locations are predicted. It differs from our previous studies in which we focused on false negative errors [8], errors in distinguishing benign and malignant masses [12], and errors in assessment of BI-RADS features [13]. Predicting false positive locations is a difficult task as it requires analysis of the entire image and finding those locations that might cause difficulty to the trainee while dismissing all the locations that will not. While our experiments confirm the high difficulty of the task, they also show the promise of our approach. One practical application of our approach is to identify locations that would result in false positive errors for each trainee so that they can focus their training on such locations, potentially improving their training.

2. Reader study and the definition of false positive errors

To validate our algorithm for predicting false positive errors, we used data from a reader study in which 10 radiology trainees along with 3 expert radiologists interpreted 100 mammographic cases independently. Among the 10 trainees, 7 were radiology residents with at least four weeks of formal breast imaging training and 3 were novices (2 medical imaging researchers and 1 medical student) with no formal training. We included the three novices to simulate radiology residents at the very beginning of their residency program. The three expert radiologists were all fellowship trained in breast imaging with 7–14 years of experience. The experts and the trainees were not aware of patients' age and medical history. The 100 mammographic cases are balanced with 50 cases originally deemed as normal and 50 abnormal cases. Each case contained 4 standard mammographic views: left craniocaudal (LCC), right craniocaudal (RCC), left mediolateral oblique (LMLO), and right mediolateral oblique (RMLO). All participants were asked to identify actionable abnormalities by clicking on them. We asked the participants to ignore microcalcifications as the focus of our study was on masses. Institutional Review Board approval was secured for this study.

We used the marks provided by the three experts to find the actual actionable masses. Specifically, if a region contained at least two out of three experts' marks and the distance between two marks was smaller than a predefined threshold T_d, we considered this region to be associated with an actionable mass. The centers of actual actionable masses were determined as the centroids of the expert annotations. Consequently, if the distance between a trainee's mark and its nearest actionable mass center is bigger than T_d, this mark is defined as a false positive error. Otherwise, it is defined as a true positive. Because the average radius of the breast masses is 9 mm [21] and the pixel spacing of the images used in the reader study was 0.0941 mm, the threshold was set to T_d = 9 mm/0.0941 mm = 96 pixels in our study.

3. The algorithm for prediction of false positive locations

3.1. Overview

In this paper, we propose an algorithm that searches through an entire mammographic image to find locations where the trainee made false positive errors. The proposed algorithm accepts an entire mammographic image as the input and returns locations that are more likely to be associated with a false positive error as the output. The algorithm is composed of 3 steps:

Step 1. Step 1: The Difference of Gaussian (DoG) filter [2] is adopted to identify suspicious false positive locations of an image, and then rubber band and region growing methods are used to segment suspicious false positive regions using local maximum points extracted from the DoG filter response map;
Step 2. Step 2: Features are extracted to describe the properties of each region and its context; and
Step 3. Step 3: A classifier is applied to predict the likelihood of a predicted location being a false positive error made by the trainee using the extracted features.

The flowchart of the proposed algorithm is shown in Fig. 1. The three steps of the algorithm are described in the subsections below.

3.2. Step 1: Identifying suspicious locations

Difference of Gaussian (DoG) filter, which has been widely used for breast mass detection [17,6], is adopted in our study as the first step to identify the suspicious locations where the trainees may make false positive errors (i.e., click on the location). After calculating the DoG filter response for the entire image, we extract local maximum points from the DoG filter response map and consider these locations suspicious. Then, by using the identified suspicious locations as reference points, three segmentation methods (dynamic programming-based rubber band, region growing with adaptive threshold, and region growing with fixed threshold) are applied to segment the abnormality or the abnormality resembling region. These segmentations will be later used to determine features of locations. The segmentation algorithms used are described below.

The dynamic programming-based rubber band method [20] can transform a round image region to a rectangular region in a polar coordinate system. Gradient, size, and intensity information extracted from the image in the polar system are combined to form a cost matrix. The boundary of the region is the path that has the lowest cost in the cost matrix determined by dynamic programming. The region growing method [1] segments a region by computing the similarity between the given seed region and its neighboring pixels iteratively. If the similarity is smaller than a predefined threshold, the seed region is grown by including its neighboring pixels. The method stops when no new pixels can be included. Two seed region growing strategies of region growing method were adopted in this study: one with a fixed seed region and the other with an adaptive seed region that is updated at each iteration. The purpose of using three different segmentation methods (the two variations of the region growing algorithm was treated as two different segmentation algorithms) is to be able to compute features indicating segmentation difficulty by comparing the three segmentation results.