Ultrasound in Med. & Biol., Vol. ■, No. ■, pp. 1–15, 2015

0301-5629/$ - see front matter

ELSEVIER

http://dx.doi.org/10.1016/j.ultrasmedbio.2015.07.020

● *Original Contribution*

GOING BEYOND A FIRST READER: A MACHINE LEARNING METHODOLOGY FOR OPTIMIZING COST AND PERFORMANCE IN BREAST ULTRASOUND DIAGNOSIS

Santosh S. Venkatesh,* Benjamin J. Levenback,† Laith R. Sultan,† Ghizlane Bouzghar,† and Chandra M. Sehgal†

*Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, Pennsylvania, USA; and
†Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

(*Received* 27 *February* 2015; *revised* 16 *June* 2015; *in final form* 16 *July* 2015)

Abstract—The goal of this study was to devise a machine learning methodology as a viable low-cost alternative to a second reader to help augment physicians' interpretations of breast ultrasound images in differentiating benign and malignant masses. Two independent feature sets consisting of visual features based on a radiologist's interpretation of images and computer-extracted features when used as first and second readers and combined by adaptive boosting (AdaBoost) and a pruning classifier resulted in a very high level of diagnostic performance (area under the receiver operating characteristic curve = 0.98) at a cost of pruning a fraction (20%) of the cases for further evaluation by independent methods. AdaBoost also improved the diagnostic performance of the individual human observers and increased the agreement between their analyses. Pairing AdaBoost with selective pruning is a principled methodology for achieving high diagnostic performance without the added cost of an additional reader for differentiating solid breast masses by ultrasound. (E-mail: sehgalc@uphs.upenn.edu) © 2015 World Federation for Ultrasound in Medicine & Biology.

Key Words: **Breast ultrasound, Breast cancer, Adaptive boosting, Computer-aided diagnosis, Artificial intelligence.**

INTRODUCTION

Considerable effort is being devoted to improve breast ultrasound for differentiating solid malignant and benign masses (Candelaria et al. 2013; Sehgal et al. 2006). Progress toward this goal is being made *via* the integration of ultrasound and mammography imaging modes (Padilla et al. 2013) and the introduction of new modes of ultrasound imaging like elastography (Chang et al. 2013; Golatta et al. 2013), 3-D imaging (Cho et al. 2005, 2006; McDonald 2011; Ruiter et al. 2012; Watermann et al. 2005) and computer-aided tomography (Duric et al. 2007). These technological developments have spurred the evolution of new computer-based algorithms to assist radiologists with breast cancer diagnosis using clinical ultrasound. These studies have been reviewed (Huang 2009; Sehgal et al. 2006). More recent investigations have extended the use of computer features for automated mass detection and classification in three dimensions with ultrasound images (Cheng et al. 2010; Liu et al. 2010). The state of the art is, however, still not entirely satisfactory; despite these advances in both breast imaging technology and image analysis, the biopsy yield continues to be low, and as many as 70% to 85% of biopsies prove to be benign. The costs of this low yield are the emotional trauma experienced by patients whose masses are ultimately determined to be benign and the socioeconomic costs to society as a whole imposed by a large number of unneeded procedures (Kopans 1992). One of the main reasons for such low yield is that the false negatives have major consequences related to patient mortality. Improving the accuracy of prediction to reduce the number of unneeded biopsies while keeping the false-negative rates to a minimum, possibly approaching zero, continues to be an important objective in the state of the art.

Several studies have indicated that multiple readings of mammograms improve diagnostic performance for breast cancer diagnosis (Georgian-Smith 2007; Gromet 2008; Taylor and Potts 2008; Waldmann et al. 2012). Waldman et al. reported that double reading of

Address correspondence to: Chandra M. Sehgal, Department of Radiology, University of Pennsylvania, 3400 Spruce Street, Philadelphia, PA 19104, USA. E-mail: sehgalc@uphs.upenn.edu

mammograms increased the tumor detection rate from 14.6 to 16.4 per 1000 cases. On the basis of their results they concluded that double reading is crucial not only for screening, but also for lesion characterization. Gromet also observed that with the double reading process, sensitivity rose from 81.4% to 88.0%, with a gain of 8.2% cancers detected. Similarly, Taylor et al. reported that double reading with arbitration increased the detection rate and decreased the recall rate. Experience with mammography suggests that the benefits of double or multiple readings can be expected to extend to ultrasound imaging. A recent study found that double reading of breast ultrasound improved diagnostic performance (Bouzghar et al. 2014). Unfortunately, it is not a simple matter to deploy multiple readers in clinical settings. As may be expected, the generation of multiple readings is labor intensive, costly, time consuming and limited by the scarcity of specialized radiologists skilled in interpreting breast ultrasound images. It is in this context that the advent of automated methods for feature extraction provides a viable real-time, low-cost alternative to a second reader to help augment physicians' interpretations of images.

The availability of this computerized resource opens the door to a principled combination of visual and machine-generated features using an adaptive machine learning procedure that incorporates the strengths of each feature set, while simultaneously identifying the small set of cases for whom the images are intrinsically ambiguous and merit further evaluation by additional imaging. In this study we describe such a computer-based system to complement a radiologist's interpretation of the ultrasound Breast Imaging Reporting and Data System ($BI\text{-}RADS_{US}$) as a second reader. The automated process of machine generating a second feature set is not only a low-cost procedure, but has the added advantage of generating features that are independent of those generated visually by a trained radiologist in the sense that the features extracted are qualitatively different. The computerized system combines *a priori* information and expert human knowledge in Bayesian settings with the logistic regression probability of computerized features to improve diagnostic decisions.

Next, we provide an overview of methodologic tools and algorithms. Then, we describe the results obtained by using the algorithms and patient data. We discuss the role of adaptive boosting and selective pruning in reducing cost and improving diagnostic performance and make our conclusions.

METHODS AND ALGORITHMS

Overview

Two independent feature sets representing readers 1 and 2 were constructed and classified for each image in the library of breast ultrasound images using a radiologist's interpretation of ultrasound $BI\text{-}RADS_{US}$ and computerized features extracted from the images (Fig. 1, step a). The radiologist's interpretations of ultrasound images ($BI\text{-}RADS_{US}$) were combined with the computer-generated features using *adaptive boosting* or the consensus method (Fig. 1, step b). This process was implemented to expand the discriminatory region of each feature set by incorporating the strengths of each set. Despite combining the regions of strength of each of the two independent feature sets, some cases remained persistently ambiguous. These cases representing the low-confidence group were pruned from the data set for further evaluation by additional imaging (Fig. 1, step c). The remaining cases, representing the high-confidence group, were evaluated for their diagnostic performance (Fig. 1, step d).

General methods

Ultrasound images of 264 solid masses from 246 patients were analyzed for this study with approval

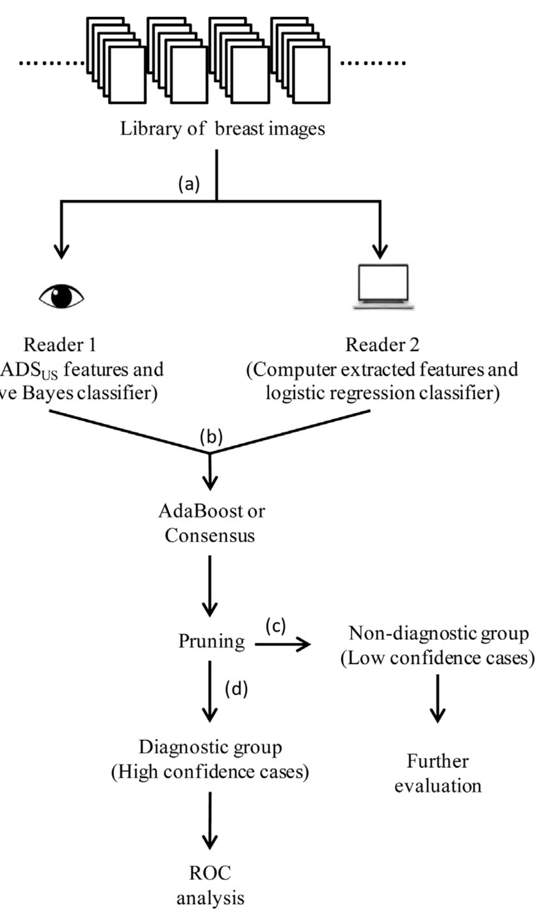

Fig. 1. Overview of the methods and procedures. $BIRADS_{US}$ = ultrasound Breast Imaging Reporting and Data System, ROC = receiver operating characteristic.