Expert Systems With Applications 46 (2016) 106–121

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Classifier ensemble generation and selection with multiple feature representations for classification applications in computer-aided detection and diagnosis on mammography

Jae Young Choi [a,1], Dae Hoe Kim [a], Konstantinos N. Plataniotis [b], Yong Man Ro [a,*]

[a] Image and Video Systems Lab, Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Yuseong-Gu, Daejeon 305-701, Republic of Korea
[b] Knowledge Media Design Institute, Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario M5S 3GA, Canada

ARTICLE INFO

Keywords:
Ensemble learning
Ensemble selection
Classification
Mammographic masses
Multiple feature representations
Computer-aided Detection (CADe)
Computer-aided Diagnosis (CADx)

ABSTRACT

This paper presents a novel ensemble classifier framework for improved classification of mammographic lesions in Computer-aided Detection (CADe) and Diagnosis (CADx) systems. Compared to previously developed classification techniques in mammography, the main novelty of proposed method is twofold: (1) the "combined use" of different feature representations (of the same instance) and data resampling to generate more diverse and accurate base classifiers as ensemble members and (2) the incorporation of a novel "ensemble selection" mechanism to further maximize the overall classification performance. In addition, as opposed to conventional ensemble learning, our proposed ensemble framework has the advantage of working well with both weak and strong classifiers, extensively used in mammography CADe and/or CADx systems. Extensive experiments have been performed using benchmark mammogram dataset to test the proposed method on two classification applications: (1) false-positive (FP) reduction using classification between masses and normal tissues, and (2) diagnosis using classification between malignant and benign masses. Results showed promising results that the proposed method (area under the ROC curve (AUC) of 0.932 and 0.878, each obtained for the aforementioned two classification applications, respectively) impressively outperforms (by an order of magnitude) the most commonly used single neural network (AUC = 0.819 and AUC =0.754) and support vector machine (AUC = 0.849 and AUC = 0.773) based classification approaches. In addition, the feasibility of our method has been successfully demonstrated by comparing other state-of-the-art ensemble classification techniques such as Gentle AdaBoost and Random Forest learning algorithms.

1. Introduction

Breast cancer is the most common form of cancer among women and is the second leading cause of death (Kopans, 2007). To reduce the workload of radiologists and to improve the specificity and sensitivity in detection of breast cancer, two different types of automated screening systems are being developed (Suri & Rangayyan, 2006): (1) Computer-aided Detection (CADe) and (2) Computer-aided Diagnosis (CADx). Table 1 provides a brief review of CADe and CADx systems. Current CADe and/or CADx systems have been clearly shown to be quite sensitive in its ability to detect cancer, but one of their main drawbacks is the high number of FPs (defined in Table 1) (Suri & Rangayyan, 2006; Sampat, 2005). Hence, high FP rate for mass detection and diagnosis remains to be one of the major problems to be resolved in CADe/CADx study (Suri & Rangayyan, 2006; Sampat, 2005; Tang, Rangayyan, Xu, Naqa & Yang, 2009).

In typical CADe (or CADx) systems, classifier design is one of the key steps for determining FP rates (Suri & Rangayyan, 2006; Sampat, 2005). Thus far, research efforts have mostly been focused on the design of the *single* classifier in both CADe and CADx systems (Suri & Rangayyan, 2006; Sampat, 2005; Tang et al., 2009; Chan, Sahiner, Wagner, & Petrick, 1999). It should be noted that there are two critical limitations within the classifier design process in mammogram images. First, the large variability in the appearance of mass patterns (Cheng et al., 2005; Velikova, Lucas, Samulski, & Karssemeijer, 2013) – due to its irregular size, obscured borders, and complex mixtures of margin types – makes classification task quite difficult. Second, research in mammography is characterized by a restricted training data, due to cost, time, and availability to patient medical information

* Corresponding author. Tel: +82 42 350 3494; fax: +82 42 350 7619.
E-mail addresses: jygchoi@kaist.ac.kr, jyoung.choi@jwu.ac.kr (J.Y. Choi), dhkim10@kaist.ac.kr (D.H. Kim), kostas@ece.utoronto.ca (K.N. Plataniotis), ymro@ee.kaist.ac.kr (Y.M. Ro).
[1] Present address: Department of Biomedical Engineering, Jungwon University, Chungcheongbuk-do, Republic of Korea.

http://dx.doi.org/10.1016/j.eswa.2015.10.014
0957-4174/

Table 1
A brief review of Computer-aided Detection (CADe) and Diagnosis (CADx).

	Computer-aided Detection (CADe)	Computer-aided Diagnosis (CADx)
Clinical role	Aiding radiologists find breast cancer on screening mammograms by localizing suspicious regions in an image	Finding radiologists decide whether a suspicious lesion is benign or malignant on diagnostic mammograms by providing a malignancy estimate of a given lesion
Classification application	Classifying suspicious regions into lesion versus normal tissue to reduce the number of false-positives that are produced at the end of the stage of detection of suspicious regions	Classifying known lesions into malignant and benign categories to incorporate the classification output into radiologists' final decision on the likelihood of malignancy
Positive case	Lesion (regardless of whether it is a malignant or benign)	A biopsy-proven malignant lesion
Negative case	A normal tissue	A biopsy-proven benign lesion; benign lesion proven by follow up
False positive case	A region being normal tissue but interpreted by the automatic algorithm as a suspicious one	A benign lesion but interpreted by the automatic algorithm as a malignant one.

and mammography images (Suri, & Rangayyan, 2006; Bilska-Wolak & Floyd, 2004). On the other hand, the number of available features (arising from the integration of multiple heterogeneous feature types) is large (Cheng et al., 2005; Jesneck, Nolte, Baker, Floyd, & Lo , 2006; Wei et al., 1997) (typically, in the thousands) relative to the number of training samples, so-called curse of dimensionality (Kuncheva, 2004). For these reasons, a *single classifier design* may face a great challenge in achieving a level of FP reduction that meets the requirement of clinical applications.

In this paper, to overcome the aforementioned limitations, we propose a new and novel ensemble classifier framework for classification applications (explained in Table 1) in mammographic CADe and CADx. This paper improves and extends preliminary work presented in Choi, Kim, Plataniotis, and Ro, (2012). In particular, this paper presents a new *ensemble selection* approach for selecting an optimal subset of base classifiers, aiming to further improve generalized (testing) classification performances. An improved ensemble generation technique is also outlined in the paper by introducing an advanced mechanism that allows the use of strong classifiers extensively used in mammography computer-aided detection and diagnosis systems. In addition, more insightful discussion of our ensemble generation on the local learner hypothesis viewpoint is provided. Moreover, we report integrated experimental results that are more extensive and rigorous in the following aspects: (1) additional assessment of our proposed ensemble classification on computer-aided diagnosis application; (2) the comparison of other state-of-the-art ensemble classification techniques; (3) comprehensive analysis using more classifier models.

The contents of the paper are organized as follows: Section 2 reviews previous work on classification of breast masses on mammograms in CADe and CADx systems. In Section 3, the region-of-interests (ROIs) segmentation and feature extraction methods used in our study are briefly described. Section 4 explains in detail the proposed ensemble classification framework. Section 5 contains the details of the image databases, and experimental setup and condition. In Section 6, we present a series of experimental results to demonstrate the effectiveness of the proposed method. Finally, concluding remarks are provided in Section 7.

2. Related work

In past years, considerable research efforts have been directed to classifier design aiming at classification applications in mammography. Wei et al. (1997) used global and local texture features extracted from manually selected ROIs of digitized mammograms, and linear discriminant analysis (LDA) to classify the masses from normal glandular tissues to minimize FP detections. Sahiner et al. (1996) proposed a convolution neural network (NN) for the task of discriminating between masses and normal tissues using texture features. Wei et al. (1995) developed a NN classifier based on multiresolution texture features extracted from the spatial gray level dependence (SGLD) matrices for distinguishing masses from normal tissues. In Mudigonda, Rangayyan, and Desaultels (2001), the four texture features, namely contrast, coherence ratio, entropy of orientation, and variance of coherence-weighted angular estimates, were extracted based on textual flow-field analysis and were used to reduce FP detections. In Junior, Rocha, Gattass, Silva, and Paiva (2013), the features of ROI regions were generated by using spatial diversity texture analysis; these generated features were subsequently applied to support vector machine (SVM) classifier for FP reduction. Krishnan, Banerjee, Chakraborty, Chakraborty, and Ray (2010) incorporated the best-possible training scheme into the SVM-based classifier. In their method, the kernel function is first selected and then an appropriate training-test partition is applied for improving classification.

Various classification techniques have been also employed for classifying masses into malignant and benign in CADx systems. Most of the typical pattern classifiers have been explored, including decision tree, k-nearest-neighbors (k-NN), NN, LDA, and SVM (Sampat, 2005; Tang et al., 2009). Among them, NN has been the most widely used (Verma, 2008; Sampat, 2005), while the SVM is the state-of-the-art classifier for CADx systems (Mavroforakis, Georgiou, Dimitropoulos, Cavouras, & Theodoridis, 2006; Nanni, Brahnam, & Lumini, 2012). The following is a brief review of some representative methods for classification of malignant versus benign masses. In Georgiou et al. (2007), morphological shape features and fractal dimension analysis were applied to various classifier approaches including LDA, k-NN, multi-layered perceptron NN, and SVM. It has been reported in (Georgiou et al., 2007) that SVM achieved the highest performance for the classification of masses into benign and malignant. The authors in (Verma, 2008) proposed a new NN architecture and a training strategy for the diagnosis of breast masses on mammograms. They suggested the introduction of additional neurons in hidden layer for benign and malignant classes and adjustment solution during the calculation of weights. Way, Sahiner, Hadjiiski, and Chan (2010) investigated the effect of finite sample size on LDA and SVM classifiers on the performance for classifying malignant versus benign masses. In Mavroforakis et al. (2006), the localized texture features together with a wide range of linear and nonlinear classification architectures (LDA, NN, and SVM, etc.) were used for characterizing mammographic masses. Results showed that SVM outperforms any linear approaches under consideration. Verma, McLeod, and Klevansky (2010) developed a so-called soft clustered based direct learning NN classifier that makes use of clustering algorithms to create soft clusters (sub-classes) within benign and malignant classes, aiming to correctly classify suspicious regions into benign and malignant patterns. Sahiner, Chan, Petrick, Helvie, and Hadjiiski (1998a) proposed the combined use of the LDA classifier and the rubber band straightening transform (RBST) texture features. In Nanni et al. (2012), a set of SVM classifiers are produced using an ensemble of texture descriptors, such as Local Ternary Pattern (LTP). Multiple results of individual SVMs are then combined using sum rule for classifying benign and malignant breast tissues.

In recent years, ensemble classification techniques have gradually gained interesting attention and their popularity among CADe and/or