

Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data

Wenqing Sun^a, Tzu-Liang (Bill) Tseng^b, Jianying Zhang^{c,d}, Wei Qian^{a,d,*}

^a Department of Electrical and Computer Engineering, University of Texas at El Paso, 500 West University Avenue, El Paso, TX, 79968, United States

^b Department of Industrial, Manufacturing & Systems Engineering, University of Texas at El Paso, 500 West University Avenue, El Paso, TX, 79968, United States

^c Department of Biological Sciences, University of Texas at El Paso, 500 West University Avenue, El Paso, TX, 79968, United States

^d Sino-Dutch Biomedical and Information Engineering School, Northeastern University, No. 11, Lane 3, Wenhua Road, Heping District, Shenyang, Liaoning, 110819, China

ARTICLE INFO

Article history:

Received 5 November 2015

Received in revised form 14 July 2016

Accepted 18 July 2016

Keywords:

Semi-supervised learning

Deep learning

Computer aided diagnosis

Convolutional neural network

Unlabeled data

ABSTRACT

In this study we developed a graph based semi-supervised learning (SSL) scheme using deep convolutional neural network (CNN) for breast cancer diagnosis. CNN usually needs a large amount of labeled data for training and fine tuning the parameters, and our proposed scheme only requires a small portion of labeled data in training set. Four modules were included in the diagnosis system: data weighing, feature selection, dividing co-training data labeling, and CNN. 3158 region of interests (ROIs) with each containing a mass extracted from 1874 pairs of mammogram images were used for this study. Among them 100 ROIs were treated as labeled data while the rest were treated as unlabeled. The area under the curve (AUC) observed in our study was 0.8818, and the accuracy of CNN is 0.8243 using the mixed labeled and unlabeled data.

© 2016 Published by Elsevier Ltd.

1. Introduction

Computerized breast cancer diagnosis (CADx) system has played an important role in the early cancer diagnosis and treatment for the last two decades, which combined with improved treatment strategies significantly reduced patients' mortality and morbidity rates (Tabar et al., 2001; Smith et al., 2015). In CADx development, groups of researchers have designed and tested many computational features including morphological features, texture features and intensity features. For example, area, circularity, ratio of semi-axis are very typical and useful morphological features (Qian et al., 2007); average intensity, mean gradient of region boundary, density uniformity are very common density features used for mass detection (Qian et al., 1999); wavelet features, gray-level co-occurrence matrix (GLCM) features, run length features are powerful texture features we used in our previous researches (Sun et al., 2015a, 2015b, 2014a, 2016a, 2016c). From the combined fea-

ture pool, the feature selection method was usually used to choose certain combination of features and the selected features were sent to classifiers for training and prediction. However, many factors are still restricting the efficacy of screening mammography and CADx system due to the large variability of breast abnormalities, overlapping dense fibro-glandular tissues, and the low detection rate (Tan et al., 2014). Feature design is considered as an essential module for most existing CADx which it is a time consuming and complicated task. Recent days, deep learning algorithms demonstrated the superior performance in many computer vision tasks. And researchers started to seek for solutions by migrating the deep learning applications into medical image areas, and the results are promising (Cireşan et al., 2013; Kallenberg et al., 2016; Sun et al., 2016b; Sun and Tseng, 2016). Instead of using hand-crafted features, these deep structured algorithms can automatically generate the features at multiple levels of abstraction.

However, deep learning algorithms require large amount of training data for fine tuning the parameters. But for most medical image projects, acquiring data base with enough labeled data is very difficult (Shiraishi et al., 2009), and the process of making diagnosis on such large number of cases is a heavy burden for radiologists. To label all these data correctly, we need several radiologists to evaluate all the data individually, then compare their evaluation results with one another, and discuss the results case by case to

* Corresponding author at: Department of Electrical and Computer Engineering, Director: Medical Imaging Informatics Laboratory, College of Engineering, University of Texas, El Paso, 500 West University Avenue, El Paso, Texas, zip code 79968, United States.

E-mail address: wqian@utep.edu (W. Qian).

URL: <http://ee.utep.edu/facultyqian.htm> (W. Qian).

reach a final conclusion. Besides these heavy workload, confidentiality agreement is another obstacle for acquiring these labeled data, and typically neither the doctors nor the patients are willing to reveal this information. However, at the same time, abundance of unlabeled data is relatively easier to acquire in most situations and will significantly reduce the problem of confidential sensitivity (Jamieson et al., 2010).

In this study we hypothesize that combining a small amount of labeled data with a large amount of unlabeled data is one solution to collect enough data to train the deep learning algorithms. Using both labeled data and unlabeled data at the same time is not the new concept, and this technique is called semi-supervised learning (SSL) in machine learning area. The procedure allows computer to integrate the information from labeled data and unlabeled data at the same time without any user intervention. It starts with the initial labeled dataset and gradually adding unlabeled data to the labeled dataset. Co-training (Mitchell and Blum, 1998) is one of the most popular and promising SSL algorithms, it trains two classifiers by labeling the unlabeled data for the other classifier, and makes decisions based on the agreement of these two classifiers. The drawback of co-training is it requires two sufficient and redundant feature subsets for each classifier, which can be hardly satisfied in CADx systems (Kiritchenko and Matwin, 2001; Wan, 2009).

To test this hypothesis, we designed a graph based SSL algorithm and a deep convolutional neural network (CNN), and we tested the performance of the algorithms using a small amount of labeled data and large amount of unlabeled data. The CNN is a deep learning algorithm that does not require unsupervised pre-training, and the structure we developed in this study is based on LeCun's model (LeCun et al., 1989, 1998). The proposed SSL scheme consists of three modules: data weighing, feature selection and a newly proposed dividing co-training data labeling. The details of our proposed methods and the results are described in the following sections.

2. Materials and methods

2.1. Data

1874 pairs of mammogram images were used from our established in-house full-field digital mammography (FFDM) image database, and each case included one pair of cranio-caudal (CC) view and one pair of mediolateral-oblique (MLO) view breast images. The ages of participants range from 32 to 74, and the average age is 51. From these mammogram images, 3158 region of interests (ROIs) with each containing a mass extracted based on the radiologists' notations, and 1434 of them were diagnosed as malignant and the rest 1724 were diagnosed as benign. All these masses were extracted from the 2472 images including 1580 images with one detected mass, 368 images with two detected masses, and 36 images with three or more detected masses. All the data we used in this research are labeled data, and for the evaluation purpose, the truth files (practitioner diagnosis) from some of the data were removed and treated as unlabeled data. The distribution of Breast Imaging Reporting and Data System (BIRADS) ratings, mass types and ROI types are shown in Fig. 1.

2.2. Methodology

The proposed method uses only a small amount of labeled data and large amount of unlabeled data for the deep neural network algorithm. All the mammograms with detected masses mentioned above were preprocessed using the techniques from our previous publications (Qian et al., 2001; Sun et al., 2015a). The truth files of the majority of the data are covered to simulate the unlabeled

data in this study, and the algorithm labeled data were combined with originally labeled data for the training deep neural network. Fig. 2 demonstrates the flowchart of our proposed method, and the details of each module are described below.

2.2.1. Extracted features

For both labeled and unlabeled ROIs, all the computational features were identical, including the features extracted from inside mass areas and outside. Besides the features inside the mass regions, the information from the surrounding areas or even whole breast, like breast density, spiculations, can be used as indicators of women having or developing breast abnormalities and cancers (Wolfe, 1976; Sampat et al., 2006; Qian et al., 1999). According to the reported work by Wang et al. (2012), using the automatically computed mammographic density features in CAD schemes made it possible to achieve better performance compared to using manually assessed density by the radiologists. The local features were calculated from the mass region itself; the global features measure the global density from the whole breast image and the textures of the surrounding mass areas.

For each mass we extracted 21 computational features respectively. There were 7 basic local features calculated inside of each mass region: area, circularity, ratio of semi-axis, longest radial length, mean intensity, intensity variation, and mass boundary irregularity factor. Then we calculated the grey level co-occurrence matrix (GLCM) (Haralick et al., 1973) of the mass region, and measured mean, variance, contrast, correlation, entropy and homogeneity features of the matrix. For the global features we computed eight features: number of spiculations around masses, average speculation length, number of spiculations exceeds certain threshold, spiculation angular distribution, mean and variation of intensity of the surrounding background of the masses, mean and variation of intensity of the whole breast areas.

From these 21 features, we applied dimension reduction method to them to reduce the feature correlations and redundancies. In this study, we tested three popular dimension reduction methods: Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Multidimensional Scaling (MDS).

2.2.2. Data weighing and confident data labelling

Since the data labelling process starts with the initial labeled data, the influence of noisy data or atypical data (outliers) will be amplified after labelling more data based on the biased criterions. To minimize such effect, we initialized the SSL procedure with data weighing module. Inspired by DROP3 instance selection method (Hall, 1999), we developed three data weighing equations using exponential function, Gaussian function, and Laplacian function. Compared to the instance selection methods, we assigned different weights to different data instead of eliminating the atypical data. For each case L_i , we calculated the accuracy A_k of keeping the case in the original labeled dataset and accuracy A_r of removing the cases from the labeled dataset. Then we assigned the weight to each case using the weighing function, and for the comparison reason, we tested three data weighing formulas:

$$W_i = e^{A_k - A_r} \quad (1)$$

$$W_i = \frac{1}{2} e^{(A_k - A_r + 1)^2} \quad (2)$$

$$W_i = \frac{1}{2} e^{|A_k - A_r + 1|} \quad (3)$$

With the weights of all instances computed, the unlabeled data were gradually labeled based on graph based semi-supervised learning algorithm (He et al., 2007; Park et al., 2013). The original co-training method begins with training two independent attribute subsets, and whenever the two classifiers reached an agreement

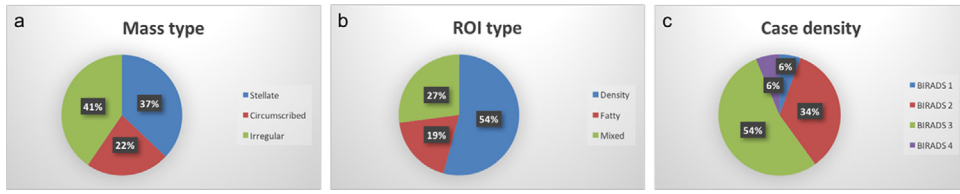


Fig. 1. The distribution of (a) mass types, (b) ROI types and (c) case densities.

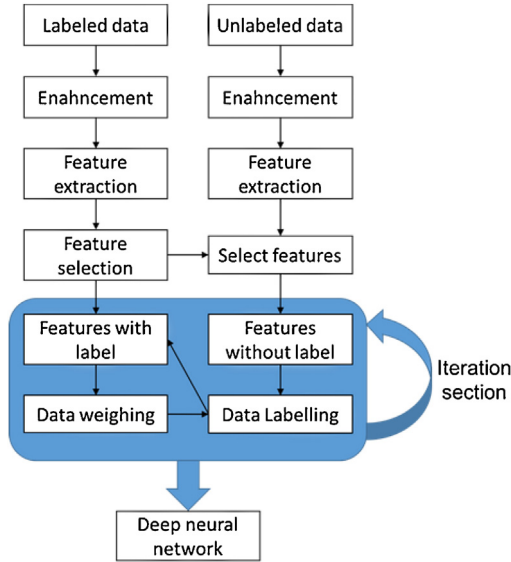


Fig. 2. A brief flowchart of our proposed scheme.

the unlabeled data were being labeled. But in medical imaging area, preparing two sufficient and redundant attribute subsets is difficult. In this study, we divided the labeled data into several equal-sized subsets instead of struggling to create two redundant feature sets extracted from one dataset.

In this study we use L denote labeled dataset and U denote unlabeled dataset, and x_i is the feature vector for one sample and y_i is the corresponding label. We constructed a weighted graph to measure the similarity of each labeled data and unlabeled data, the nodes in Fig. 3 reflect the labeled or unlabeled data, and the connections show the similarities between two samples. The similarity function is shown in Eq. (4), and compared to previous work (He et al.,

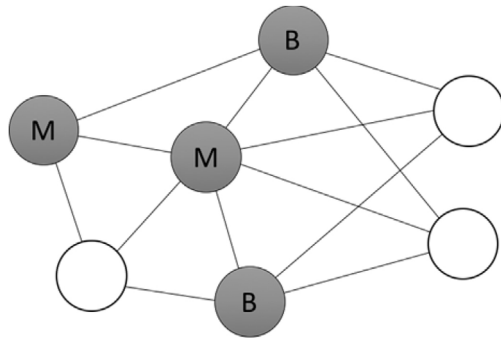


Fig. 3. An example of constructed graph model, the node with letter “B” shows the benign data, the node with letter “M” shows the malignant data, and the blank node means unlabeled data. The connections between the nodes represent the similarity measurements.

2007; Park et al., 2013) the weights were incorporated into the calculations.

$$\begin{cases} s_{ij} = \exp \left(\frac{-(x_i - x_j)^T (x_i - x_j)}{w_i w_j a^2} \right) & \text{if } i \neq j \\ s_{ij} = 0 & \text{if } i = j \end{cases} \quad (4)$$

For the unlabeled data, there is no weight available so we manually assigned 1 as their weights, and k -nearest-neighbors (kNN) was used to locate the adjacent nodes (in this experiment, we set $k=5$). As our scheme is based on the so-called “cluster assumption” algorithm which is summarized by (Chapelle et al., 2003): if there is a path connecting two points that only passes through regions of high density, the possibility of them belonging to the same class is high. The following function (Eq. (5)) first introduced by (Choi et al., 2013) satisfies this assumption, and its solution is shown in Eq. (6).

$$\min_f (f - y)^T (f - y) + \mu f^T L f, \quad (5)$$

$$f = (I + \mu L)^{-1} y \quad (6)$$

where f is the predicted score vector, μ is the tradeoff parameter, I is identity matrix, and L is the graph Laplacian matrix defined as $L = \text{diag}(d_i) - \sum_i s_{ij}$.

2.2.3. Semi-supervised learning procedure

Since the semi-supervised learning is very sensitive on the initial labeled data, we designed a dividing co-training procedure to try its best to make confident labelling. The concept is similar to majority voting, but simply splitting the labeled dataset into many groups result in making each subgroup too small to build a reliable classifier (Sun et al., 2014b). In our proposed procedure, we shuffled and divided the original dataset into two equal-sized subgroups five times, so each subgroup was just half of the original size.

At first, the initial 10 subgroups of labeled data were used to train the 10 classifiers with the graph based data labelling method described above. All these 10 classifiers were used for every unlabeled data according to Eq. (6), and the common confident data were appended to the labeled dataset. The common confident data were chosen by setting a threshold of minimum number of classifiers. Then the whole process kept repeating until no more common confident data appeared, the final labeled dataset was used for classifying the test data. To control the quality of the classifier and the accuracy of newly labeled data for every iteration t , we calculated the prediction accuracy Acc_t^c of the new classifier c_t^c using the combination of initial labeled data and 30% of most confident newly labeled data. Then we compared the same classifier accuracy in the previous iteration $t-1$, if $Acc_t^c < Acc_{t-1}^c$ the expanded labeled dataset was re-shuffled and re-divided until the prediction accuracy improved.

Since our algorithm only added the confident data to the labeled dataset, not all unlabeled data was eventually appended to the labeled dataset. The rest unlabeled data were treated as noisy data

Table 1

The accuracy of newly labeled data using different dimension reduction method and settings.

Method	Dimension			
	3	6	9	12
PCA	79.47%	81.35%	81.74%	81.46%
LDA	74.22%	74.22%	74.22%	74.22%
MDS	79.93%	81.29%	82.43%	82.20%

or atypical data, and they will not be used to build the deep neural network.

2.2.4. Deep convolutional neural network

Deep convolutional neural network has a hierarchical architecture with several levels of composition of non-linear operations in the learned functions. Compared to other traditional CADx systems, it doesn't require us to manually design the hand-crafted features. The architecture of the CNN model we used in this study is shown in Fig. 4. The network begins with sub-patches extracted from the ROIs and the last layer is a fully-connected layer connecting several convolutional layers and subsampling layers (LeCun et al., 1989, 1998). The structure contains three pairs of convolutional layers, with each convolutional layer followed by max-pooling layer. For each convolutional layer, 12, 8 and 6 kernels were applied for the previous feature maps respectively, and the kernel size was set to 5 by 5. The batch size was 100 for the input, subsampling rate was 2, and the learning rate was set to 0.1 for 100 epochs.

2.3. Experimental design

From the 3158 ROIs, we randomly selected 2400 ROIs as training data and the rest 758 ROIs as testing data. Among the 2400 training ROIs, we randomly removed truth files of 2300 ROIs, so 2300 unlabeled data and 100 labeled data were used for this study. Then the proposed data labeling method was applied to these 2400 ROIs, and the majority of the unlabeled data were automatically labeled by our algorithm. The initially labeled data and newly labeled data were used to train the deep neural network. Since the CNN requires its input of the same size, so we processed the original ROIs according to the following criterion: if the small dimension of segmented mass region is less than 52 pixels, it was placed to the center of the 52 by 52 pixel area; otherwise it was downsampled to the size of 52 by 52 pixels. The resized images were rotated to four different directions, and then vertically mirrored and rotated again. After the data augmentation, one image generated eight images after this step. The testing images were resized by the same procedure, but they will not be mirrored and rotated.

3. Results

We first evaluated the influences of three different dimension reduction methods. Table 1 shows the data labeling accuracy using PCA, LDA and MDS. For the comparison reason, we tested the reduced dimensions at 3, 6, 9, and 12. Since LDA only produces at most number of classes -1 feature projections, only 1 feature projection can be produced in this study. That explains the reason LDA has the same accuracy with different dimension settings. The performance of different method was measured by calculating the accuracy of newly labeled data.

Table 2 shows the accuracy of data labelling accuracy with three different data weighing equations (Eq. (1)–(3)). And the results show that exponential function has the highest accuracy, sensitivity and specificity.

We tested the performance of different thresholds of minimum classifiers in majority voting. Using the minimum classifiers from 5

Table 2

Performance comparison using three different data weighing functions.

Weighing function	Exponential function	Gaussian function	Laplacian function
Labeling Accuracy	82.43%	80.79%	79.53%
Sensitivity	81.00%	80.29%	79.18%
Specificity	72.26%	69.35%	70.92%

to 9 as the threshold, we counted the amount of final labeled data, and we used these data for CNN, and calculated their accuracies on testing data.

Fig. 6 shows the performance of CNN, support vector machine (SVM) and artificial neural network (ANN) using different amount of initial labeled data. And we also compared the performance using labeled data only and mixed data. Since using labeled data only cannot provide enough data to properly train the CNN, we excluded the comparison of CNN using labeled data only from Fig. 6.

To compare the performance using labeled data only and mixed data, we uncovered the truth files for all data in our dataset. Then we used all the uncovered data to train the CNN and compared the accuracy using mixed data. The accuracy of using fully unlabeled data is 85.52% compared to 82.43% using same amount of mixed data but with only 100 being labeled. The first layers of learnt features from both datasets are shown in Fig. 7.

4. Discussion

In this study we developed a semi-supervised deep convolutional neural network for breast cancer diagnosis, which can use large amount of unlabeled data to improve the accuracy. This study gives a new way to incorporate deep learning algorithms to medical image studies. Different from other image classification tasks, medical images are much more difficult to collect and label. The method we introduced in this study only requires a small amount of labeled data, and the abundant resources of unlabeled data can be exploited.

To develop a system can use the unlabeled data as efficient as possible, we tested different settings for the proposed scheme, and the highest accuracy we get in this study is 82.43% using 100 labeled data and all unlabeled data. From the comparison of three dimension reduction methods (PCA, LDA and MDS), we noticed that the MDS has the best performance when we set the value of reduced dimension to 9, and generally MDS and PCA have very similar performance. Besides, we also tested three different weighing functions: exponential function, Gaussian function, and Laplacian function. And exponential function generates the best performance in terms of accuracy, sensitivity and also specificity. Based on the comparisons of different amounts of labeled data (Fig. 6), it can be concluded that the amount of labeled data has a significant impact on the performance of the system. The accuracy and AUC grows with the size of labeled data, but the accuracy difference and AUC difference decrease when the amount of labeled data becomes larger. And we noted this trend for all the three tested ANN, SVM and CNN algorithms. Based on these observations we conclude that our proposed scheme is less sensitive to initial labeled data than schemes using the labeled data only, since the unlabeled data provides additional information for the training; also the influence of small dataset will become more significant when the size of labeled data is extremely small (for example 40 in our study), then the ability of incorporating information from unlabeled data would become limited. Because splitting the initial labeled dataset is the essential part of our algorithm, if the size is extremely small the split dataset will be even smaller and makes the scheme more sensitive to noise. Fig. 6 also shows the algorithms performance of using labeled data only and both labeled and unlabeled data, and the AUC

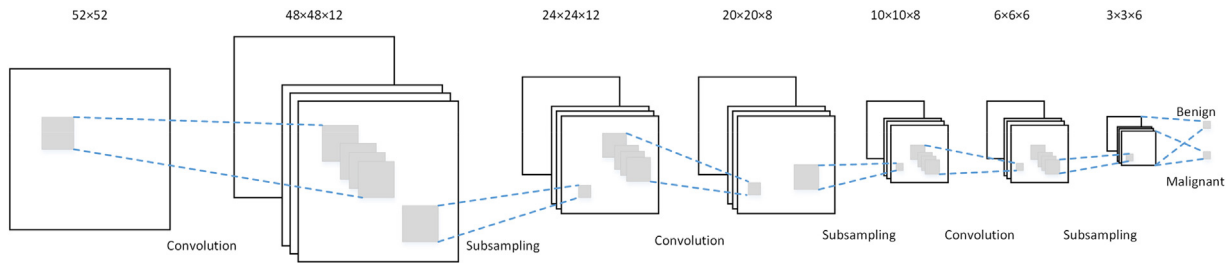


Fig. 4. The structure of CNN we used in this study.

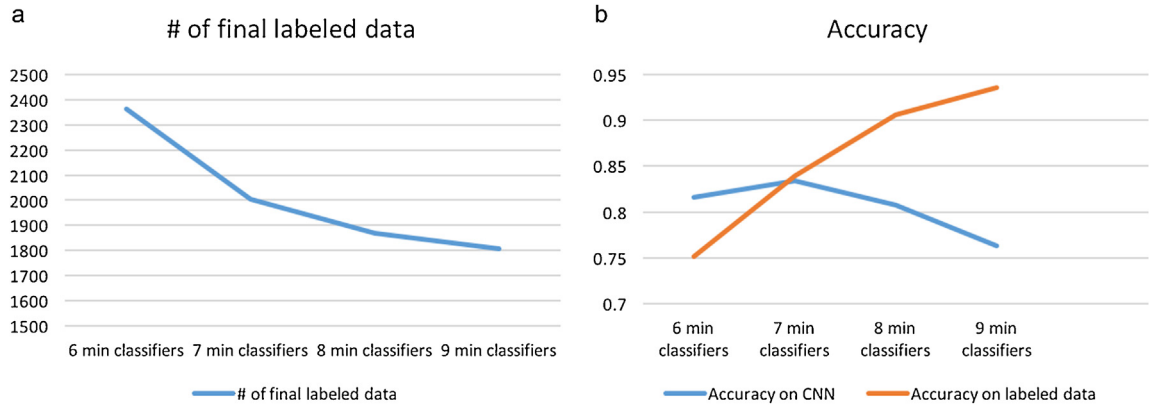


Fig. 5. (a) The number of final labeled data using different minimum classifier threshold. (b) The accuracy of newly labeled data and CNN using different minimum classifier threshold.

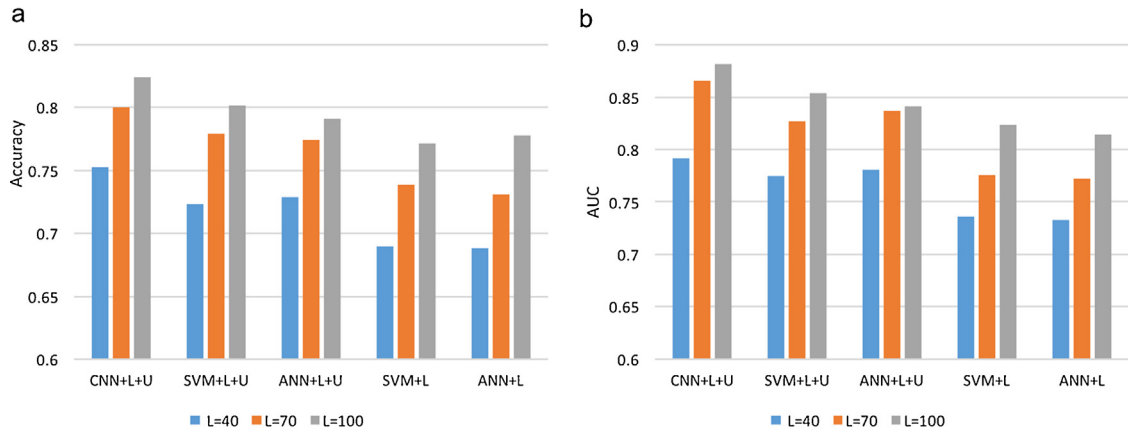


Fig. 6. The (a) accuracy and (b) AUC of CNN, SVM and ANN using labeled data, mixed data, and performance using different number of labeled data.

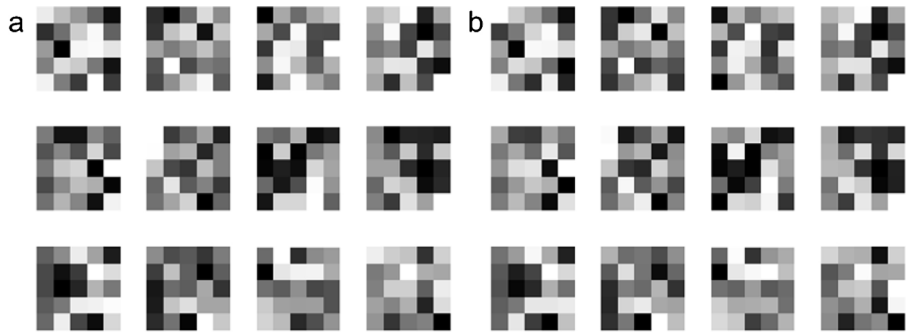


Fig. 7. The automatically generated features by CNN using all (a) mixed data and (b) same amount of all labeled data.

of SVM and ANN increased 3.63% and 3.23%. As for the CNN, if we only use the labeled data, the dataset is too small to train the CNN, so no comparison was based on this data. For the majority voting, we tested different thresholds of minimum classifiers for adding

the unlabeled data into the labeled dataset. As we set the threshold value higher, the accuracy on newly labeled data increased, but the total amount of newly labeled data decreased, which affected the overall CNN performance. And from the test results shown in Fig. 5, we set the threshold to 7.

We also compared the CNN performances using mixed data and same amount of all labeled data. As we expected, the scheme using all labeled data generated 3.75% higher accuracy compared to using a small portion of labeled data and large portion of unlabeled data. Fig. 7 shows the trained features from CNN using mixed data and same amount of unlabeled data, and the feature patterns trained from both datasets are very similar. The unlabeled data could never replace labeled data, and using unlabeled data is only a supplement. Labeled data contains more accurate information from the radiologist compared to computer automatically labeled data, which will definitely help improve the accuracy. However, in some situations, when the labels are not all available, then our proposed scheme can help achieve an acceptable performance for research purpose. In our experiments, the SSL method enables us to use unlabeled data to train the CNN, and it achieved the AUC of 0.8818; if we use the labeled data only, the best AUC is 0.8236 observed by using SVM; using the mixed data for the same SVM scheme, the AUC is 0.8535, which is also higher than using labeled data only.

In summary, this study we developed a new SSL algorithm enables us to use small amount labeled data and a relatively large amount of unlabeled data to train deep convolutional neural network. The results showed that using unlabeled data can dramatically increase the performance of the system. For deep learning algorithms, large dataset is required for the training procedure, which is a great obstacle to apply deep learning algorithms to medical images. Our designed scheme allows the users to include the unlabeled data into the deep learning training dataset and increase the overall accuracy. Future work includes testing our scheme on other deep structured algorithms and optimizing the scheme for each individual deep learning method.

Conflict of interest statement

No conflict of interest.

Acknowledgements

This work was partially supported by National Institutes of Health (SC1CA166016), National Science Foundation (DUE-TUES-1246050) and Department of Education (Award #P031S120131). We also thank the Border Biological Research Center (BBRC) Core Facilities at The University of Texas at El Paso (UTEP) for their support, which were funded by RCMI-NIMHD-NIH grant (8G12MD007592).

References

- Chapelle, O., Weston, J., Schölkopf, B., 2003. Cluster kernels for semi-supervised learning. *Adv. Neural Inf. Process. Syst.* 7 (15), 1.
- Choi, I., Park, K., Shin, H., 2013. Sharpened graph ensemble for semi-supervised learning. *Intell. Data Anal.* 17 (3), 387–398.
- Cireşan, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J., 2013. Mitosis detection in breast cancer histology images with deep neural networks. *Medical Image Computing and Computer-Assisted Intervention-MICCAI*, 411–418.
- Hall, M.A., 1999. Correlation-based feature selection for machine learning. *Methodology* 21i195–i20 (April), 1–5.
- Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. *IEEE Trans. Syst. Man. Cybern.* 3 (6), 610–621.
- He, J., Carbonell, J., Liu, Y., 2007. Graph-based semi-supervised learning as a generative model. *IJCAI International Joint Conference on Artificial Intelligence*, 2492–2497.
- Jamieson, A.R., Giger, M.L., Drukker, K., Pesce, L.L., 2010. Enhancement of breast CADx with unlabeled data. *Med. Phys.* 37 (8), 4155–4172.
- Kallenberg, M., Petersen, K., Nielsen, M., Ng, A., Dia, P., Igel, C., Vachon, C., Holland, K., Karssemeijer, N., Lillholm, M., 2016. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. *IEEE Trans. Med. Imaging* 0062, 1–10, Special issue on deep learning.
- Kiritchenko, S.S., Matwin, S., 2001. Email classification with co-training. *Proceedings of the 2001 Conference of the Centre for Advanced Studies on Collaborative Research*, 8.
- LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D., 1989. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* 1 (4), 541–551.
- LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86 (11), 2278–2323.
- Mitchell, T., Blum, A., 1998. Combining labeled and unlabeled data with co-training. *Proc. Elev. Annu. Conf. Comput. Learn. Theory*, 92–100.
- Park, K., Ali, A., Kim, D., An, Y., Kim, M., Shin, H., 2013. Robust predictive model for evaluating breast cancer survivability. *Eng. Appl. Artif. Intell.* 26 (9), 2194–2205.
- Qian, W., Li, L., Clarke, L.P., 1999. Image feature extraction for mass detection in digital mammography: influence of wavelet analysis. *Med. Phys.* 26, 402.
- Qian, W., Sun, X., Song, D., Clark, R., 2001. Digital mammography: wavelet transform and kalman-filtering neural network in mass segmentation and detection. *Acad. Radiol.* 8 (11), 127–135.
- Qian, W., Song, D., Lei, M., Sankar, R., Eikman, E., 2007. Computer-aided mass detection based on ipsilateral multiview mammograms. *Acad. Radiol.* 14, 530–538.
- Sampat, M.P., Whitman, G.J., Stephens, T.W., Broemeling, L.D., Heger, N.A., Bovik, A.C., Markey, M.K., 2006. The reliability of measuring physical characteristics of spiculated masses on mammography. *Br. J. Radiol.* 79 (2).
- Shiraishi, J., Pesce, L.L., Metz, C.E., Doi, K., 2009. Experimental design and data analysis in receiver operating characteristic studies: lessons learned from reports in radiology from 1997 to 2006. *Radiology* 253 (3), 822–830.
- Smith, R.A., Manassaram-Baptiste, D., Brooks, D., Doroshenko, M., Fedewa, S., Saslow, D., Brawley, O.W., Wender, R., 2015. Cancer screening in the United States, 2015: a review of current American Cancer Society guidelines and current issues in cancer screening. *CA. Cancer J. Clin.* 65 (1), 30–54.
- Sun, W., Tseng, T.-L.(Bill), 2016. A preliminary study on breast cancer risk analysis using deep neural network. *International Workshop on Digital Mammography*, 385–391.
- Sun, W., Zheng, B., Lure, F., Wu, T., Zhang, J., Wang, B.Y., Saltzstein, E.C., Qian, W., 2014a. Prediction of near-term risk of developing breast cancer using computerized features from bilateral mammograms. *Comput. Med. Imaging Graph.* 38 (5), 348–357.
- Sun, W., Tseng, T.-L., Zheng, B., Lure, F., Wu, T., Francia, G., Cabrera, S., Zhang, J., Vélez-Reyes, M., Qian, W., 2014b. Using undiagnosed data to enhance computerized breast cancer analysis with a three stage data labeling method. *SPIE Med. Imaging*, 90350T.
- Sun, W., Tseng, T.-L.B., Zheng, B., Zhang, J., Qian, W., 2015a. A new breast cancer risk analysis approach using features extracted from multiple sub-regions on bilateral mammograms. *SPIE Med. Imaging Int. Soc. Opt. Photonics* 9414, 941422.
- Sun, W., Tseng, T.-L.(Bill), Qian, W., Zhang, J., Saltzstein, E.C., Zheng, B., Lure, F., Yu, H., Zhou, S., 2015b. Using multiscale texture and density features for near-term breast cancer risk analysis. *Med. Phys.* 42 (6), 2853–2862.
- Sun, W., Tseng, T.-L.(Bill), Zheng, B., Zhang, J., Qian, W., 2016a. A novel breast cancer risk assessment scheme design using dual view mammograms. *International Workshop on Digital Mammography*, 392–399.
- Sun, W., Zheng, B., Qian, W., 2016b. Computer aided lung cancer diagnosis with deep learning algorithms. *SPIE Med. Imaging* 9785, 97850Z–97858.
- Sun, W., Huang, X., Tseng, T.-L., Zhang, J., Qian, W., 2016c. Computerized lung cancer malignancy level analysis using 3D texture features. *SPIE Med. Imaging*, 978538.
- Tabar, L., Vitak, B., Tonychen, H.H., Yen, M.F., Duffy, S.W., Smith, R.A., 2001. Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. *Cancer* 91 (9), 1724–1731.
- Tan, M., Pu, J., Zheng, B., 2014. A new and fast image feature selection method for developing an optimal mammographic mass detection scheme. *Med. Phys.* 41 (8), 081906.
- Wan, X., 2009. Co-training for cross-lingual sentiment classification. *Proc. Jt. Conf. 47th Annu. August*, 1–9.
- Wang, X., Li, L., Xu, W., Liu, W., Lederman, D., Zheng, B., 2012. Improving performance of computer-aided detection of masses by incorporating bilateral mammographic density asymmetry: an assessment. *Acad. Radiol.* 19 (3), 303–310.
- Wolfe, J.N., 1976. Breast patterns as an index of risk for developing breast cancer. *Am. J. Roentgenol.* 126 (6), 1130–1139.