

LEARNING TO COMBINE DECISIONS FROM MULTIPLE MAMMOGRAPHY VIEWS

Alan Joseph Bekker[†] Moran Shalhon^{*} Hayit Greenspan^{*} Jacob Goldberger[†]

[†] Faculty of Engineering, Bar-Ilan University, Israel

^{*} BioMedical Engineering, Tel-Aviv University, Israel

ABSTRACT

In this paper we address the problem of differentiating between malignant and benign tumors based on their appearance in the CC and MLO mammography views. We describe a two-step classification method that is based on a view-level decision, implemented by a logistic regression classifier, followed by a stochastic combination of the two view-level indications into a single malignant or benign decision. The EM algorithm is used to find the parameters of the proposed model. Our method was evaluated on a standardized digital database for screening mammography (DDSM). Experimental results demonstrate the effectiveness of optimally combining the decisions based on the two views.

Index Terms— Mammography, Microcalcifications, Computer-aided diagnosis

1. INTRODUCTION

According to the American Cancer Society [1], breast cancer is the second most common cancer among US women, accounting for 29% of newly diagnosed cancers. Following recommended screening guidelines for detecting breast cancer at an early stage can increase both survival rate and treatment options. Mammography is currently the most effective tool for early detection and diagnosis of breast cancer. The presence of clustered microcalcifications (MCs), which are tiny deposits of calcium, is one of the most important and sometimes the only sign of cancer on a mammogram.

To date the characterization of MCs as benign or malignant based on their appearance in mammograms is a difficult task even for expert radiologists, resulting in a high number of unnecessary breast biopsies [2]. The difficulty of classifying benign and malignant MCs by their appearance alone is clearly seen from their visual similarity, as illustrated in the set of example images shown in Fig. 1. Therefore, computer-aided diagnosis (CADx) systems are being developed to support radiologists to better differentiate benign from malignant mammographic lesions (e.g., [3], [4]).

A screening mammographic examination usually consists of four images, corresponding to each breast scanned in two views—mediolateral oblique (MLO) view and craniocaudal (CC) view. The MLO projection is taken in a 45° angle

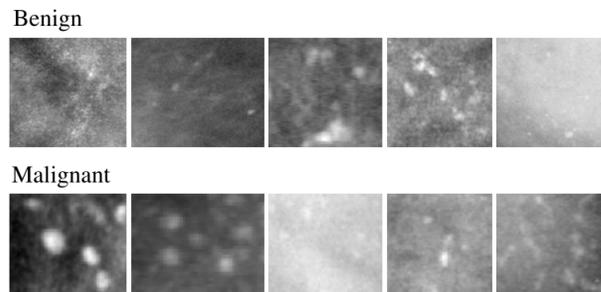


Fig. 1: Examples of benign and malignant MC clusters (from the DDSM dataset).

and shows part of the pectoral muscle. The CC projection is a topdown view of the breast. In reading mammograms, radiologists judge whether or not a lesion is present by comparing both views and breasts. Recent studies, (e.g. [6]) demonstrated the superior performance of a multi-view CAD system over its single-view counterpart.

Although diverse algorithms for mammography computer-aided detection and diagnosis of breast cancer can be found in the literature, it is still an active research field. The performance of current commercial CAD systems still needs to be improved so that they can meet the requirements of clinics and screening centers. The main focus of most previous multi-view methods was on improving the localized detection of breast cancer or on building an extended feature sets using both views. In this work we present a model that automatically learns how to integrate information collected from multiple views into a single decision. We tested our approach on CC and MLO breast images of patients from the digital database for screening mammography (DDSM) [7]. We show the improved performance of the proposed method over several alternative strategies for combining view-level information.

2. A TWO-STEP CLASSIFICATION MODEL

In real life diagnosis, the radiologist considers both CC and MLO views to determine whether the patient has a malignant MC cluster and should undergo a biopsy. In some cases only one of the views is indicative and the radiologist bases the

decision on that view, while in other cases only by analyzing both views together a decision can be made. In our labeled training data we usually do not have direct information whether there is indication of malignant tumors from each view. Instead, we have a biopsy based indication that is correlated with the mammography based diagnosis.

We approach the benign/malignant decision task, based on the two views, as a two step procedure. In the first step separate classifiers are applied on the CC and MLO views to determine whether, based only on the specific view, the cluster is malignant. In the second step, the final decision is determined by integrating the decisions of both views. In this study we try to learn not only how each view can forecast the biopsy result, but also how to combine the two different views into an optimal decision.

We model each view-level binary decision via logistic regression. In a logistic regression model with parameter w , the probability of an input vector x to be labeled as 1 is:

$$p(y = 1|x; w) = \frac{1}{1 + e^{-w^T x}}. \quad (1)$$

Assume that for each patient we have feature vectors x_{cc} and x_{mlo} extracted from the two mammography views, CC and MLO, respectively. Each feature vector is used as an input to a different logistic regression classifier with its own parameter denoted by w_{cc} and w_{mlo} respectively. Let y_{cc} and y_{mlo} be the binary output of the view-level classifiers that provide view-level benign/malignant decisions (we use the convention 0-benign and 1-malignant). Since we only have ground-truth labels based on biopsy, we further assume that we cannot directly observe y_{cc} and y_{mlo} . Instead, we can only observe a stochastic (noisy) combination of them denoted as z ($z = 1$ is a malignant biopsy and $z = 0$ is a benign biopsy). The two-views fusion step is defined by the following parameter:

$$\theta_{ij} = p(z = 1 | y_{cc} = i, y_{mlo} = j)$$

i.e., θ_{ij} is the probability of a malignant biopsy $z = 1$ given view-level decisions $y_{cc} = i$ and $y_{mlo} = j$. According to this model, given features x_{cc} and x_{mlo} extracted from the two views, the probability of obtaining a biopsy based decision z is:

$$p(z = 1 | x_{cc}, x_{mlo}; w_{cc}, w_{mlo}, \theta) = \quad (2)$$

$$\sum_{i=0,1} \sum_{j=0,1} \theta_{ij} \cdot p(y_{cc} = i | x_{cc}; w_{cc}) p(y_{mlo} = j | x_{mlo}; w_{mlo}).$$

The model is illustrated in Figure 2.

The two-step model parameters can be learned from a labeled training data. Assume we are given n (CC, MLO) pairs of feature vectors:

$$(x_{1,cc}, x_{1,mlo}), \dots, (x_{n,cc}, x_{n,mlo})$$

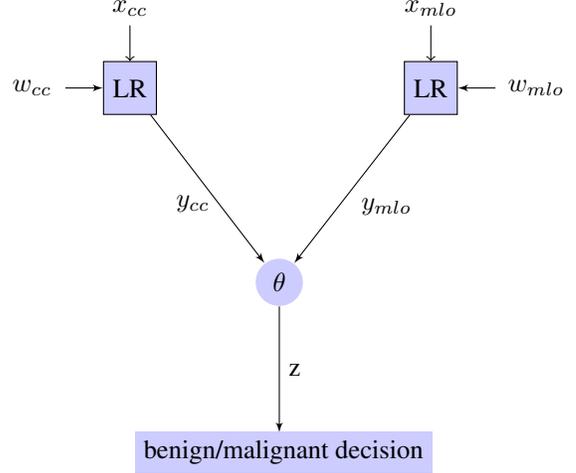


Fig. 2: A diagram of the two-step classification model. In the first step the CC and MLO views are used as input to logistic regression (LR) classifiers that produce binary decisions y_{cc} and y_{mlo} . In the second step a final decision z is obtained as a stochastic combination of the two view-level decisions.

with corresponding binary labels z_1, \dots, z_n obtained from biopsy information. The log-likelihood function of the model parameters is:

$$l(w_{cc}, w_{mlo}, \theta) = \sum_t \log p(z_t | x_{t,cc}, x_{t,mlo}; w_{cc}, w_{mlo}, \theta) \quad (3)$$

The goal of the learning task is to find the view integration parameters $\{\theta_{ij}\}$, and the logistic-regression parameters w_{cc} and w_{mlo} that maximize the likelihood function (3). Since we need to find maximum-likelihood parameters for a model with hidden random variables $\{(y_{t,cc}, y_{t,mlo}) | t = 1, \dots, n\}$, we apply the EM algorithm [8].

The EM algorithm iterates between the E-step that estimates the hidden random variables and the M-step that updates the values of the model parameters. Due to lack of space we do not provide a full derivation of the EM algorithm for our model. Instead, we state the resulting equations. In the E-step we use the current parameter values to compute a soft estimation of the hidden random variables:

$$c_{tij} = p(y_{t,cc} = i, y_{t,mlo} = j | z_t, x_{t,cc}, x_{t,mlo}; w_{cc}, w_{mlo}, \theta) \quad (4)$$

for all $t = 1, \dots, n$, $i = 0, 1$ and $j = 0, 1$. This computation can be easily carried out using Bayes' rule and Eq. (2).

In the M-step we re-estimate the model parameters. The updated values of the parameters θ are:

$$\theta_{ij} = \frac{\sum_{\{t|z_t=1\}} c_{tij}}{\sum_t c_{tij}} \quad i = 0, 1 \quad j = 0, 1 \quad (5)$$

There is no closed-form solution for the updated logistic-regression parameters w_{cc} and w_{mlo} . The M-step update of

the value of w_{cc} is the one that maximizes the following cost function:

$$L(w_{cc}) = \sum_t \sum_{i=0,1} c_{ti} \log p(y_{t,cc} = i | x_{t,cc}; w_{cc}) \quad (6)$$

where

$$p(y_{t,cc} = 1 | x_{t,cc}; w_{cc}) = \frac{1}{1 + e^{-w_{cc}^T x_{t,cc}}} \quad (7)$$

and

$$c_{ti} = p(y_{t,cc} = i | z_t, x_{t,cc}, x_{t,mlo}) = c_{ti0} + c_{ti1} \quad (8)$$

such that c_{ti0} , c_{ti1} were computed in the E-step. c_{t1} is the posterior probability, based on the current parameter values, that there is an indication from the features of the CC image $x_{t,cc}$ that the tumor is malignant.

The cost function $L(w_{cc})$ is actually a soft (weighted) version of the cost function that is maximized while training a standard logistic regression. Unlike standard logistic regression with 0/1 labels, we have here a probabilistic information on the label value (defined in Eq. (8)). Like the standard logistic-regression cost function, $L(w_{cc})$ is a concave function with a single maximum point that can be easily found using a gradient ascent procedure. The derivative of the cost function $L(w_{cc})$ is:

$$\frac{dL(w_{cc})}{dw_{cc}} = \sum_t (c_{t1} - p(y_{t,cc} = 1 | x_{t,cc}; w_{cc})) x_{t,cc}.$$

The parameter updating of w_{mlo} is done in a similar way. This concludes the M-step.

The EM algorithm is notoriously known to get stuck in a local maximum point of the likelihood function. Hence, it is important to choose a meaningful way to initialize the model's parameters. A reasonable initialization is to use the biopsy labels z_1, \dots, z_n to directly train a logistic regression with inputs $x_{1,cc}, \dots, x_{n,cc}$ to get w_{cc} . We can initialize w_{mlo} in a similar way. Since our method is based on combination of EM algorithm and Logistic Regression classifier we denote it EM-LR.

3. EXPERIMENTS

3.1. Data-set and Preprocessing

We based our study on the DDSM dataset [7], which provides the largest number of available annotated mammograms with a biopsy-proven diagnosis. For each patient the dataset includes, the demographic data of the patient, the density of the breast (BI-RADS scale), and information on the boundaries of the annotated regions, where the lesion was detected and the biopsy was done (some cases contain more than one region of interest (ROI)). Almost all cases contain images from both the craniocaudal (CC) view and the mediolateral oblique

(MLO) view. The contours of the lesions are provided by a chain code which we used to extract ROIs. We chose patients in the DDSM dataset that had both CC and MLO views in order to test our model. We separated the mammograms into two different tissue-density categories : fatty tissues (ratings 1 and 2), and dense tissues (ratings 3 and 4) and developed the model for each of them separately. Overall, the dataset we used was comprised of a fatty-tissue dataset, which included 260 pairs of CC, and MLO images, of which 139 were benign and 121 were malignant, and a dense-tissue dataset that included 445 pairs of CC, MLO images, of which 233 were benign and 212 were malignant.

In our mammographic data the feature vectors x_{cc} and x_{mlo} are extracted from the CC and MLO views respectively. The feature vectors used for our study were based on the Curvelet transform [9]. MCs need to be treated the same in different orientations. Therefore, we used rotation invariant texture features. The features were extracted from the Curvelet coefficients at intermediate scales (in our study, two scales), and included for each scale the four features mentioned in [10], with three additional features: entropy, skewness and kurtosis. Overall each extracted ROI was represented by 14 features. We tried other classic texture features (such as GLCM, GLRLM, Wavelets and more) and we chose the Curvelet based features that yielded the best performance.

3.2. Results

We compared the proposed EM-LR method with several baseline methods that use the biopsy given labels to directly train logistic-regression and SVM classifiers. We trained a classifier separately on the MLO features and on the CC features. At the test phase we combined the CC and MLO classifiers results using either OR or AND gates. Another baseline method that directly uses the biopsy labels is training a classifier on a concatenation of the MLO and CC features (denoted by CC+MLO). We also implemented several variants of our two-step approach. In addition to the full model, where θ is learned from the training data, we also examined the variant where in the EM-LR algorithm θ is set to be an OR gate (i.e. $\theta_{ij} = i \vee j$) and the case where θ is set to be an AND gate.

We evaluated the algorithm performance using ROC curves, calculating area under curve (AUC) and using classification accuracy, sensitivity and specificity measures. The results were computed using 10 fold cross validation. Table 1 and Table 2 show classification results for the two types of breast tissue (Fatty/Dense). As can be seen from the classification results, the variants of the proposed EM-LR approach outperformed all base-line methods.

Table 3 shows the maximum-likelihood values of θ for dense category based on the entire set of fatty images. The maximum-likelihood values for the fatty category were similar. Note that the first index of θ corresponds to the CC view, e.g. θ_{01} is the probability of a malignant decision given that

Table 1: Classification results (benign vs. malignant) for fatty breast tissues.

	method	CC+MLO	AND	OR	θ_{ML}
Accuracy	Logistic	69.6	67.3	67.7	-
	SVM	70.0	70.4	71.1	-
	EM-LR	-	70.2	71.0	72.5
AUC	Logistic	0.74	0.73	0.74	-
	SVM	0.76	0.76	0.75	-
	EM-LR	-	0.75	0.76	0.77
Sensitivity	Logistic	69.5	67.2	67.4	-
	SVM	70.2	70.4	70.9	-
	EM-LR	-	70.7	70.1	72.3
Specificity	Logistic	69.7	67.4	69.7	-
	SVM	69.9	70.0	71.2	-
	EM-LR	-	69.6	72.0	72.7

Table 2: Classification results (benign vs. malignant) for dense breast tissues.

	method	CC+MLO	AND	OR	θ_{ML}
Accuracy	Logistic	57.1	58.0	60.9	-
	SVM	63.1	61.2	62.1	-
	EM-LR	-	64.7	64.3	64.7
AUC	Logistic	0.71	0.71	0.72	-
	SVM	0.72	0.68	0.70	-
	EM-LR	-	0.73	0.73	0.73
Sensitivity	Logistic	57.0	58.2	60.8	-
	SVM	63.2	61.1	62.1	-
	EM-LR	-	63.5	64.6	64.8
Specificity	Logistic	58.1	57.5	61.1	-
	SVM	63.0	61.3	62.0	-
	EM-LR	-	64.9	63.7	64.1

the CC view based decision is benign and the MLO based decision is malignant. As can be seen from Table 3, the maximum-likelihood θ is more general than the AND and OR operators. The values $\theta_{00} = 0.054$ and $\theta_{11} = 0.999$ mean that whenever the two view-based decisions agree, the final decision will follow that decision. The parameter value $\theta_{10} = 0.546$ means that if, based on the CC view, the cluster looks malignant and, based on the MLO view, it looks benign then it looks malignant then the probability of a malignant biopsy slightly more than one half. Finally, $\theta_{01} = 0.567$ means that if, based on the CC view, the cluster looks benign and, based on the MLO view, it looks malignant then the probability of a malignant biopsy is also slightly more than one half.

To conclude, in this study we presented a method for automatically combining mammography view-level decisions into a global benign/malignant classification. We showed the improved performance of our approach over several alternative

Table 3: Maximum likelihood parameter values of θ for the dense dataset.

Parameter	CC	MLO	ML value
θ_{00}	0	0	0.054 ± 0.036
θ_{01}	0	1	0.567 ± 0.135
θ_{10}	1	0	0.546 ± 0.116
θ_{11}	1	1	0.999 ± 0.047

ways of combining the view-level information. The work presented here can be extended in several directions. In our experiment we use a specific set of features extracted from the image ROI. However, the proposed method can be applied to any other feature set. In the algorithm presentation we focused on the most common mammography views namely CC and MLO. Our method, however, can be easily extended to mammography with more than two views.

4. REFERENCES

- [1] *Breast Cancer Facts and Figures 2013-2014*, American Cancer Society, 2014.
- [2] D. B. Kopans, "The positive predictive value of mammography," *American Journal of Roentgenology*, vol. 158, no. 3, pp. 521–526, 1992.
- [3] S. Singh, V. Kumar, H. K. Verma, and D. Singh, "SVM based system for classification of microcalcifications in digital mammograms," in *Engineering in Medicine and Biology Society (EMBS), Annual International Conference of the IEEE*, 2006, pp. 4747–4750.
- [4] M. Elter and A. Horsh, "CADx of mammographic masses and clustered microcalcifications: A review," *Med. Phys.*, vol. 36, pp. 2052–2068, 2009.
- [5] M. Samulski and N. Karssemeijer, "Optimizing case-based detection performance in a multi-view CAD system for mammography," *IEEE Trans on Med Imaging*, vol. 30, pp. 1001–1009, 2011.
- [6] M. Heath, K. Bowyer, D. Kopans, R. Moore, and W. P. Kegelmeyer, "The digital database for screening mammography," *Proceedings of the Int. Workshop on Digital Mammography*, pp. 212–218, Medical Physics Publishing, 2001.
- [7] A. P. Dempster, N. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 1–38, 1977.
- [8] E. Candés, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *SIAM Multiscale Model. Simul.*, pp. 861–899, 2006.
- [9] Y. Shang, Y. Diao, and C. Li, "Rotation invariant texture classification algorithm based on curvelet transform and SVM," in *Int. Conf. on Machine Learning and Cybernetics*, 2008, vol. 5, pp. 3032–3036.