Bag of visual words based approach for the classification of benign and malignant masses in mammograms using voting-based feature encoding

Conference Paper · July 2018

DOI: 10.1117/12.2316307

4 authors:

Zobia Suhail
Aberystwyth University

Erika R.E. Denton
Norfolk and Norwich University Hospitals NHS Foundation Trust

Reyer Zwiggelaar
Aberystwyth University

Arif Mahmood
Information Technology University (ITU)

Bag of Visual Words Approach for Classification of Benign and Malignant Masses in Mammograms Using Voting Based Feature Encoding

Zobia Suhail[a], Arif Mahmood[b], Erika R.E. Denton[c], and Reyer Zwiggelaar[a]

[a]Department of Computer Science, Aberystwyth University, Wales, UK.
[b]Department of Computer Science and Engineering, Qatar University, Doha.
[b]Norfolk and Norwich University Hospitals NHS Foundation Trust.
Department of Radiology, Norwich, U.K.

Received: date / Accepted: date

ABSTRACT

Classification of benign and malignant masses in mammograms is a challenging problem. It has wide applications in the development of Computer Aided Diagnosis (CAD) systems, however many challenges still need to be addressed. Due to the risk associated with segmenting the mass region, focus is shifting from selecting the features just from the mass area, to the whole Region of Interest (RoI) containing that mass. Bag of Visual Words (BoVW) techniques are gaining attention for classification tasks in medical imaging by considering RoI as a set of local features. In general BoVW aims to construct a global descriptor based on the extracted local features. In this work, we investigate the performance of BoVW for the classification of benign and malignant mammographic masses. Several features have been explored as the local features and different methods are applied for building the code-book. Subsequently we propose a voting-based approach to encode the features. The proposed approach is evaluated on a subset of DDSM dataset. Initial results reveal classification accuracy as high as 87% and Area Under the Curve (AUC) as 0.93, which are better than the current state-of-the-art approaches applied to the same problem.

Keywords: Classification, Mammogram, Masses, Computer Aided Diagnosis. Bag of Visual Words, Benign, Malignant

Further author information: Send correspondence to Zobia Suhail.
Zobia Suhail: E-mail: zoa1@aber.ac.uk
Arif Mahmood: E-mail: rfmahmood@gmail.com
Erika R.E. Denton: E-mail:erika.denton@nnuh.nhs.uk
Reyer Zwiggelaar: E-mail: rrz@aber.ac.uk

1. INTRODUCTION

Breast cancer is one of the most common types of cancer in women and it is the second major death cause after lung cancer.[1] Early detection is a key to prevention of breast cancer deaths.[2] Mammography is considered to be the most reliable and effective screening method for early breast cancer detection.[3] However, the specificity associated with the classification of benign and malignant masses in mammograms is still low, therefore benign breast biopsies after false positive mammographic assessment is still a major concern. According to a study, breast cancer found in approximately 4 out of 1000 women who has gone through mammographic screening process with Positive Predictive Value (PPV) of 41%.[4] Masses and micro-calcification are two commonly found abnormalities in mammograms. Unlike micro-calcification that appears as tiny spots of calcium, masses appear as more volumetric object on mammography images. Classification of these lesions as either benign/malignant is itself a challenging task and existing work has taken the shape[5,6] of the mass into account or has been based on global features.[7] An alternative to the latter is the BoVW approach.[8,9] In medical imaging, Visual Words (VW) has been used for image classification,[10] automating the process of image annotation[11] and image retrieval.[9] In mammogram analysis BoVW has been used for the classification of breast parenchymal tissue,[12] breast density classification[13] and breast tissue (normal/micro-calcification) classification.[14] In this paper, we explored the BoVW approach for the classification of benign and malignant masses in mammograms by using voting-based encoding of VW. Local Binary Patterns (LBP), Intensity-based statistical feature and Histogram of Oriented Gradients (HOG) has been used as local image features that are then encoded using soft assignment technique to k-nearest code-words. Our experiments demonstrate excellent performance of the proposed algorithm compared to the current state of the art techniques. To the best of our knowledge the proposed technique is novel and has not been used before us for the classification of mammograms.

2. DATASET

A subset of the Image Retrieval in Medical Application (IRMA) project[15] has been used in this study containing mammographic patches. The size of each patch is 128×128 and the images in the dataset are in PNG format. The dataset provides four breast tissue density classes (fatty, fibro-glandular, heterogeneously dense and extremely dense) with three assessment categories (normal, benign and malignant) for each tissue density class. Total number of mammogram patches are 233 per assessment category for each tissue density, that makes 2796 patches in total.

In this work, we performed experiments on 400 mammographic patches belonging to fatty tissue density class (200 for benign and 200 for malignant class). The whole data is split into test and training in ratio of 75%(300) and 25%(100). The sample mammogram patches from the reference database belonging to the fatty class could be seen in Fig. 1.

3. THE PROPOSED BOVW ALGORITHM

The proposed algorithm consists of five steps: feature extraction, features pre-processing, code-book generation, features encoding and feature fusion. The overall architecture of the proposed method can be seen in Figure 2, and the details are given in the following sections.

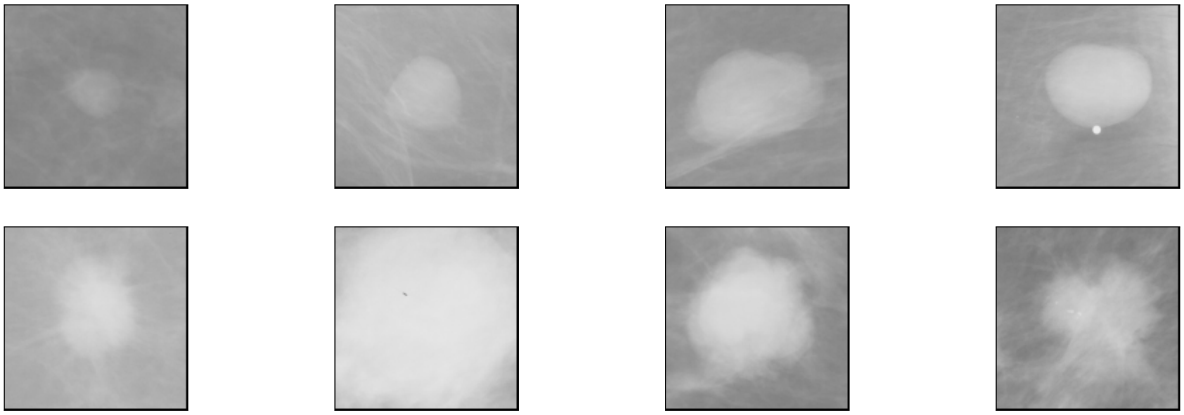

Figure 1: Sample images from the reference database: top row: benign mammogram patches, bottom row: malignant mammogram patches

3.1 Feature Extraction

Local features are extracted from each Region of Interest (RoI) belonging to each of the benign and malignant class for which following local descriptors were used.

- Intensity-Based Statistical Features (IBSF)
- Local Binary Patterns (LBP)
- Histogram of Oriented Gradients (HOG)

The selection of these particular features are according to the initial classification results achieved by using these local representations. The set of intensity-based features used in this work includes Mean (μ), Skewness ($skew$), Kurtosis (kur), Entropy (E) and Root Mean Square (rms). Local Binary Patterns (LBP) is a particular type of textural descriptor used for the image classification.[16] For this particular work, we computed LBP by setting (P,R) = (2,16), whereas the number of bins for the histogram is set to 18. In addition only "Uniform" LBP has been considered in this experiment, because ULBP provides a maximum of the 3×3 texture patterns in the underlying texture surface.[17] HOG in another useful image descriptor commonly used for the purpose of object detection.[18] Like LBP, the HOG parameters need to be tuned for improved results. For this study, we used 32×32 cell size, 4×4 cells per block and 9-bins histogram.

3.2 Feature Preprocessing

It should be noted that all the 3 feature-sets, i.e. IBSF, LBP and HOG are normalized to unit magnitude. In this step, preprocessing is applied on the extracted features in order to convert the low-level local features that are highly correlated to a set of uncorrelated variables that are linearly independent.[19] In addition, whitening technique is used with Principal Component

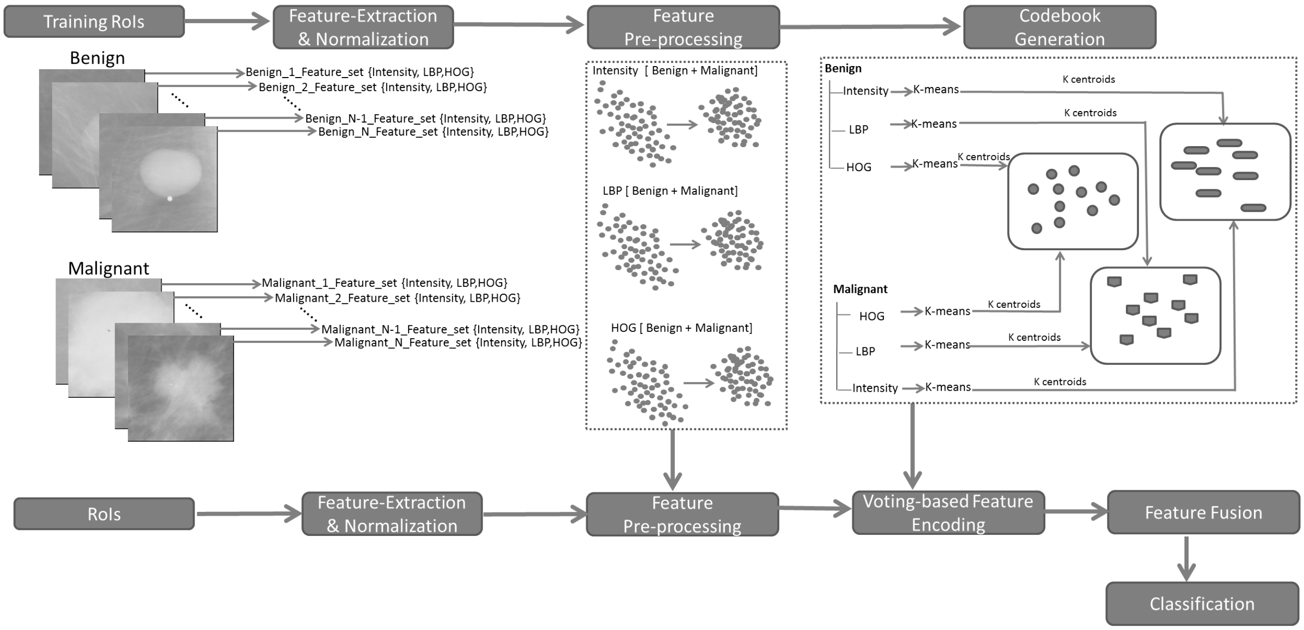

Figure 2: Figure showing the main steps involved in the current method of classifying benign and malignant RoIs. Training phase includes steps of feature extraction, features pre-processing, generating codebook. After generating codebook, test and train RoIs are encoded by using voting-based method to the k-nearest codewords from the codebook.

Analysis (PCA) to ensure that the uncorrelated outputs have almost the same component-wise variances.

The features are extracted from the training data corresponding to both benign and malignant classes. In this step the PCA dimensionality reduction technique is applied to each of the feature-set by combining features for both classes (benign and malignant) corresponding to each feature category (IBSF, LBP, HOG) separately. Subsequently the Principal Components (PCs) for each individual feature-set are computed. In addition, whitening technique is used with PCA to ensure that the uncorrelated outputs have the same component-wise variance. Let $f_j^i \in \mathcal{R}^r$ be a particular type of feature in r dimensional space, $i \in \{IBSF, LBP, HOG\}$ shows feature type, and $1 \leq j \leq m$ is feature number in the training set having m total number of features,

$$\frac{1}{m}\sum_{j=1}^{m}(f_j^i - \mu^i)(f_j^i - \mu^i) = P^i \Lambda^i P^{i^\top}, \tag{1}$$

where μ^i is the mean of a particular type of training features, $P^i = [p_1^i \; p_2^i \; \cdots p_r^i] \in \mathcal{R}^{m\times m}$ is the orthonormal basis matrix containing eigenvectors or Principal Components (PC) and $\Lambda^i \in \mathcal{R}^{r\times r}$ is the diagonal matrix of eigenvalues arranged in decreasing order, $\lambda_j^i \geq \lambda_{j+1}^i$ on the diagonal. The number of PCs are selected $s = 2$ for IBSF, $s = 5$ for LBP and $s = 70$ for HOG features. The selection of PCs are made in a way to capture overall energy $\sum_{j=1}^{s}\lambda_j / \sum_{j=1}^{r}\lambda_j \geq 0.90$. After computing PCs for each of the individual feature-set, the local descriptors (IBSF, LBP, HOG) for each of the RoIs belonging to benign and malignant classes are transformed to the reduced dimensions as computed in Equation 1.

$$q_j^i = \begin{pmatrix} p_1^i/\sqrt{\lambda_1^i} \\ p_2^i/\sqrt{\lambda_2^i} \\ \vdots \\ p_s^i/\sqrt{\lambda_s^i} \end{pmatrix} d_j^i \tag{2}$$

where $q_j^i \in \mathcal{R}^s$ is representing the transformed features, $d_j^i \in \mathcal{R}^r,\ r \geq s$ is the original extracted local feature and $p_1^i,\ p_2^i, \cdots p_s^i$ are the selected PCs for feature type i, and $s \in \{2, 5, 70\}$ in all of our experiments. The benefits achieved by applying this pre-processing step is 2-Fold. Firstly the dimension of the data is reduced that also contributes in minimizing the overall computational complexity. Secondly, the hand-crafted features may contains some noise, this step may refine the features by removing those noise factors.

3.3 Codebook Generation

In this step, we generate a codebook that will be used as a basic model and for generating the global representation of the RoIs. In order to build a codebook each of the reduced feature-space (IBSF, LBP, HOG) for each of the benign and malignant class is partitioned into c_k clusters using k-means clustering algorithm.[19] K-means clustering is used to partition the data points into c_k clusters in such a way that the intra-cluster Euclidean distance of a data point q_j^i to the cluster centroid μ_c^i in minimized, where $j \leq m$ and $c \leq c_k$. The objective function can be defined as:

$$\underset{i}{\mathrm{J}} = \sum_{c=1}^{c_k} \sum_{j=1}^{m} \| q_j^i - \mu_c^i \|_2^2 \tag{3}$$

In order to minimize the objective function defined in Equation 3, each point is assigned to a cluster in a way that it has a minimum distance to that cluster centroid compared to all other clusters. We set value of $c_k = 8$ for this study after performing certain experiments to improve the overall performance. For each of the feature-set (IBSF, LBP, HOG) 8 clusters are computed for each of the benign and malignant classes. Instead of creating a single codebook for all the features, we created separate codebooks for each of the three features. After applying k-means clustering 8 clusters-centroids for IBSF features are combined for both benign and malignant class to form codebook for Intensity-features with 16 items. Likewise the cluster centroids for other two features (LBP and HOG) are also combined for each of the benign and malignant class separately to form the codebook for LBP and HOG features. In this way, the codebook for each feature is composed of 16 code-words (8 benign & 8 malignant). Subsequently, the generated codebooks are used for building global feature representations.

3.4 Features Encoding

In this step, we generate descriptors that will be used for the classification of benign and malignant masses. For the 25% of data (100 images 50 for each benign and malignant class), the same pre-processing is repeated as we did for the codebook generation and as a result we obtain reduced dimensionality feature vectors $q_j^i \in \mathcal{R}^s$. At this stage we used the generated codebooks for encoding these feature separately for a particular feature type, $i \in$ {IBSF, LBP, HOG}. Voting-based encoding method is used here to encode the features in which each of the feature descriptor vote for the code-word for the generation of a global descriptor. Instead of doing hard assignments of the code-words (either 0 or 1), a soft-assignment strategy has been used here, where a normalized weight vector is constructed corresponding to each of the feature q_j^i. Here we used SA-k (Soft-Assignment k) that only counts the weight of the k-nearest code-words for the purpose of encoding.[20] The normalized weight $w_j^i \in \mathcal{R}^{16}$ for feature q_j^i and code-words c_x^i where $1 \leq x \leq 16$ is the code-word number in $i-$th codebook:

$$w_{j,x}^i = \frac{exp(-\beta \left\| q_j^i - c_x^i \right\|_2^2)}{\sum_{x=1}^{16} exp(-\beta \left\| q_j^i - c_x^i \right\|_2^2)}, \tag{4}$$

where β is used as a smoothing factor that controls the softness of the assignment. We set the value of this smoothing parameter β equal to 1.00 based on the experiments to achieve good results. The normalized weights of all the codewords in the codebook are computed for a particular feature q_j^i, $1 \leq j \leq t$, and then only the k-highest weights $w_{j,x}^i$ are retained as encoded feature and rest of the entries are made zero in the weight vector w_j^i. The final encoded feature w^i is then computed by summation over t normalized weight vectors computed from each RoI

$$w_i = \sum_{j=1}^{t} w_j^i \in \mathcal{R}^{16},\ i \in \{\text{IBSF, LBP, HOG}\}. \tag{5}$$

At the end of this step, we end up with three encoded features for each RoI. In the next step the process to build a global feature descriptor is defined.

3.5 Feature Fusion

Feature fusion is effectively a method of combining multiple local features that result in a global representation of the RoI.[21] Here we used descriptor level feature fusion, in which the encoded features for each RoI is concatenated as a global descriptor that is then used for the classification of benign and malignant mass.

$$v_{ROI} = [w_{IBSF}^\top \mid w_{LBP}^\top \mid w_{HOG}^\top]^\top \in \mathcal{R}^{48}, \tag{6}$$

The resultant fused vectors v_{ROI} for the training RoI are used to train kernel SVM that are then used for testing.

3.6 Experiments and Results

Large numbers of experiments are conducted to evaluate the performance of the proposed algorithm for the task of benign and malignant classification. The encoded features of 100 RoIs from the data (50 benign and 50 malignant) are used to test the algorithm's performance. No instance from the dataset that are used for the model testing have been used in the generation of the codebook. In addition the principal components (in pre-processing) have been found for the

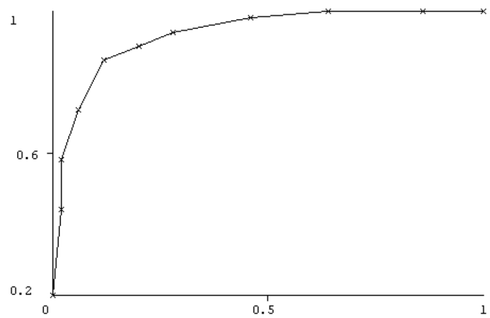

Figure 3: ROC curve for the proposed method

	Benign	Malignant
Benign	44	6
Malignant	7	43

Table 1: Confusion Matrix.

training data. Subsequently, the same PCs has been used to reduce the dimensionality of the whole dataset. 10-Fold Cross validation (FCV) scheme has been used to access the performance of the data. The overall Classification Accuracy (CA) achieved is 87% with Area Under the ROC curve(A_z) equal to 0.93. The ROC curve for the current method can be seen in Figure 3, that is plotted against False Positive Rate (FPR) to the True Positive Rate (TPR), whereas the confusion matrix for the total 100 instances can be seen from Table.1.

4. FUTURE WORK AND CONCLUSION

In this work we explored BoVW approach for the classification of benign and malignant masses using features selected from whole RoI as global descriptors. In future we will try extracting features from small patches of the RoI as local descriptors to improve overall classification accuracy. The dataset used in the experiments belong to fatty breast tissues. The work will be extended to include all breast tissue densities including fatty, fibro-glandular, heterogeneously dense and extremely dense tissues. In conclusion, the presented work provides good classification results with an accuracy of 87% and A_z=0.93, which is in line with the current state of the art methods.

REFERENCES

[1] J. Ferlay, A. Héry, P. Autier, *et al.*, “Global burden of breast cancer,” in *Breast cancer epidemiology*, 1–19, Springer (2010).

[2] L. Humphrey, M. Helfand, B. Chan, *et al.*, “Breast cancer screening: a summary of the evidence for the us preventive services task force,” *Annals of internal medicine* **137**(5-1), 347–360 (2002).

[3] J. Khatcheressian, A. Wolff, T. Smith, *et al.*, “American society of clinical oncology 2006 update of the breast cancer follow-up and management guidelines in the adjuvant setting,” **24**(31), 5091–5097, American Society of Clinical Oncology (2006).

[4] M. Morton, D. Whaley, K. Brandt, *et al.*, “Screening mammograms: interpretation with computer-aided detectionprospective evaluation,” *Radiology* **239**(2), 375–383 (2006).

[5] P. Valarmathie, V. Sivakrithika, and K. Dinakaran, “Classification of mammogram masses using selected texture, shape and margin features with multilayer perceptron classifier,” *Biomedical Research* (2016).

[6] G. Ertas, H. Gulcur, E. Aribal, *et al.*, "Feature extraction from mammographic mass shapes and development of a mammogram database," in *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, **3**, 2752–2755, IEEE (2001).

[7] Y. Li, H. Chen, G. Rohde, *et al.*, "Texton analysis for mass classification in mammograms," *Pattern Recognition Letters* **52**, 87–93 (2015).

[8] S. Selvi and C. Kavitha, "Radiographic medical image retrieval system for both organ and pathology level using bag of visual words," *Int J Eng Sci Emerg Technol* **6**(4), 410–416 (2014).

[9] J. Wang, Y. Li, Y. Zhang, *et al.*, "Boosted learning of visual word weighting factors for bag-of-features based medical image retrieval," in *Image and Graphics (ICIG), 2011 Sixth International Conference on*, 1035–1040, IEEE (2011).

[10] M. Awedh, "Medical image classification using multivocabulary," *Asian Journal of Applied Sciences* **8**(1), 71–78 (2015).

[11] R. Bouslimi, A. Messaoudi, and J. Akaichi, "Using a bag of words for automatic medical image annotation with a latent semantic," *arXiv preprint arXiv:1306.0178* (2013).

[12] J. Wang, Y. Li, Y. Zhang, *et al.*, "Bag-of-features based classification of breast parenchymal tissue in the mammogram via jointly selecting and weighting visual words," in *Image and Graphics (ICIG), 2011 Sixth International Conference on*, 622–627, IEEE (2011).

[13] C. Hiba, Z. Hamid, and A. Omar, "An improved breast tissue density classification framework using bag of features model," in *Information Science and Technology (CiSt), 2016 4th IEEE International Colloquium on*, 405–409, IEEE (2016).

[14] I. Diamant, H. Greenspan, and J. Goldberger, "Breast tissue classification in mammograms using visual words," in *Electrical & Electronics Engineers in Israel (IEEEI), 2012 IEEE 27th Convention of*, 1–4, IEEE (2012).

[15] T. Deserno, M. Soiron, J. Oliveira, *et al.*, "Towards computer-aided diagnostics of screening mammography using content-based image retrieval," in *Graphics, Patterns and Images (Sibgrapi), 2011 24th SIBGRAPI Conference on*, 211–219, IEEE (2011).

[16] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," in *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, **1**, 582–585, IEEE (1994).

[17] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence* **24**(7), 971–987 (2002).

[18] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, **1**, 886–893, IEEE (2005).

[19] C. Bishop, "Pattern recognition," *Machine Learning* **128**, 1–58 (2006).

[20] J. V. Gemert, C. Veenman, A. Smeulders, *et al.*, "Visual word ambiguity," *IEEE transactions on pattern analysis and machine intelligence* **32**(7), 1271–1283 (2010).

[21] X. Peng, L. Wang, X. Wang, *et al.*, "Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice," *Computer Vision and Image Understanding* **150**, 109–125 (2016).