

HYBRID TECHNIQUE BASED ON N-GRAM AND NEURAL NETWORKS FOR CLASSIFICATION OF MAMMOGRAPHIC IMAGES

Pradnya Kulkarni¹, Andrew Stranieri¹, Siddhivinayak Kulkarni¹, Julien Ugon¹ and Manish Mittal²

¹Centre for Informatics and Applied Optimisation, Federation University
PO Box 663, Ballarat 3353, Victoria

p.kulkarni, a.stranieri, s.kulkarni, j.ugon@federation.edu.au

²Lakeimaging, Ballarat, Victoria
mittalm@lakeimaging.com.au

ABSTRACT

Various texture, shape, boundary features have been used previously to classify regions of interest in radiological mammograms into normal and abnormal categories. Although, bag-of-words or n-gram model has been effective in text representation for classification or retrieval of text, these approaches have not been widely explored for medical image processing. Our purpose is to represent regions of interest using an n-gram model, then deploy the n-gram features into a back-propagation trained neural network for classifying regions of interest into normal and abnormal categories. Experiments on the benchmark miniMIAS database show that the n-gram features can be effectively used for classification of mammograms into normal and abnormal categories in this way. Very promising results were obtained on fatty background tissue with 83.33% classification accuracy.

KEYWORDS

N gram, Bag of Phrases, Neural Network, Mammograms, Image Processing

1. INTRODUCTION

Breast cancer is one of the major causes of death. Early detection is possible with the help of mammography [1]. There are vast amount of mammographic images generated every day. Detection of abnormality is a difficult and demanding job for the radiologists as the features of the abnormality are obscured or can be similar to those in a normal breast. There are always inter-observer variations among the radiologists due to difference in their experience level. There are also intra-observer variations in reading mammograms caused by fatigue and stress. Although a computer program may never be able to achieve the level of knowledge and cognitive capability of a radiologist, it is better at doing certain tasks repetitively and consistently. Computer aided detection systems can thus be used as a second opinion for the radiologists to improve their interpretation accuracy.

Computer aided detection systems usually use segmentation techniques to locate the region of interest, get the features and use the features to discover if the region is normal or abnormal. Different texture features such as histograms, co-occurrence matrix, gabor filters and wavelet transforms have been previously used for classification of regions of interest [2][3][4]. Classification of regions of interest is also useful for image retrieval task where a radiologist looks for similar past cases.

In the field of information retrieval, the bag-of-words concept was originally used in text retrieval systems. The document can be represented with the histograms of the words present in it. Forming phrases by grouping n sequential words, we get a bag-of-phrases model of a document. Thus n -gram is composed of n sequential words. The use of n -grams is already proven to be very efficient in the natural language processing [5]. The same concept can be applied to the image retrieval or classification problem. An image can be represented with the help of n -grams or bag-of-visual-phrases model. Pedrosa [6] shows that the n -gram representation can improve the retrieval performance over bag-of-words model for medical as well as non-medical image datasets.

The purpose of this paper is to extract n -gram features for the regions of interest in the miniMIAS dataset and use these features along with a back-propagation neural network to classify the regions into normal and abnormal categories. The paper is organised as follows. Section 2 details the literature review, research methodology is described in Section 3, Section 4 demonstrates experimental results and analysis of these results by varying n , effect of increasing grey levels and effect of background tissue on classification accuracy and Section 5 concludes the paper.

2. LITERATURE REVIEW

For detection of mammographic masses, the methods reported to date mainly utilize morphological features to distinguish a mass from the normal mammographic background [7]. Linear discriminant analysis is used in [8], where 76% sensitivity and 64% specificity was achieved on the mammographic dataset acquired from Michigan hospitals. Spatial grey level dependence matrix (SGLD) features along with artificial neural network were used to classify the regions of interest from miniMIAS dataset into normal and abnormal categories [9]. In this case 119 regions were used for the training (jack and knife method) and the remaining 119 regions were used for testing purpose. Radial basis function neural network (RBFNN) provided classification accuracy of 78.15% whereas the Multilayer perceptron neural network (MPNN) provided classification accuracy of 82.35%. In [10] Wong tried to classify mammographic regions of interests into mass or non-mass using texture features and artificial neural network. Four significant Grey level co-occurrence matrix (GLCM) features (correlation (0), angular second moment (0), inverse difference moment (0) and correlation (45)) were used on the miniMIAS dataset. Using leave-one-out resampling on 50 ROIs, classification accuracy of 86% was achieved. An effort to find the most distinguishing textural features between normal and abnormal ROIs for miniMIAS dataset can be seen in work [11]. Using the multivariate t test they found that the ASM, Correlation, Sum_Var and Diff_Entropy have the discriminating power for GLCM of distance 1, whereas Angular second moment is the only discriminating feature for GLCM of distance 3.

Nithya experimented on a DDSM database of mammography to classify the ROIs into normal and cancerous categories using artificial neural network [12]. Their method used textural features (Correlation, Energy, Entropy, Homogeneity and sum of Square variance) to train a neural network with 200 mammograms (100 normal, 100 cancer) and tested this with 50 mammograms (25 normal, 25 cancer). Mass classification using gabor filter features and SVM for classification is reported in [13].

A bag of words model used effectively in text retrieval is further extended as Bag-of-visual-words for image classification and retrieval [14][15][16][17][18][19][20]. Researchers have taken various approaches for representing an image with the bag-of-words model. It has been used to analyse mammographic images in [21]. In this method using histogram intersection and SVM a classification accuracy of 73.3% was achieved. Wang reports that this bag-of-visual-words approach is successfully applied for classifying the breast images into BI_RADS categories [21]. This approach has also shown to be useful in lung image classification [23]. The bag of words model has proven to be useful for classification (88% classification accuracy) of normal versus microcalcifications in mammographic images [24]. This approach is also been shown useful in the classification of focal liver lesions on CT images in [25]. In [26] raw intensities without normalization are used as local patch descriptors. The raw patches are then sampled densely with the stride of one pixel in the liver lesion region to form the BoW representation.

Extension of a bag-of-visual-words model is a bag-of-visual-phrase model also called n-gram model. This model has been investigated on various medical and non-medical datasets [6]. They used scale invariant Fourier transform (SIFT) features by identifying keypoints and then represented the image as a Bag of Visual Phrases model. This work shows that the bag-of-visual-phrase model is more efficient in representing image semantics and improves the retrieval performance.

3. RESEARCH METHODOLOGY

This section describes various stages used for extracting n-gram features and classifying mammographic images into two classes: a) Normal and b) Abnormal. Figure 1 shows each component of the proposed technique.

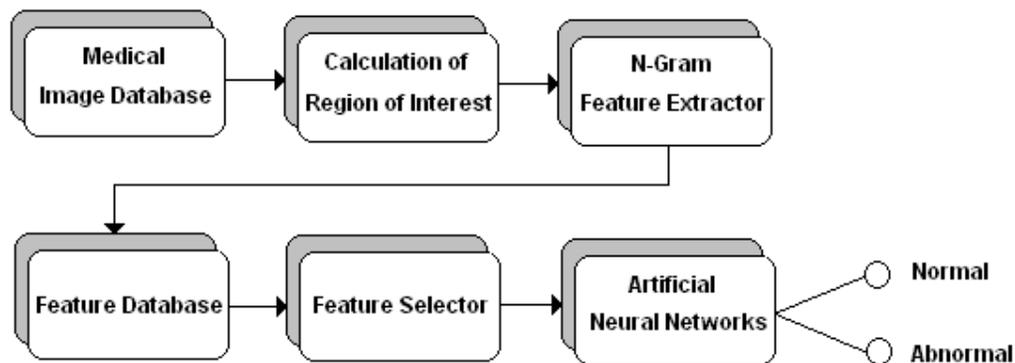


Figure 1. Block diagram of the proposed technique of n-gram feature extractor and ANN classifier

3.1 Calculation of Region of Interest

We are using benchmark miniMIAS database [27]. In this database abnormality is specified with the help of x and y coordinates of the centre of abnormality and a radius of circle enclosing the abnormality. Normal mammograms are also present. Equal size regions of 140×140 pixels around the abnormality are extracted. Figure 2 shows an example image with extraction of region of interest from a mammogram. For normal mammograms regions of 140×140 pixels are cropped from the centre of the mammogram.

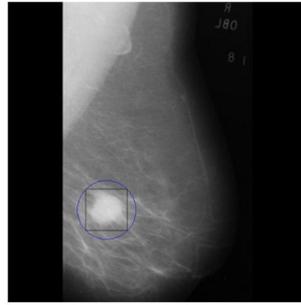


Figure 2: Example ROI extraction containing a circumscribed mass.

3.2 Grey scale reduction

The images are in portable grey map format. Each pixel is represented with a grey level between 0 to 256. For reducing computational complexity, the images are reduced to G number of grey level bins. To reduce the images to G number of grey level bins, the i^{th} bin is composed of grey levels illustrated by equation (1).

$$\text{Bin } i = (i - 1) * \text{binsize to } (i * \text{binsize}) - 1 \quad (1)$$

Where

$$\text{binsize} = \frac{\text{highest grey level} - \text{lowest grey level}}{\text{number of bins}} \quad (2)$$

3.3 N gram feature extractor

In text retrieval context, Word n-grams are a phrases formed by a sequence of n-consecutive words. E.g. 1-gram representation is composed of words such as {mass, network, reporting}. On the other hand, 2-gram is represented by sequence of two words for example {benign mass, neural network, structured reporting}. Phrases formed by n consecutive letters provide more complete representation of the document thereby increasing the retrieval performance. Character n-grams are phrases formed by n consecutive characters. For example, the 3-grams in the phrase “the fox” are “the, he_, e_f, _fo, fox”; the four grams are “the_f, he_f, e_fo, _fox”. Algorithms for generating and counting n-grams are fast and efficient.

For our experiments, we consider grey scale intensity of a pixel is equivalent to one visual character. We can then represent the image in terms of sequence of n-consecutive intensity levels. We call this representation visual character n-gram model. One gram features would thus be similar to the histograms of each grey levels present in an image. Two gram features would consider every possible combination of the two grey levels adjacent to each other. This would closely relate to co-occurrence matrix features with an adjacency distance of 1. We restrict experiments in this study to three and four gram features.

The number of possible phrases is dependent on the grey level bins used to represent the image and can be calculated with the following equation.

$$\text{num of possible phrases in } n \text{ gram} = (\text{num of greylevel bins})^n \quad (3)$$

Thus with the images reduced to 8 grey level bins we have 8 one gram features, 64 two gram features, 512 three gram features and 4096 four gram features for every region of interest

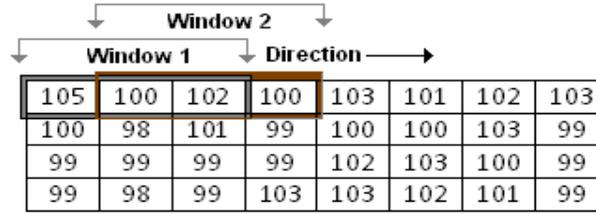


Figure 1: Sliding window for calculating three gram features

We then use sliding window of size n so that we are looking at phrase formed by n consecutive visual characters at a time. The count of how many times the visual phrase is repeated in an image is stored as its n gram features. Figure 3 shows how a sliding window of size 3 pixels is used to calculate 3 gram features of an image.

3.4 Feature Normalization and Feature Database

After calculating the n -gram features we normalize the features by dividing the count with the largest count in the available features so that feature values range between zero to one. The normalized features are stored in the feature database. For improving the classification accuracy, we selected the features which clearly distinguish between normal and abnormal ROIs.

3.5 Artificial Neural Network Classifier

Artificial neural networks have been used effectively for classification of images. Multilayer perceptron and radial basis function network have been used for classification of mammograms [9]. Different neural network algorithms are characterised by learning method and architecture of the network. The supervised learning neural network is efficient to learn the extracted n -gram features from normal and abnormal images. Supervised neural network is used for classifying the images into two categories (normal and abnormal). The input neurons are varied according to the number of features calculated for n -grams.

The normalized features are given as input to the neural network. The corresponding outputs were set to 1 depending upon normal or abnormal ROI features present at the input. Classification accuracy is calculated with the following equations, where n_{ab} is the number of correctly classified abnormal ROIs, n_n is the number of correctly classified normal ROIs, N_{ab} is the total number of abnormal ROIs present and N_n is the total number of normal ROIs.

$$\text{Abnormal Classification Accuracy} = n_{ab}/N_{ab} \quad (4)$$

$$\text{Normal Classification Accuracy} = n_n/N_n \quad (5)$$

$$\text{Overall Classification Accuracy} = \frac{(n_{ab}+n_n)}{N_{ab}+N_n} \quad (6)$$

4. EXPERIMENTAL RESULTS AND ANALYSIS

This section details various experiments and analysis of the results conducted on Mini-MIAS benchmark medical image database.

4.1 Mini-MIAS Database and Image Pre-processing

The Mammographic Image Analysis Society, UK has provided Mini- MIAS as benchmark medical image database for research purposes [27]. The mammograms have been reduced to 200 micron pixel edge and clipped or padded so that every image is 1024×1024 pixels. Abnormality area or region of interest (ROI) is specified with the help of x, y image coordinates of centre of abnormality and the radius of a circle enclosing the abnormality. Information about the type of abnormality such as calcification, mass (circumscribed, speculated or ill-defined), architectural distortion or asymmetry is specified in the database. Type of Background tissue such as Fatty, Fatty-glandular, Dense-glandular has also been mentioned for each mammogram. Images of normal mammograms are also present in the database. The images are in portable grey map format which can be directly read using matlab software.

For our experiment we cropped a square of size 140×140 pixels around the centre of the abnormality. For normal mammograms a square of size 140×140 pixels was cropped from the centre of the mammogram. Example abnormal and normal ROIs are shown in Figure 4.

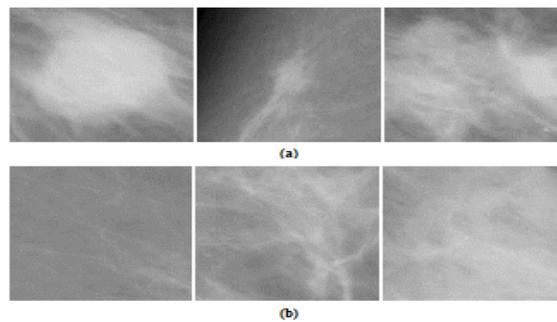


Figure 4 (a): Abnormal ROIs and (b): Normal ROIs

4.2 N gram feature extraction

The regions of interests were reduced in grey scale levels for reducing computational complexity. The n-gram features are then calculated from the grey scale reduced regions of interests. However, all possible phrases may not be present in a corpus and hence the vocabulary of visual phrases reduces significantly reducing the computational complexity of the system. Table 1 shows the number of possible visual phrases and the actual phrases present in the corpus with different grey level bins used. The features which are helpful in clearly distinguishing between normal and abnormal ROIs are selected.

Table 1: Number of possible and actual phrases for 3 gram representation

Grey levels	Number of possible visual phrases	Number of visual phrases present in the corpus
8	512	35
16	4096	82
24	13824	137

4.3 Classification using artificial neural network

The selected n-gram features are used as inputs for the neural network. The output of the neural network then shows the classification of regions into normal and abnormal categories. Various experiments are carried out to see the effect of changing n, background tissues and grey scale levels. Learning rate, momentum and number of iterations are varied to get the maximum output accuracy. The optimum results were obtained with the following settings for six input neurons and 2 output neurons as shown in Table 2.

Table 2: Neural network parameters

Neural Network Parameters	Values
Learning Rate	0.3
Momentum	0.7
Iterations	3000

4.4 Effect of number of words in n grams on classification accuracy

To see the effect of increasing the number of words in a visual phrase, we used 102 regions of interests with fatty background tissue. We used 3 fold cross validation technique to test the classification accuracy. Out of 102, 69 images were used for training and 33 images were used for testing purposes in each fold. To analyse the effect of increasing the grey level bins, three gram features were used. The following table shows that classification accuracy did not increase significantly with the increase in the number of grey level bins used.

Table 1: Effect of increase in number of grey level bins on classification accuracy

Number of grey level bins	Classification accuracy		
	Abnormal	Normal	Overall
8	76.19%	66.66%	69.56%
16	76.19%	68.75%	71.01%
24	76.19%	64.58%	68.11%

Hence we reduce the ROIs to 8 grey level bins. Table 4 shows the effect of increasing n on the classification accuracy.

Table 4: Effect of number of words in n-gram on classification accuracy

Number of consecutive visual words in a phrase	Number of possible phrases	Classification accuracy		
		Abnormal	Normal	Overall
Onegram	8	76.19%	64.58%	68.11%
Twoqram	64	76.19%	64.58%	68.11%
Threegram	512	80.95%	66.66%	73.81%
Fourgram	4096	85.71%	70.83%	78.27%

4.5 Effect of background tissue change on classification accuracy

There are three types of background tissues in mammograms. To analyse the effect of background tissue on classification accuracy, we used 33 training pairs and 12 testing pairs with 3 fold cross validation technique. We found that there are six pure sequence three gram features which distinguish between normal and abnormal regions quite effectively. These six pure sequence three gram features are used as inputs to the neural network. Table 5 shows the results.

Table 5: Effect of background tissue on classification accuracy

Background tissue	Classification accuracy		
	Abnormal	Normal	Overall
Fatty	83.33%	83.33%	83.33%
Fatty Glandular	50.00%	61.90%	58.62%
Dense Glandular	50.00%	50.00%	50.00%

5. CONCLUSION AND FUTURE RESEARCH

A new feature extraction technique called visual character n-gram is proposed for classification of mammograms into normal and abnormal categories. The regions of interests were reduced in grey scale levels to reduce computations complexity. A back-propagation neural network was used for classification purpose. Three fold cross validation technique was used to test the classification accuracy. Number of experiments was conducted for extracting n-gram features and classification of these features using neural networks. Very promising results were obtained on fatty background tissue with 83.33% classification accuracy on Mini-MIAS benchmark database. Although, direct comparisons with other neural network and co-occurrence matrix approaches reported in the literature are not immediately possible because each study deploys different image sets and ROI, however the classification accuracies obtained here are promising and warrant further examination. As the number of visual words in a phrase increases, significant improvement in overall classification accuracy is observed. Results are analysed by changing various parameters. Future research will incorporate investigating distinguishing n-gram features for improving classification accuracy for fatty glandular and dense glandular background tissue.

We would like to use n-gram text features of the associated radiology reports and investigate effect of fusion of n-gram image features and n-gram text features to improve classification accuracy.

REFERENCES

- [1] Bovis, K., Singh, S., Fieldsend, J., & Pinder, C. (2000) "Identification of masses in digital mammograms with MLP and RBF nets." *Proceedings of the IEEE-INNS-ENNS International Joint Conference (IJCNN)*, Vol 1, pp. 342-347.
- [2] Müller, H., Michoux, N., Bandon, D., & Geissbuhler, A. (2004) "A review of content-based image retrieval systems in medical applications—clinical benefits and future directions." *International Journal of Medical Informatics*, Vol. 73, No. 1, pp. 1-23.
- [3] Cheng, H. D., Shi, X. J., Min, R., Hu, L. M., Cai, X. P., & Du, H. N. (2006) "Approaches for automated detection and classification of masses in mammograms." *Pattern Recognition*, Vol. 39, No. 4, pp. 646-668.
- [4] Akgül, C. B., Rubin, D. L., Napel, S., Beaulieu, C. F., Greenspan, H., & Acar, B. (2011) "Content-based image retrieval in radiology: current status and future directions." *Journal of Digital Imaging*, Vol. 24, No. 2, pp. 208-222.
- [5] Suen, C. (1979) "n-gram statistics for natural language understanding and text processing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 1, No. 2, pp. 164–172.
- [6] Pedrosa, G. V., & Traina, A. J. (2013) "From Bag-of-Visual-Words to Bag-of-Visual-Phrases using n-Grams." In *26th Graphics, Patterns and Images (SIBGRAPI)*, pp. 304-311.
- [7] Chan, H. P., Wei, D., Helvie, M. A., Sahiner, B., Adler, D. D., Goodsitt, M. M., & Petrick, N. (1995) "Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in texture feature space." *Physics in medicine and biology*, Vol. 40, No. 5, pp. 857-876.
- [8] Petrosian, A., Chan, H., Helvie, M., Goodsitt, M., Adler, D. (1994) "Computer-aided diagnosis in mammography: classification of mass and normal tissue by texture analysis." *Physics in Medicine and Biology*, Vol. 39, No. 12, pp. 2273–2288.
- [9] Christoyianni, I., Dermatas, E., & Kokkinakis, G. (1999) "Neural classification of abnormal tissue in digital mammography using statistical features of the texture." In *Electronics, Circuits and Systems, Proceedings of ICECS'99*, Vol 1, pp. 117-120.
- [10] Wong, M., Nguyen, H., & Heh, W. (2006) "Mass Classification in Digitized Mammograms Using Texture Features and Artificial Neural Network." *Journal of Pattern Recognition*, Vol. 39, No. 4, pp. 646-668.
- [11] Wei, C. H., Li, C. T., & Wilson, R. (2005) "A general framework for content-based medical image retrieval with its application to mammograms." In *Proceedings of SPIE*, Vol. 5748, pp. 135-142.
- [12] Nithya, R., & Santhi, B. (2011) "Classification of normal and abnormal patterns in digital mammograms for diagnosis of breast cancer." *International Journal of Computer Applications*, Vol. 28, No. 6, pp. 21-25.
- [13] Hussain, M., Khan, S., Muhammad, G., Berbar, M., & Bebis, G. (2012) "Mass detection in digital mammograms using gabor filter bank." In *IET Conference Image Processing (IPR 2012)*, pp. 1-5.
- [14] Caicedo, J., Cruz-Roa, A., and Gonzalez, F. (2009) "Histopathology image classification using bag of features and kernel functions," *International Conference on Artificial Intelligence in Medicine, Lecture Notes in Computer Science*, Vol. 5651, pp. 126–135.
- [15] Jégou, H., Douze, M. and Schmid, C. (2010) "Improving bag-of-features for large scale image search," *International Journal of Computer Vision*, Vol. 87, No. 3, pp. 316–336.
- [16] Nister, D. and Stewenius, H. (2006) "Scalable recognition with a vocabulary tree," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2161–2168.
- [17] Philbin, J., Chum, O., Isard, M., Sivic, J. and Zisserman, A. (2007) "Object retrieval with large vocabularies and fast spatial matching," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1-8.
- [18] Rahman, M., Antani, S. and Thoma, G. (2011) "Biomedical cbir using "bag of keypoints" in a modified inverted index," *International Symposium on Computer-Based Medical Systems*, pp. 1–6.
- [19] Sivic, J. and Zisserman, A. (2003) "Video google: A text retrieval approach to object matching in videos," *International Conference on Computer Vision*, Vol. 2, pp. 1470–1477.

- [20] Wang, J., Li, Y., Zhang, Y., Xie, H. and Wang, C. (2011) "Boosted learning of visual word weighting factors for bag-of-features based medical image retrieval,"