VISUAL CHARACTER N-GRAMS FOR CLASSIFICATION AND RETRIEVAL OF RADIOLOGICAL IMAGES

Pradnya Kulkarni[1], Andrew Stranieri[1], Siddhivinayak Kulkarni[1], Julien Ugon[1] and Manish Mittal[2]

[1]Centre for Informatics and Applied Optimisation, Federation University
PO Box 663, Ballarat 3353, Victoria
p.kulkarni,a.stranieri,s.kulkarni,j.ugon@federation.edu.au
[2]Lakeimaging, Ballarat, Victoria
mittalm@lakeimaging.com.au

ABSTRACT

Diagnostic radiology struggles to maintain high interpretation accuracy. Retrieval of past similar cases would help the inexperienced radiologist in the interpretation process. Character n-gram model has been effective in text retrieval context in languages such as Chinese where there are no clear word boundaries. We propose the use of visual character n-gram model for representation of image for classification and retrieval purposes. Regions of interests in mammographic images are represented with the character n-gram features. These features are then used as input to back-propagation neural network for classification of regions into normal and abnormal categories. Experiments on miniMIAS database show that character n-gram features are useful in classifying the regions into normal and abnormal categories. Promising classification accuracies are observed (83.33%) for fatty background tissue warranting further investigation. We argue that Classifying regions of interests would reduce the number of comparisons necessary for finding similar images from the database and hence would reduce the time required for retrieval of past similar cases.

KEYWORDS

N-gram, Bag of Phrases, Neural Network, Mammograms, Classification, Radiological Images

1. INTRODUCTION

Cancer is one of the major causes of death. Cancer detection is possible with the help of radiology. Radiological images are stored in digital format now a day to reduce storage burden and ease of access. It has been observed that diagnostic radiologists struggle to maintain high interpretation accuracy while trying to fulfil the efficiency goals [1]. Misdiagnosis by radiologists due to non-medical reasons are reported to be in the range of 2% to 4% [2][3]. The misinterpretation of cancer lesions increases patients' anxiety, health-care costs and also results in death in rare cases. There are vast amount of radiological 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 of a normal case. There are inter-observer variations among the radiologists due to difference in their experience level. There are also intra-observer variations in reading radiological images caused by fatigue and stress. The two types of errors a radiologist is prone to make. First, *perception error* where he or she is not able to notice the abnormality and second, *interpretation error* where he or she notices the abnormality but fails to diagnose it correctly due to lack of experience. 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. The perception

errors can be handled using image processing algorithms and the abnormalities can be notified to the radiologist. While interpreting an image the radiologist makes use of two types of knowledge. One is formalized *domain knowledge* which is present in texts and other written documentation. Second is implicit or *tacit knowledge* consisting of radiologists' individual expertise, organisational practices and past cases [4]. A junior radiologist lacks the experience and hence retrieval of similar past cases would help improve the interpretation accuracy.

Text based search approach could work for retrieving the past similar cases only if all the previous images have been perfectly annotated using the keywords in controlled vocabulary. Annotating every single image in the database is a very tedious job. Content based image retrieval (CBIR) is based on retrieving similar images by using image content. Three levels of features can be used for describing content of an image. Low level features such as colours, textures, shapes, Medium level features such as objects, and high level features such as scene which has semantic meaning. CBIR helps users to find similar images in multimedia applications [5]. CBIR techniques have been successfully applied in various real world applications such as security, arts, entertainment, geographical information systems. This saves users' time considerably in browsing unstructured data and image related information. Radiology images pose challenges compared to multimedia images as they contain rich information, specific features that need to be carefully recognized for medical image analysis. For example, Cysts and solid nodules usually have uniform internal density, whereas complex lesions have heterogeneous characteristics. These images are mostly grey scale images and hence colour information is not available. Radiologists are often interested in areas where the abnormalities exist, called as regions of interests (ROI). The region similarity can be calculated using different texture, shape, and margin features.

In order to retrieve similar images the region of interest from the query image is segmented manually or automatically and then the features of ROIs are compared. Comparing every single ROI in the database with the query ROI would take considerable amount of time. Classification of ROI into different categories will help reduce the number of comparisons necessary to find out the similar images. Different texture features such as histograms, co-occurrence matrix, gabor filters and wavelet transforms have been previously used for classification of regions of interest[6][7][8].

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 [9]. 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 [10] shows that the visual n-gram representation can improve the retrieval performance over bag-of-words model for medical as well as non-medical image datasets. In languages such as Chinese, where there are no specific word boundaries a concept called *character n-gram* is very useful for document retrieval. Because it is hard to find specific word boundaries for an image, we define a novel visual *character n-gram model* for representing an image for image retrieval.

The purpose of this paper is to extract character 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 classification done this way can thus help retrieval of past similar images and would be helpful as a diagnostic support aid in radiology. The paper is organised as follows. Section 2 details the literature review, section 3 gives information about proposed character n-gram model for image representation, research methodology is described in Section 4, Section 5 demonstrates experimental results and analysis of these results by varying n and effect of background tissue on classification accuracy and Section 6 concludes the paper along with future directions.

2. LITERATURE REVIEW

Content Based Image Retrieval has been experimented in radiology for teaching, research and diagnostic purposes. Various features and approaches have been investigated. In [11] training images are partitioned into 8×8 equal grid of 64 non-overlapping regions. Colour features composed of mean and standard deviation of each channel in RGB colour space and texture features composed of second order texture moments are used. Two different global thesauruses in the form of similarity and a correlation matrix where each element of the matrices defines concept similarities or co-occurrence relationships respectively are utilized in a quadratic form of distance measure to compare a query and database images. Results show that concept based feature representation performs much better compared to Colour layout descriptor (CLD) and Edge histogram descriptor (EHD). The quadratic similarity matching performs better when compared to Euclidean and Cosine similarity matching. However, a limitation of this system is the construction of the global matrix which is prohibitively difficult for large collections. Arakeri et al [12] classify brain tumours based on the rotation invariant shape features (circularity and irregularity) and texture features from tumour region into benign and malignant classes and then use the wavelet based Fourier descriptors and local binary patterns for checking similarity. Shape similarity is calculated using Euclidean distance and the texture similarity is calculated using Chi-square distance. Experiments show that this system [12] works better than the earlier developed systems [13][14][15]. Liu [16] developed a CT lung image retrieval system for assisting differential diagnosis. They found that most important feature of lung cancer is density (texture). This system uses fixed size rectangles to mark the Region of Interest (ROI) area. A three dimensional lung-based coordinate system, which was suggested by the radiologist, was constructed to describe the position information of selected blocks in order to eliminate the variance inherent in lung's size and shape. Major 2D FFT coefficients representing texture features of each marked region are then extracted. The centre information of the marked region is used for position matching. The experimental results show that in simple query, 96.6% of images can be correctly retrieved with the displacement up to 22% of the block size. ASSERT proposes physician-in -the-loop approach in which the human delineates the pathology bearing regions (PBR) and a set of anatomical landmarks in the image when the image is entered into the database. Features such as Texture (grey level co-occurrence matrix), Shape features (longer axis, shorter axis Fourier descriptors and moments) and the position information of a PBR are used. Indexing is based on hashing function based on feature grouping. Results show that the retrieval efficiency is quite good [17]. medGIFT considers image as a document containing several visual words. The image is described in terms of a set of words by mapping the visual features derived from local and global colour or texture cues to these keywords. Thus the image retrieval problem is converted into a standard text based information retrieval. Around 1000-3000 features per image are extracted such as global colour histograms colour blocks, histogram of Gabor filters. Limitation of this system is it relies on the annotation tool. Lim and Chevallet [18] adopt a content description approach similar to medGIFT. A statistical learning framework is used to extract vocabularies of meaningful medical terms (VisMed terms) associated with visual appearance of image samples. Manually cropped image regions are used as a training set to statistically learn the terms Image retrieval in medical applications. IRMA [19] aims at classifying the medical images according to image modality, body orientation, anatomic region and biological region and uses about 1024 Edge based and texture based features. Similarity is calculated for the whole region and hence similarity based on ROI is not possible. In [20] Tao et al developed a medical image search engine for diffuse parenchymal lung disease. The system combines offline learning aspect of classification based CAD systems with online learning aspect of CBIR systems. The Local Binary Patterns (LBP) algorithm is used to extract volumetric texture-based features. 3D bullae size histogram is also used. Patient-specific disease-specific features are used as supplement to cross-patient disease-specific features extracted offline to achieve better results.

In [21] a tool was proposed for uterine cancer research called cervigramFinder. The user can either delineate the boundary of the region of interest or choose from the list of clinically important regions. The limitations of this system are from the usability point of view. Huang et al [22] describes a diagnostic support tool for brain disease of children. Geometric, texture features and Fourier coefficients are used. In this system, the user selects a seed pixel, so that the computer can retrieves ROI from the query image. The database consists of Brain MR images of known brain lesions (Children's hospital Los Angeles) from 2500 patients aged between 0 and 18, Relevant patient information, Radiological reports, Histopathology reports, and other relevant reports. 60% of the match results contain true diagnoses or that judged by expert neuro-radiologist to represent plausible alternatives. In [23] Muller et al integrated an image retrieval system into a radiology teaching file system, and the performance of the retrieval system was evaluated, with use of query topics that represent the teaching database well, against a standard of reference generated by a radiologist. The results of this evaluation indicate that content-based image retrieval has the potential to become an important technology for the field of radiology, not only in research, but in teaching and diagnostics as well. CBIR have been deployed in teaching contexts by exporting interesting cases from PACS or viewing station. Casimage [24], myPACS (http://www.mypacs.net) are some of the example systems that offer maximum flexibility. Further a content based search could be used to compliment text-based methods of accessing data. Students could be able to browse the data in an easy way by using such systems. Lecturers can find cases similar in appearance but different in diagnosis. Shyu et al developed a system for teaching [25].

For detection of mammographic masses, the methods reported to date mainly utilize morphological features to distinguish a mass from the normal mammographic background [26]. Linear discriminant analysis is used in [27], 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 [28]. In this case 119 regions were used for the training (jack and knife method) and the remaining 119 regions of 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 [29] 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 [30]. 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 [31]. 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 [32].

A bag of words model used effectively in text retrieval is further extended as Bag-of-visual-words for image classification and retrieval [33][34][35][36][37][38][39]. Researchers have taken various approaches for representing an image with the bag-of-words model. It has been used to analyse mammographic images in [40]. 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 [41]. This approach has also shown to be useful in lung image classification [42].

The bag of words model has proven to be useful for classification (88% classification accuracy) of normal versus microcalcifications in mammographic images [43]. This approach is also been shown useful in the classification of focal liver lesions on CT images in [44]. In [45] 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. Case based fracture image retrieval system [46] used bag-of-visual words model with two sampling strategies for evaluation. One used a dense 40x40 pixel Grid sampling and the second one used the standard SIFT features. It is observed that a dense sampling strategy performed better than the SIFT detector–based sampling. Similar conclusion is obtained by [42]. Moreover, it is concluded from the experiments presented in [46], that SIFT based systems with optimized parameters slightly outperformed the GIFT (GNU Image finding tool) system.

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 [10]. 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-phrases model is more efficient in representing image semantics and improves the retrieval performance. Character n-gram model has not been tried before for image retrieval.

3. CHARACTER N-GRAMS FOR IMAGE REPRESENTATION

In text retrieval context, Word n-grams are 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}. The 1 gram representation is the bag-of-words approach. Phrases formed by n consecutive words enrich the semantics and provide more complete representation of the document thereby increasing the retrieval performance.

Image representation is very important for image classification and retrieval application. Many studies have tried to come up with different ways of representing image. Low level image features such as colour, texture and Shape have been used in the past. One of the state of the art techniques for image representation is bag-of-visual-words. In this approach image is represented by using dictionary composed of different visual words. Visual words are local image patterns, which represent relevant semantic information. Image is then represented with the number of occurrences of each visual word present in the image. This approach is similar to bag-of-words model used for document representation. The bag-of-visual-words model has following limitations:

- Simple counts of visual words occurring in an image are used and hence the spatial relations between the words are lost. Different objects can share same words but could have different layout
- Noisy words due to coarseness of the vocabulary construction process.
- High dimensional nature of key point descriptors causes very high computation times.

Bag-of-visual-phrases model has been proposed which is similar to bag-of-phrases model for text retrieval where a visual phrase is formed by combining more than one visual word adjacent to each other. The image would then be represented with the number of occurrences of the visual phrase in an image. This is a more powerful representation as it takes into account the spatial relationship between consecutive visual words. Couple of problems mentioned earlier such as noisy words and high computation times because of high complexity are still present for this approach.

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". For languages such as Chinese, where there are no specific word boundaries, the character n-gram features have been very efficient in document retrieval. If we consider an image, it is very hard to find specific visual word boundary. We would therefore argue that the visual character

n-gram representation of images would work better for image classification and retrieval. To our knowledge, this concept has not been investigated till date. For this purpose we would consider the grey level value for a pixel is a visual character and use a sliding window of n characters to count the number of occurrences of sequence of n grey level values. These counts would be able to catch the shape, margin and texture properties of an image. Moreover we take into account every pixel value and hence information is not lost in vocabulary construction. Also the spatial relationships are taken into consideration by using this strategy. Most importantly the algorithm for calculation of visual character n-grams are simple and hence do not present computational overhead.

4. 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.
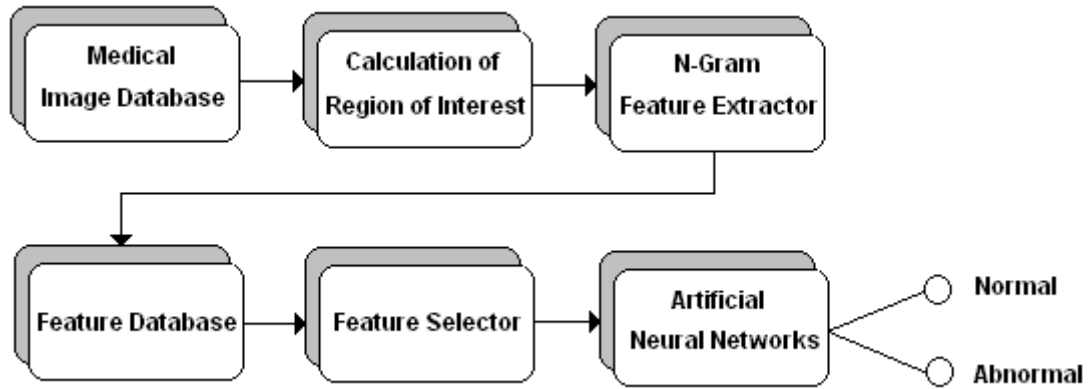

Figure 1. Block diagram of the proposed technique

4.1 Calculation of Region of Interest

We are using benchmark miniMIAS database [47]. 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.
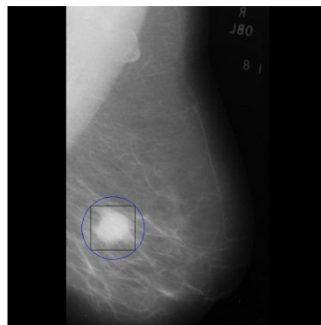

Figure 2: Example ROI extraction containing a circumscribed mass.

4.2 Grey scale reduction

The images are in portable grey map format. Each pixel is represented with a grey level between 0 and 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 ith bin is composed of grey levels illustrated by equation (1). Figure 3 shows the grey scale reduction process using 8 grey level bins.

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

Where

$$binsize = \frac{highest\ grey\ level - lowest\ grey\ level}{number\ of\ bins} \tag{2}$$

30	55	62	131	140	102	145
28	48	25	75	78	83	58
25	180	200	205	104	108	88
63	196	220	210	99	74	69

1	2	2	5	5	4	5
1	2	1	3	3	3	2
1	6	7	7	4	4	3
3	7	8	8	4	3	3

Original Image   Gray scale reduced image with 8 bins

Figure 3: Grey Scale reduction using 8 bins

4.3 N gram feature extractor

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 adjaceny 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.

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

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.

Window 2 ————
Window 1 ————
Direction ---

1	2	2	5	5	4	5
1	2	1	3	3	3	2
1	6	7	7	4	4	3
3	7	8	8	4	3	3

Figure 4: 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 4 shows how a sliding window of size 3 pixels is used to calculate three gram features of an image.

4.4 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.

4.5 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 [28]. 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.

$$Abnormal\ Classification\ Accuracy = n_{ab}/N_{ab} \tag{4}$$

$$Normal\ Classification\ Accuracy = n_n /N_n \tag{5}$$

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

5. EXPERIMENTAL RESULTS AND ANALYSIS

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

5.1 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 [47]. 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 with different background tissues are shown in Figure 5.
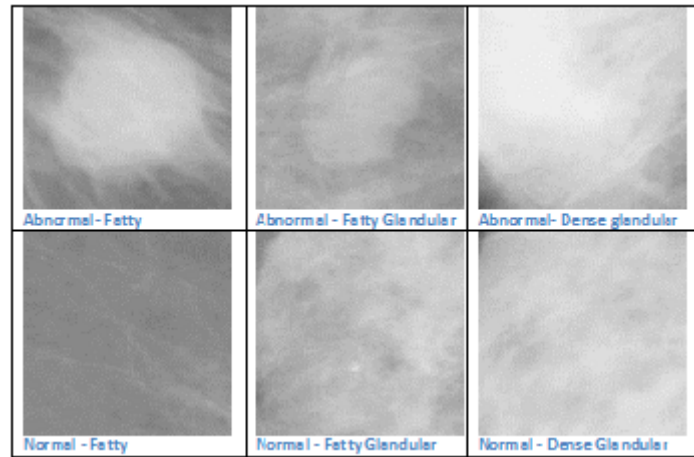
Figure 5 Abnormal and Normal ROIs with different background tissues

5.2 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

5.3 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. In the training phase the normalized n-gram features are presented to input and the outputs are set to one or zero depending on the ground truth available from the miniMIAS database. Thus weights are learned by using supervised learning strategy. In testing phase normalized n-gram features from the query image are applied at the input. Neural network then uses the learned weights to classify the query region of interest into normal or abnormal category.

5.4 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. Table 2 shows the effect of increasing the grey level bins, using three gram features. It was observed that the classification accuracy did not increase significantly with the increase in the number of grey level bins used. Hence we reduce the ROIs to 8 grey level bins. Table 3 shows the effect of increasing n on the classification accuracy.

Table 2: 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%

Table 3: 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%
Twogram	64	76.19%	64.58%	68.11%
Threegram	512	80.95%	66.66%	73.81%
Fourgram	4096	85.71%	70.83%	78.27%

5.5 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 for all background tissue types. These six pure sequence three gram features are used as inputs to the neural network. Table 4 and Figure 5 summarize the results.

Table 4: Effect of background tissue on classification accuracy

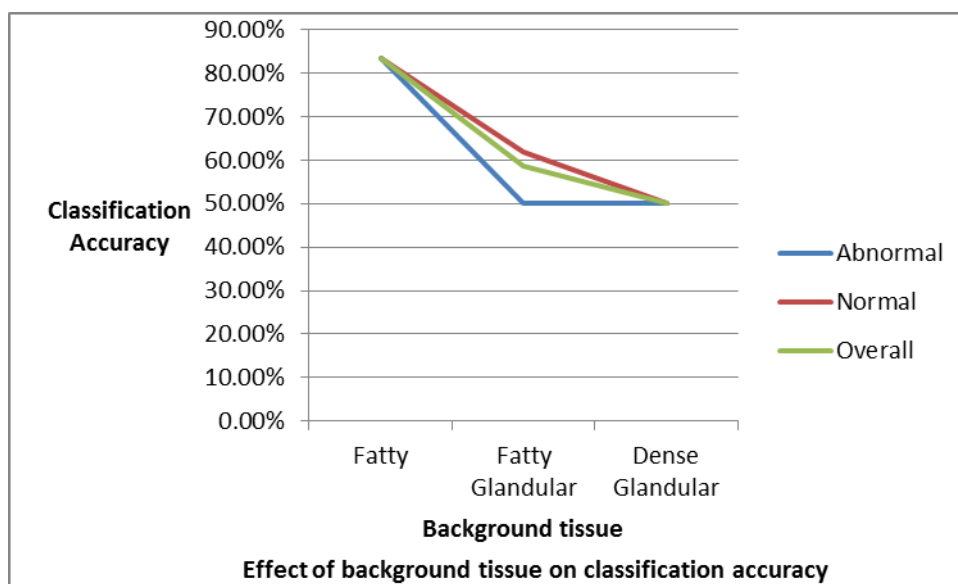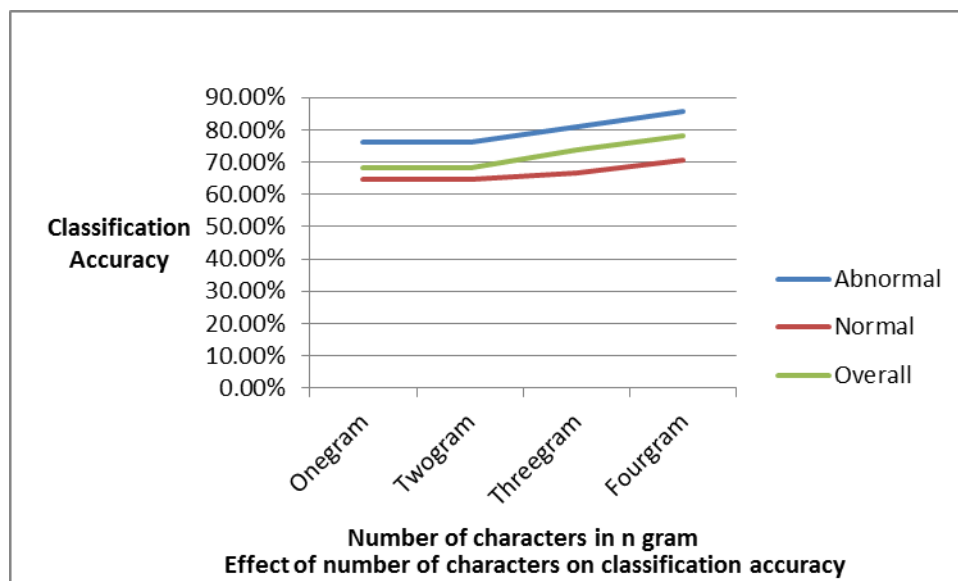|	Classification accuracy		
Background tissue	Abnormal	Normal	Overall
Fatty	83.33%	83.33%	83.33%
Fatty Glandular	50.00%	61.90%	58.62%
Dense Glandular	50.00%	50.00%	50.00%

Figure 5 Effect of N and Background tissue on classification accuracy

6. CONCLUSION AND FUTURE RESEARCH

Retrieval of past similar cases can help improve the interpretation accuracy of inexperienced radiologists. Representation of an image is vital for image retrieval applications. Classification of images into different categories is essential in order to reduce the computation time during retrieval process. A novel technique for representation of image, called visual character n-gram is proposed. N-gram features are experimented for classification of mammograms into normal and abnormal categories. The regions of interests were reduced in grey scale levels to reduce computations complexity. A supervised back-propagation neural network was used for classification purpose. We used Three fold cross validation technique to test the classification accuracy. Numbers of experiments were 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. It was observed that the classification accuracy for fatty background was much more than the one for fatty-glandular or dense-glandular background. As the number of visual words in a phrase increase, significant improvement in overall classification accuracy is observed. Although, direct comparisons with other neural network and co-occurance matrix approaches reported in the literature are not immediately possible because each study deploys different image sets and ROI, the classification accuracies obtained here are promising and warrant further investigation. Future research will incorporate investigating calculation of n-gram features in different directions for improving classification accuracy for fatty glandular and dense glandular background tissue. Dealing with different resolutions and image sizes is another challenge. We would like to investigate a similarity measure for retrieval of these images and lastly, fusion of n-gram image features and text features of the associated radiology reports

7. REFERENCES

[1] Rubin, G. D. (2000). Data explosion: the challenge of multidetector-row CT. *European journal of radiology*, *36*(2), 74-80.

[2] Siegle, R. L., Baram, E. M., Reuter, S. R., Clarke, E. A., Lancaster, J. L., & McMahan, C. A. (1998). Rates of disagreement in imaging interpretation in a group of community hospitals. *Academic radiology*, *5*(3), 148-154.

[3] Soffa, D. J., Lewis, R. S., Sunshine, J. H., & Bhargavan, M. (2004). Disagreement in interpretation: a method for the development of benchmarks for quality assurance in imaging. *Journal of the American College of Radiology*,*1*(3), 212-217.

[4] Montani, S., & Bellazzi, R. (2002). Supporting decisions in medical applications: the knowledge management perspective. *International Journal of Medical Informatics*, *68*(1), 79-90.

[5] Kulkarni, S. (2010), Natural Language based Fuzzy Queries and Fuzzy Mapping of Feature Database for Image Retrieval, *Journal of Information Technology and Applications*, 4(1), pp. 11-20.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] Rahman, M., Antani, S. and Thoma, G. (2010): Local concept based medical image retrieval with correlation enhanced similarity matching based on global analysis, *Computer Vision and Pattern Recognition Workshops, IEEE Computer Society Conference,* San Francisco, CA, 87 – 94

[12] Arakeri, M. and Reddy, R., (2012). Medical image retrieval system for diagnosis of brain tumor based on classification and content similarity, *India Conference (INDICON), 2012 Annual IEEE* 2012,Kochi, India 416-421.

[13] Emmanuel, M., Ramesh Babu ,D., Potdar,G.,Sonkamble, B.and Game, P. (2007) : Content based medical image retrieval, *International Conference on Information and Communication Technology in Electrical Sciences*, IEEE, 712-717

[14] Dube, S.,El-Saden, S., Cloughesy, T. and Sinha, U. (2006): Content based image retrieval for MR image studies of brain tumors, *IEEE International conference on Engineering in Medicine and Biology*, 3337–3340

[15] Rahman,M., Bhattacharya, P. and Desai, B. (2007): A framework formedical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback, *IEEE Transaction on Information Technology in Biomedicine*, 11(1): 58–69

[16] Liu,C., Tai,P.,Chen,A., Peng, C., Lee, T., Wang, J. (2001): A content –based CT lung image retrieval system for assisting differential diagnosis image collection, *IEEE International Conference on Multimedia and Expo*, Tokyo, Japan. 174 – 177

[17] Shyu, C. R., Brodley, C., Kak, A., Kosaka, A., Aisen, A. and Broderick, L. (1999) ASSERT, A physician-in-the-loop content-based image retrieval system for HRCT image databases", *Computer Vision and Image Understanding*, 75(½) 111-132

[18] Lim J, Chevallet J. (2005): Vismed: A visual vocabulary approach for medical image indexing and retrieval. in Second Asia Information Retrieval Symposium. Jeju Island, Korea

[19] Lehmann, T.,Guld, M., Thies,C., Plodowski, B. Keysers, D., Ott, B.,Schubert, H. (2004): IRMA--content-based image retrieval in medical applications. Stud Health Technol Inform 107(2): 842-846.

[20] Tao Y, Zhou X, Bi, J., Jerebkoa, A., Wolf, M., Salganicoff, M., Krishnana A. (2009):An adaptive knowledge-driven medical image search engine for interactive diffuse parenchymal lung disease quantification. *Medical imaging : computer-aided diagnosis*, vol 7260. SPIE, p 726007.

[21] Xue, Z., Antani, S., Long, L., Jeronimo, J., Schiffman, M., Thoma, G. (2009): CervigramFinder: a tool for uterine cervix cancer research. *Proc. AMIA Annual Fall Symposium,* San Francisco,California, USA, 1-24, Curan, Red Hook.

[22] Huang, H., Nielsen, J., Nelson, M., Liu, L. (2005) : Image-matching as a medical diagnostic support tool (DST) for brain diseases in children. *Comput Med Imaging Graph* 29(2-3): 195-202

[23] Muller, H., Rosste, A., Garcia, A., Vall_ee, J. and Geissbuhler, A. (2005): Benefits from content based visual data access in radiology. *RadioGraphics, 25*, pp. 849-858.

[24] Muller, H., Rosset, A., Vallee, J. and Geissbuhler, A. (2003): Integrating content-based visual access methods into a medical case database, *Medical Informatics Europe conference* 2003, pp. 480-485.

[25] Shyu, C., Kak, A., Brodley, C., Pavlopoulou, C., Chyan, M. and Broderick, L. (2000): A web-based CBIR-assisted learning tool for radiology education - anytime and anyplace, *IEEE international conference on multimedia and expo*

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] 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," *International Conference on Image and Graphics (ICIG)*, pp. 1035–1040.

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

[41] Cheng, E., Xie, N., Ling, H., Bakic, P. R., Maidment, A. D., & Megalooikonomou, V. (2010) "Mammographic image classification using histogram intersection," *IEEE International Symposium on Biomedical Imaging: From Nano to Macro,* pp. 197-200.

[42] Avni, U., Goldberger, J., Sharon, M., Konen, E., & Greenspan, H. (2010) "Chest x-ray characterization: from organ identification to pathology categorization", *Proceedings of the ACM International Conference on Multimedia Information Retrieval,* pp. 155-164.

[43] Diamant, I., Goldberger, J., & Greenspan, H. (2013) "Visual words based approach for tissue classification in mammograms," *SPIE International Society for Optics and Photonics Medical Imaging.* pp. 867021-867021-9.

[44]   Safdari, M., Pasari, R., Rubin, D., & Greenspan, H. (2013) "Image patch-based method for automated classification and detection of focal liver lesions on CT," *SPIE International Society for Optics and Photonics Medical Imaging,* pp. 86700Y-86700Y-7.

[45]   Yang, W., Lu, Z., Yu, M., Huang, M., Feng, Q., & Chen, W. (2012) "Content-based retrieval of focal liver lesions using bag-of-visual-words representations of single-and multiphase contrast-enhanced CT images", *Journal of Digital Imaging*, Vol. 25, No. 6, 708-719.

[47]   Suckling, J. *et al* (1994) "The Mammographic Image Analysis", *Society Digital Mammogram Database International Congress Series*, Vol. 1069. pp375-378.

[46]   Zhou, X., Stern, R., & Müller, H. (2012). Case-based fracture image retrieval.*International journal of computer assisted radiology and surgery*, *7*(3), 401-411.

Short Biography:

Pradnya Kulkarni is currently a PhD student at Centre for Informatics and Applied Optimization (CIAO), Federation University, Australia. Her PhD topic is "Integration of Image and Text Retrieval for Radiological Diagnostic Support System". Prior to this, she has completed postgraduate diploma in information technology from University Of Ballarat (now Federation University), Australia and Electronics Engineering from India. She has more than four years of teaching experience at University of Ballarat, Australia. She also possess industrial experience of five years working as Software Developer, Research Assistant (Embedded Systems, C, Visual Basic,Web development using java, SQL, XML), in India, Canada and Austalia. Her research interests are Health Informatics, Image Processing/Retrieval and Artificial Intelligence.