

Classification of Breast Tumor in Mammogram Images Using Unsupervised Feature Learning

¹Aidarus M. Ibrahim, ¹Baharum Baharudin, ¹Abas Md Said and ²P.N. Hashimah

¹Department of Computer and Information Sciences, Faculty of Science and Information Technology, Universiti Teknologi PETRONAS, 31750, Tronoh, Perak, Malaysia

²Faculty of Computing and Mathematical Sciences, Universiti Teknologi MARA, Malaysia

Article history

Received: 24-12-2016

Revised: 08-03-2016

Accepted: 10-05-2016

Corresponding Author:

Aidarus M. Ibrahim
Department of Computer and Information Sciences, Faculty of Science and Information Technology, Universiti Teknologi PETRONAS, 31750, Tronoh, Perak, Malaysia
Email: aidarus_g02440@utp.edu.my

Abstract: In this study, we propose a learning-based approach using feature learning to minimize the manual effort required to extract features. Firstly, we extracted features from equally spaced sub-patches covering the input Region of Interest (ROI). The dimensionality of the extracted features is reduced using max-pooling. Furthermore, spherical k-means clustering coupled with max pooling (k-means-max pooling) is compared with well-known feature extraction method namely Bag-of-features. The resulting feature vector is fed to two different classifiers: K-Nearest Neighbor (K-NN) and Support Vector Machine (SVM). The performance of these classifiers is evaluated to use of Receiver Operating Characteristics (ROC). Our results show that k-means-max pooling, combined with K-NN, achieved good performance with an average classification accuracy of 98.19%, sensitivity of 97.09% and specificity of 99.35%.

Keywords: Mammogram, Breast Cancer, K-Means Clustering, Max-Pooling, Bag-of-Features, Classifier

Introduction

Breast cancer is disease that is diagnosed among the women in the world (Marrocco *et al.*, 2010). Mammography is used as early screening tool for breast cancer. The studies in (Perlmutter *et al.*, 1997; Rawashdeh *et al.*, 2013; Rojas Domínguez and Nandi, 2008), reported that the radiologist may over look some abnormalities during mammogram image screening. Hence, Computer Aided (CAD) system helps the radiologist to double check the mammogram images. A lot of literature has been reviewed in the area of CAD systems for breast cancer as well as techniques for improving classification accuracy (Etehadtavakol *et al.*, 2013; Korkmaz and Korkmaz, 2015; Pak *et al.*, 2015; Ramirez-Villegas and Ramirez-Moreno, 2012; Wang *et al.*, 2014). Feature representation is one of the techniques that were developed for improving the classification accuracy of CAD. Feature representation contains set of techniques that converts the input data into more meaningful representation so that machine learning algorithms can simply use it.

Traditional manually designed feature descriptors in (Yadav *et al.*, 2015), such as gradient operators and filter banks, are not able to capture if complex variation related in frequency is found in medical

images (Rose *et al.*, 2010). This paves the way for designing efficient image descriptors. The process of learning features concentrates into main categories namely supervised and unsupervised learning. In the supervised learning algorithms, usually it performs the process of learning using classes or targets. However, unsupervised learning methods, the process of feature learning is carried out without using any classes, (Hinton *et al.*, 2006) (Mairal *et al.*, 2009), auto-encoders (Boureau and Cun, 2008), (Coates *et al.*, 2011).

To the best of our knowledge, limited work has been done on applying unsupervised feature learning to mammogram images. In this study, we adopted the unsupervised feature learning process based on unlabeled data (Coates and Ng, 2012). This approach can learn important and subtle features from the statistics of the image. By applying set of learned features coupled with labels to train the mammogram images, the features were extracted from patches of size 8×8 that represent the ROI of mammogram images. Moreover, the obtained features from the patches were pooled together to reduce the dimension of the features. Finally, classification is performed to classify between normal and abnormal using the features obtained from the unsupervised learning algorithm.

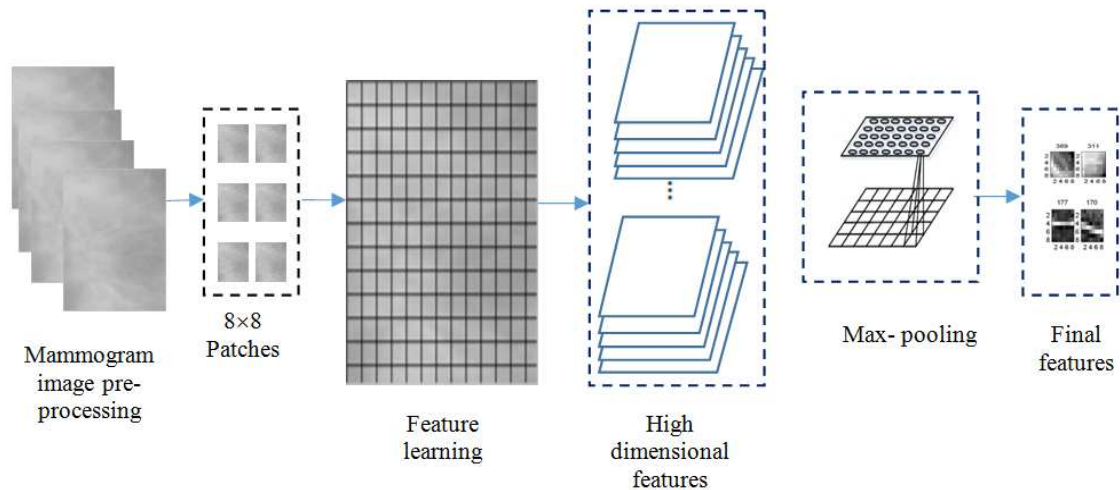


Fig. 1. Feature learning pipeline

The main contribution of this paper is the application of k-means-max pooling in mammogram images for automatic feature representation to enhance the classification of the mammogram images into normal and abnormal, as shown in Fig. 1.

Materials and Methods

The Digital Database for Screening Mammography (DDSM) is a publicly-available resource used by the image analysis community (Heath *et al.*, 2000). In this experiment, we used a total of 400 images, which represented 200 normal conditions and 200 abnormal (benign and malignant) conditions. The cropping operation was applied to the images to cut off the unwanted portions. ROIs were cropped to the size of 128x128 as shown in Fig. 2. The mammogram images were cropped manually by selecting the ROI of mammogram images. All the unnecessary parts such as background which are out of the tumor area were completely eliminated.

Feature Learning Architecture

The ROIs obtained from mammogram images were transformed into patches. Each patch of size of 8x8 were collected and stored as feature vector. However, mammogram images taken during the breast cancer screening might have variations such as brightness and contrast. Hence, to eliminate this issue image normalization. The learning architecture proposed by (Coates and Ng, 2012) uses the following procedure for feature learning representation of an image patch:

- Normalize for each patch, subtract out the mean of the intensities and divide by the standard deviation as follows:

$$x^i = \frac{x^i - \text{mean}(x^{(i)})}{\sqrt{\text{var}(x^{(i)}) + 10}} \quad (1)$$

where, $x^{(i)}$ are normalized patches and “mean” and “var” are the mean and variance of the elements of $x^{(i)}$.

Whitening Transform

Principal Component Analysis (PCA) is used to reduce the dimensionality of the data (Kambhatla and Leen, 1997). There is a similar preprocessing step called ZCA whitening (Coates and Ng, 2011), which is required for some algorithms. If we are training on images, the raw input is redundant, since neighboring pixel values are highly correlated. The aim of whitening is to make the input less redundant:

$$[V, D] := \text{eigh}(\text{cov}(x)); \text{So } SDV^T // = \text{cov}(x) \quad (2)$$

$$x^i := V(D + \epsilon_{zca} I)^{-1/2} V^T x^i, \forall_i$$

If $V D V^T = \text{cov}(x)$ is the eigenvalue decomposition of the covariance of the data points, x , then the whitening points are decomposed as $V(D + \epsilon_{zca} I)^{-1/2} V^T x$, where ϵ_{zca} is constant. In this study, to normalize the data, we set ϵ_{zca} as 0.01 for 8 by 8 pixel patches.

Spherical K-Means Clustering

Coates and Ng (2012), make a convincing case that K-means clustering is capable of learning dictionaries that can be easily used for classification. The K-means algorithm is particularly intriguing and it's very fast compared to standard K-means using Euclidean distances. Following the steps of a spherical K-means algorithm, which is much faster than using the conventional K-means (Zhong, 2005).



Fig. 2. Sample mammogram images: (a) normal images, (b) abnormal images

Max Pooling

Max pooling makes the feature learning into new reusable features that keeps significant information although removing redundant information. The typical properties of pooling are the robustness to cluster and compactness of representation (Boureau *et al.*, 2010). In this study, the following pooling steps were applied.

Notations:

- For instance, if the unpooled data are a $p \times K$ matrix of 1-of- k codes taken at P locations, we extract a single P -dimensional column v of 0s and 1s, indicating the absence or presence of the feature at each location

- The vector v is reduced by a pooling operation to a single scalar $f(v)$
- Max pooling: $f_a(v) = \frac{1}{p} \sum_{i=1}^p v_i$
- average pooling: $f_m(v) = \max_i v_i$
- Given two classes C_1 and C_2 , we examine the separation of conditional distributions:

$$p(f_m | C_1) \text{ and } p(f_m | C_2), p(f_a | C_1) \text{ and } p(f_a | C_2) \quad (3)$$

In this experimental design, we take the max of the cluster memberships over each 8x8 region. Max pooling reduces the dependency of the feature vectors on their exact placement in an image (each element of each 8x8 block gets treated about the same) and it also maintains a lot of the information that was in each of the feature vectors, especially when the feature vectors are expected to be sparse. Moreover, during the experiment design, we have taken different size of clusters (100 clusters and 150 clusters) to figure out the performance of the classifiers, as shown in section 3.

Bag-of-Features

Bag-of-features is a vector of occurrence counts of a vocabulary of local image features. The basic steps of bag-of-words when applied to images are as follows (Cheng *et al.*, 2010):

- Building a codebook for local patches
- Extracting the local patches from the ROIs of mammogram images
- representing an ROIs using the statistics of its quantized local patches
- Inference based on the statistics collected in step 3

The obtained set of ROI of mammogram images is represented and divided into testing and training sets. Then, visual vocabulary is built by clustering the patches from the training set followed by representing them as per image distributions. Moreover, each patch is represented as histogram of visual words drawn from vocabulary. In The experimental setup, 100 clusters and 150 clusters based on k-means as clustering method was applied to evaluate the classifier performance.

Classification

The learned features were stored as feature vector. The next step is to separate given classes into normal and abnormal. Two well know classifiers are adopted in this study as follows:

K-Nearest Neighbor

The KNN classifier usually applies either the Euclidean distance or the cosine similarity between the

training tuples and the test tuples. In this study, the Euclidean distance is applied in implementing the KNN($k = 1$) model for feature classification (Cunningham and Delany, 2007).

Support Vector Machine

A Support Vector Machine (SVM) is widely used in mammogram images classification due to its performance for the accuracy rate. This classifier achieves the classification rate by applying hyperplane. To see the classification performance of SVM for separating the two classes (normal Vs abnormal) and it is recommended the hyperplane has the largest distance to the nearest training data point of the two classes (Chang and Lin, 2011).

Performance Measures

The performance of the classifier is evaluated using a ten-fold cross validation method (Acharya *et al.*, 2015). The dataset is divided into ten equal parts. Each part contains the features from a similar proportion of images from normal and abnormal classes. Nine parts are used to train the classifiers. The remaining parts of the image features are used as a testing set. This process is repeated ten times using a different set in each case. In each fold, we apply several classification measures in order to obtain a more reliable comparison. Normal and abnormal mammographic images, respectively correspond to negative and positive samples. In this study, True Positive (TP) and True Negative (TN), respectively represent the number of abnormal and normal tests, which are properly classified. Similarly, False Positive (FP) and False Negative (FN), respectively, represent the number of normal and abnormal tests, which are incorrectly classified. The mathematical computation of performance measures are as follows:

$$\text{Accuracy} = \frac{TP + TN}{N} \times 100 \quad (4)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (6)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

Receiver Operating Characteristics

The Receiver Operating Characteristics (ROC) curve provides a visual representation of the tradeoff between true positive and false positives. They are the

percentage of correctly classified features with respect to the percentage of incorrectly classified negative features (Nascimento *et al.*, 2013). As shown in Fig. 4 to 7, the point (0,0) along curve represents a classifier that by default classifies all features as being negative, where a point (0,1) represent a classifier that positively classifies all features.

The experimental results is carried out using windows 8 and MATLAB R2013a. Some of the tools that were selected were image processing and statistical tool box with application of Laptop of intel Core i5-4200U CPU @1.6GHZ.

Results

Table 1 to 4 presents the classification results using KNN and SVM classifiers for k-means-max pooling and Bag-of-features. We have obtained the highest performance for k-means-max pooling as compared to Bag-of-features method with an average accuracy, sensitivity and specificity of 98.19, 97.09 and 99.35% respectively (at 150 clusters). It is also obvious from the results that the KNN classifier performs better than SVM, with a high accuracy of 98.19%. To get the best

optimum classifier with a single dataset, the classifier should not only have significant accuracy, but should also provide good sensitivity and specificity. Such a balance is required to make certain decisions that classify two classes (normal Vs. abnormal). Hence, this K-NN classifier was selected as the optimum classifier for this dataset mainly because it provided the best accuracy, sensitivity and specificity.

Figure 3 visualizes the resulting 50 clusters centers from k-means-max pooling. We sorted the data based on how often data centers are assigned to each cluster. So, the top left has the most elements and the bottom right has the least.

Figure 4 to 7 show the ROC curve for k-means-max pooling compared with Bag-of-features using KNN and SVM. It can be observed from these figures that the performance of k-means-max pooling is better in the KNN classifier at the 150 cluster. Furthermore, the ROC curve is more near to the y-axis for the KNN classifier than Bag-of-features, indicating a higher AUC value for KNN. Therefore, k-means-max pooling coupled with KNN has the highest discrimination capacity to separate the two classes (normal Vs. abnormal).

Table 1. Performance of feature learning at 100 clusters

Performance					
Classifier	Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)
KNN	K-means-max- pooling	97.50	96.48	98.77	98.49
	Bag-of-features	91.87	93.20	90.68	92.77

Table 2. Performance of feature learning at 150 clusters

Performance					
Classifier	Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)
KNN	K-means-max- pooling	98.19	97.07	99.39	98.50
	Bag-of-features	92.86	93.85	91.80	92.87

Table 3. Performance of feature learning at 100 clusters

Performance					
Classifier	Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)
SVM	K-means-max- pooling	88.6	81.77	100.0	90.36
	Bag-of-features	83.6	80.67	96.8	87.50

Table 4. Performance of feature learning at 150 clusters

Performance					
Classifier	Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)
SVM	K-means-max- pooling	89.46	82.58	100.0	83.27
	Bag-of-features	84.70	80.17	88.6	81.98

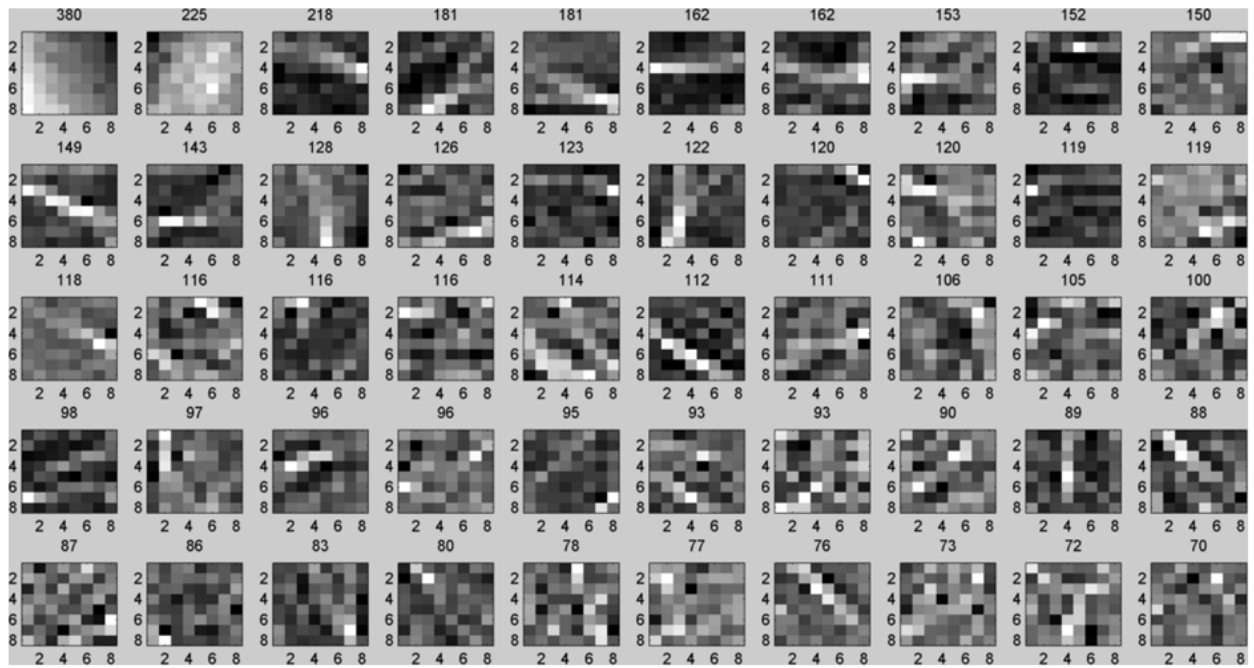


Fig. 3. Visualization cluster centroids

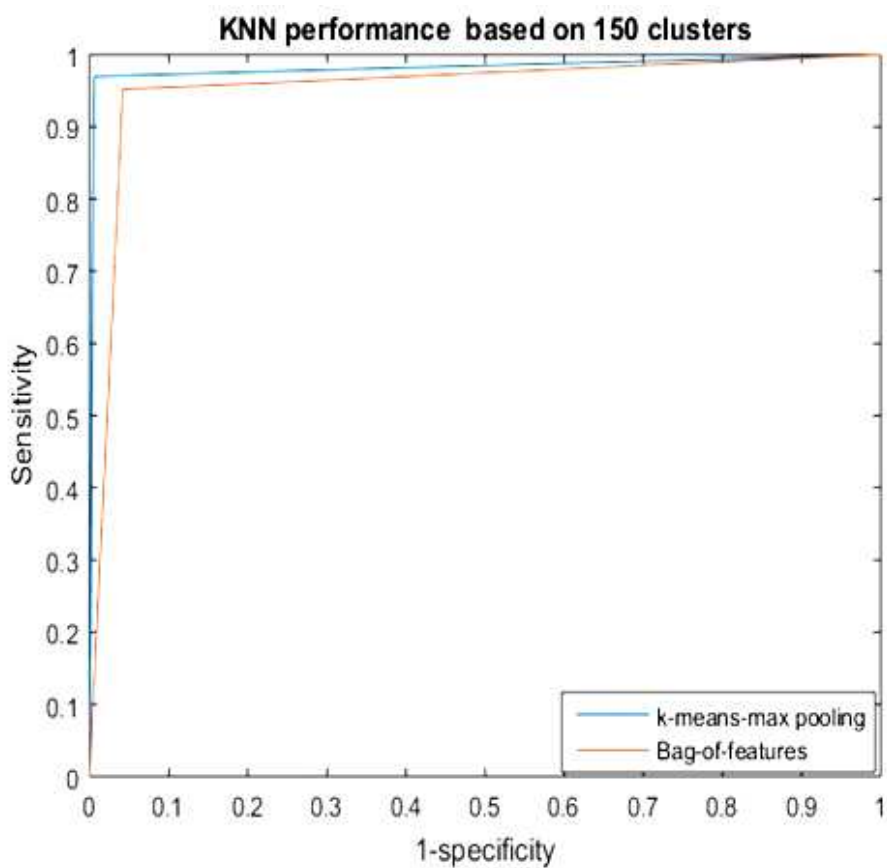


Fig. 4. ROC performance of KNN classifier (at 150)

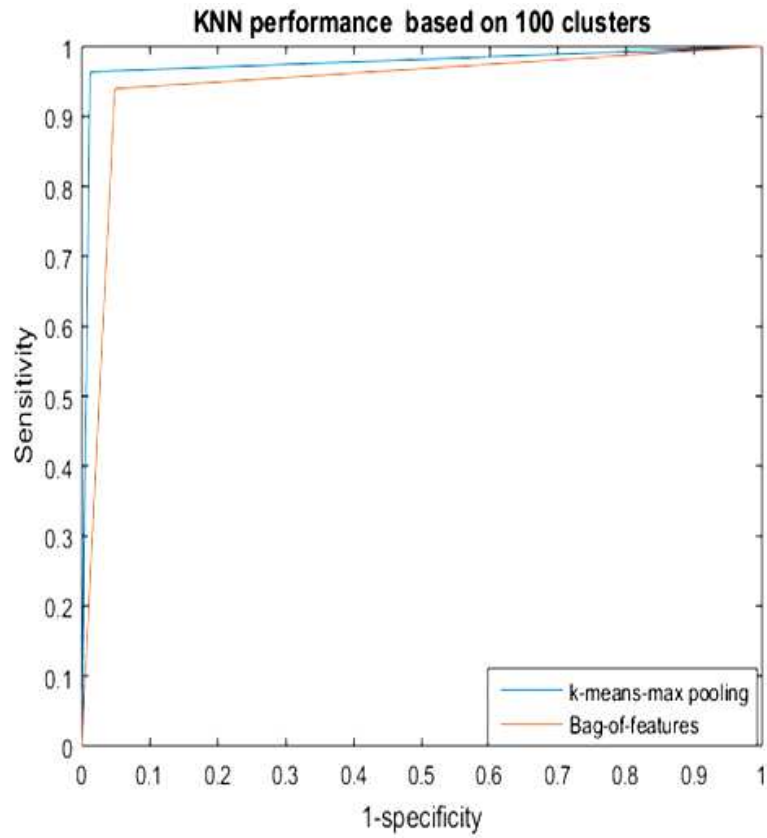


Fig. 5. ROC performance of KNN classifier (at 100 clusters)

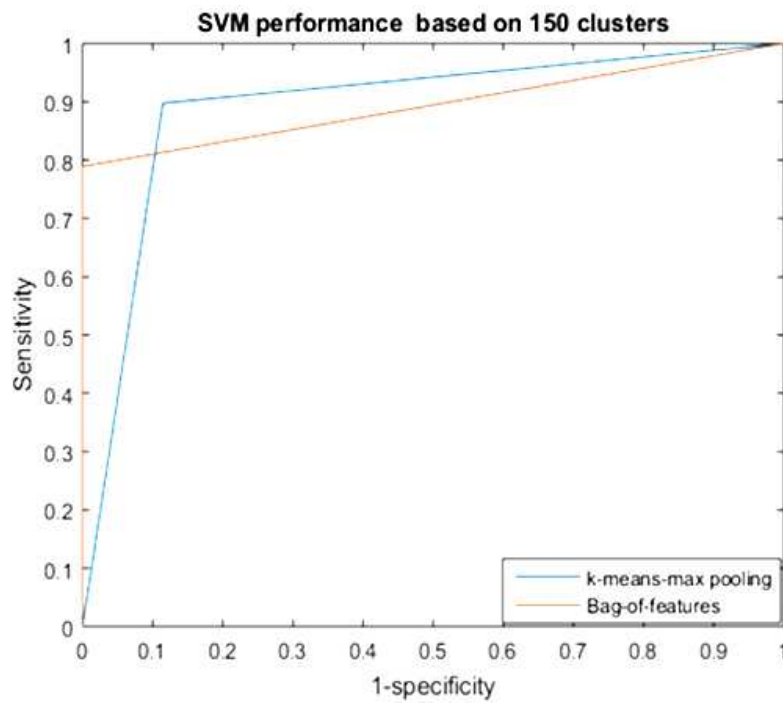


Fig. 6. ROC performance of SVM classifier (at 150) clusters

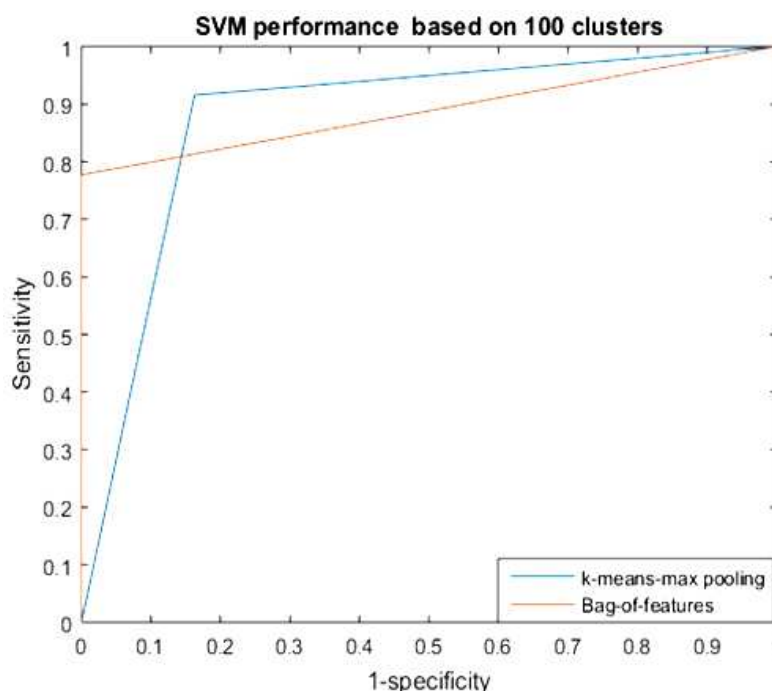


Fig. 7. ROC performance of SVM classifier (at 100 clusters)

Table 5. Summary of studies reporting automated detection of breast tumors from mammographic images

Authors	No. of Images	Features	Method used	Dataset	Classifier	Accuracy (%)
(Rocha <i>et al.</i> , 2014)	300 image	Gray Level Co-occurrence Matrixes (GLCM), Gray Level Run Length Matrixes (GLRLM) and Gray Level Gap Length Matrixes (GLGLM)	Gleason and Menhinick diversity indexes	DDSM	SVM	86
(Lim and Er, 2004)	343 images	First-order gradient distribution and Gray-level Co-occurrence Matrixes (GCMs)	First-order gradient distribution and gray-level co-occurrence matrices	DDSM	GDFNN	70
Our proposed study	400 images	Unsupervised feature learning	k-means-max pooling	DDSM	KNN, SVM	98.6-88.6

Discussion

Selecting the significant features from mammogram images is important for the mammogram image classification. Significant features increases the accuracy rate of the CAD system in breast cancer. The proposed methods to extract important features from mammogram images vary from one another and they are implemented either in supervised or unsupervised learning. In this study, we take the unsupervised feature learning method that will help in detecting breast tumors. We present the comparison of the results obtained using our technique and other techniques in the literature that also aim to diagnose mammogram breast tumors.

The summary of the studies reported by different authors is shown in Table 5. Rocha *et al.* (2014) classified 200 DDSM mammogram images into normal and malignant classes using Gleason and Menhinick diversity indexes and the SVM classifier. They have reported sensitivity and specificity of 90 and 83.33%, respectively. Lim and classified 343 DDSM

mammogram images into benign and malignant using first-order gradient distribution, gray-level co-occurrence matrices and Generalized Dynamic Neural Networks (GDFNN). They achieved a true-positive and false positive fractions of 95.0 and 52.8%, respectively. The advantages of our proposed method are as follows:

- We reported the highest sensitivity of 97.09% and a specificity of 99.35% in classifying the normal and abnormal classes compared to Bag-of-features
- The proposed system is able to detect two classes (normal Vs abnormal) with 98.19% accuracy; this will reduce the workload of clinicians
- Our proposed method is more reliable, as we used ten-fold stratified cross-validation and used 400 mammogram images in this study

Future Works

In the future, this study will focus outside the boundaries of the skin which is the area of cancer. By

performing possible multiple selection between benign and malignant

Conclusion

An accurate and fast diagnosis of breast cancer can help clinicians in their diagnosis. Hence, in this study, we proposed a CAD system based on k-means-max pooling for early detection of breast cancer. We reported an accuracy of 98.19%, sensitivity of 97.09% and specificity of 99.35% using the K-NN classifier for k-means-max pooling. We have also showed that, the features of k-means-max pooling performed better than Bag-of-feature. Our proposed system is able to clearly diagnose all normal and abnormal cases correctly (97.07% sensitivity) and reduce clinician workloads.

Acknowledgement

The authors would like to thank TM Lehmann, Dept. of Medical Informatics, RWTH Aachen, Germany and the University of South Florida for making the database available for research purposes.

Author's Contribution

All the authors have made same contribution in experimental results as well as writing the manuscript.

Ethics

We want to declare that this manuscript has not been published or is under consideration elsewhere.

References

- Acharya, U.R., E. Ng, L.W.J. Eugene, K.P. Noronha and L.C. Min *et al.*, 2015. Decision support system for the glaucoma using Gabor transformation. *Biomed. Signal Process. Control*, 15: 18-26.
DOI: 10.1016/j.bspc.2014.09.004
- Boureau, Y.L. and Y.L. Cun, 2008. Sparse feature learning for deep belief networks. *Proceedings of the Advances in Neural Information Processing Systems, (IPS' 08)*, pp: 1185-1192.
- Boureau, Y.L., J. Ponce and Y. LeCun, 2010. A theoretical analysis of feature pooling in visual recognition. *Proceedings of the 27th International Conference on Machine Learning, (CML' 10)*, Israel, pp: 111-118.
- Chang, C.C. and C.J. Lin, 2011. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2: 27-27.
DOI: 10.1145/1961189.1961199
- Cheng, E., N. Xie, H. Ling, P.R. Bakic and A.D. Maidment *et al.*, 2010. Mammographic image classification using histogram intersection. *Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, Apr. 14-17, IEEE Xplore Press, Rotterdam, pp: 197-200. DOI: 10.1109/ISBI.2010.5490381
- Coates, A. and A.Y. Ng, 2011. Selecting receptive fields in deep networks. *Proceedings of the Advances in Neural Information Processing Systems, (IPS' 11)*, pp: 2528-2536.
- Coates, A. and A.Y. Ng, 2012. Learning Feature Representations with k-Means. In: *Neural Networks: Tricks of the Trade*, Montavon, G., G.B. Orr and K.R. Müller (Eds.), Springer, New York ISBN-10: 3540653112, pp: 561-580.
- Coates, A., A.Y. Ng and H. Lee, 2011. An analysis of single-layer networks in unsupervised feature learning. *Proceedings of the International Conference on Artificial Intelligence and Statistics, (AIS' 11)*, Fort Lauderdale, FL, USA., pp: 215-223.
- Cunningham, P. and S.J. Delany, 2007. k-Nearest neighbour classifiers. *Multiple Classifier Syst.*
- Etehadtavakol, M., E. Ng, V. Chandran and H. Rabbani, 2013. Separable and non-separable discrete wavelet transform based texture features and image classification of breast thermograms. *Infrared Phys. Technol.*, 61: 274-286.
DOI: 10.1016/j.infrared.2013.08.009
- Heath, M., K. Bowyer, D. Kopans, R. Moore and W.P. Kegelmeyer, 2000. The digital database for screening mammography. *Proceedings of the 5th International Workshop on Digital Mammography, (WDM' 00)*.
- Hinton, G.E., S. Osindero and Y.W. Teh, 2006. A fast learning algorithm for deep belief nets. *Neural Comput.*, 18: 1527-1554.
DOI: 10.1162/neco.2006.18.7.1527
- Kambhatla, N. and T.K. Leen, 1997. Dimension reduction by local principal component analysis. *Neural Comput.*, 9 1493-1516.
DOI: 10.1162/neco.1997.9.7.1493
- Korkmaz, S.A. and M.F. Korkmaz, 2015. A new method based cancer detection in mammogram textures by finding feature weights and using kullback-leibler measure with kernel estimation. *Optik-Int. J. Light Electron Opt.*, 126: 2576-2583.
DOI: 10.1016/j.ijleo.2015.06.034
- Lim, W.K. and M.J. Er, 2004. Classification of mammographic masses using generalized dynamic fuzzy neural networks. *Med. Phys.*, 31: 1288-1295.
DOI: 10.1118/1.1708643

- Mairal, J., F. Bach, J. Ponce and G. Sapiro, 2009. Online dictionary learning for sparse coding. Proceedings of the 26th Annual International Conference on Machine Learning, Jun. 14-18, ACM, Canada, pp: 689-696. DOI: 10.1145/1553374.1553463
- Marrocco, C., M. Molinara, C. D'Elia and F. Tortorella, 2010. A computer-aided detection system for clustered microcalcifications. *Artif. Intell. Med.*, 50: 23-32. DOI: 10.1016/j.artmed.2010.04.007
- Nascimento, M.Z.D., A.S. Martins, L.A. Neves, R.P. Ramos and E.L. Flores *et al.*, 2013. Classification of masses in mammographic image using wavelet domain features and polynomial classifier. *Expert Syst. Applic.*, 40: 6213-6221. DOI: 10.1016/j.eswa.2013.04.036
- Pak, F., H.R. Kanan and A. Alikhassi, 2015. Breast cancer detection and classification in digital mammography based on Non-Subsampled Contourlet Transform (NSCT) and super resolution. *Comput. Meth. Programs Biomed.*, 122: 89-107. DOI: 10.1016/j.cmpb.2015.06.009
- Perlmutter, S.M., P.C. Cosman, R.M. Gray, R.A. Olshen and D. Ikeda *et al.*, 1997. Image quality in lossy compressed digital mammograms. *Signal Process.*, 59: 189-210. DOI: 10.1016/S0165-1684(97)00046-7
- Ramirez-Villegas, J.F. and D.F. Ramirez-Moreno, 2012. Wavelet packet energy, Tsallis entropy and statistical parameterization for support vector-based and neural-based classification of mammographic regions. *Neurocomputing*, 77: 82-100. DOI: 10.1016/j.neucom.2011.08.015
- Rawashdeh, M.A., R.M. Bourne, E.A. Ryan, W.B. Lee and M.W. Pietrzyk *et al.*, 2013. Quantitative measures confirm the inverse relationship between lesion spiculation and detection of breast masses. *Acad. Radiol.*, 20: 576-580. DOI: 10.1016/j.acra.2012.12.010
- Rocha, S.V.D., G. Braz Junior, A.C. Silva and A.C.D. Paiva, 2014. Texture analysis of masses in digitized mammograms using Gleason and Menhinick diversity indexes. *Rev. Brasileira Eng. Bioméd.*, 30: 27-34. DOI: 10.1016/j.acra.2012.12.010
- Rojas Domínguez, A. and A.K. Nandi, 2008. Detection of masses in mammograms via statistically based enhancement, multilevel-thresholding segmentation and region selection. *Computerized Med. Imag. Graph.*, 32: 304-315. DOI: 10.1016/j.compmedimag.2008.01.006
- Rose, D.C., I. Arel, T.P. Karnowski and V.C. Paquit, 2010. Applying deep-layered clustering to mammography image analytics. Proceedings of the Biomedical Sciences and Engineering Conference, May 25-26, IEEE Xplore Press, Oak Ridge, TN., pp: 1-4. DOI: 10.1109/BSEC.2010.5510827
- Wang, Y., J. Li and X. Gao, 2014. Latent feature mining of spatial and marginal characteristics for mammographic mass classification. *Neurocomputing*, 144: 107-118. DOI: 10.1016/j.neucom.2013.11.050
- Yadav, A.R., R. Anand, M. Dewal and S. Gupta, 2015. Multiresolution local binary pattern variants based texture feature extraction techniques for efficient classification of microscopic images of hardwood species. *Applied Soft Comput.*, 32: 101-112. DOI: 10.1016/j.asoc.2015.03.039
- Zhong, S., 2005. Efficient online spherical k-means clustering. Proceedings of the IEEE International Joint Conference on Neural Networks, Jul. 31-Aug. 4, IEEE Xplore Press, pp: 3180-3185. DOI: 10.1109/IJCNN.2005.1556436