© 2010 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

A Classifier with Clustered Sub Classes for the Classification of Suspicious Areas in Digital Mammograms

Peter Mc Leod and Brijesh Verma, Member, IEEE

Abstract—This paper presents a novel methodology for the classification of suspicious areas in digital mammograms. The methodology is based on the fusion of clustered sub classes with various intelligent classifiers. A number of classifiers have been incorporated into the proposed methodology and evaluated on the well known benchmark digital database of screening mammography (DDSM). The results in the form of overall classification accuracies, TP, TN, FP and FN have been analyzed, compared and presented. The results of all four tested classifiers with clustered sub classes on the DDSM benchmark database show that the proposed methodology can significantly improve the accuracy and reduce the false positive rate.

I. INTRODUCTION

 $\mathbf{B}_{\text{accounting}}$ for around 500,000 deaths per year worldwide [1]. In the European Union this represents 25% of all cancer cases and 19% of all cancer deaths [2]. Unfortunately the aetiology of breast cancer is not known which means that early detection is currently the most effective mechanism for reducing the mortality rate from breast cancer [1]. If detected early the survival rate is 98%; however if metastasis have spread to distant organs survival drops to 23% [3]. This means that early detection is essential for survival and the Gold standard for detection has been Screening Mammography which can reduce mortality by up to 41% [4]. Screening Mammography has however been shown to have several problems in that accuracy can vary with the skill of the radiologist and that fatigue has been shown to have an effect on the radiologists capacity to identify malignancies. CAD has been proposed as an adjunct to facilitate training and to provide a second opinion on the malignancy of an abnormality [2].

Various researchers have concluded that Computer Assisted Diagnosis (CAD) can increase the detection rate of early stage malignancies [5-10] which are more easily missed due to their small size; and in the case of masses low radiographic contrast. Other studies have concluded that CAD does not lead to an improvement in the detection and classification of breast abnormalities [11]. This has lead to many radiologists being skeptical about the efficacy of CAD systems, and that we can conclude that the main drawback to the adoption of CAD in clinical practice is the number of false positive classifications that can be obtained [12]. The reduction in the false positive rate must be achieved within the constraints of system resources (CPU cycles, memory, processing time) while maintaining a non invasive process. The research presented in this paper addresses this issue by incorporating clustered sub-classes into the classification process in order to improve classification accuracy and reduce the false positive rate.

The rest of this paper is organized as follows. Section two provides background information with a review of existing methods. Section three provides an overview of the proposed technique while section four details the experimental results obtained. Section five provides a discussion and analysis of these results while section six details the conclusions and future research.

II. BACKGROUND

Various methodologies have been proposed to address the issue of variable classification accuracy for the classification of suspicious areas in digital mammograms. The myriad number of solutions that have been examined reflect the complexity involved in optimizing this problem. Complicating factors include [13]:

- Mammograms are a two dimensional representation of a three dimensional structure.
- Mass type anomalies are often difficult to detect as they have a similar spatial density to surrounding breast tissue (Figure 1).
- Breast tissue is a complex structure which has features that can be easily mistaken for anomalies (Figure 1).
- Malignant anomalies are not easy to determine as they are not consistent in their morphological characteristics; making it almost impossible to elicit a set of rules that facilitate classification.
- Anomalies may be occluded.

Manuscript received February 4, 2010.

P. Mc Leod is in the School of Computing Sciences at CQUniversity, Bruce Highway, North Rockhampton, Queensland 4702, Australia (phone +617 46 393-633; e-mail McLeod.Ptr@gmail.com).

B. Verma is in the School of Computing Sciences at CQUniversity, Bruce Highway, North Rockhampton, Queensland 4702, Australia (phone +617 49 309-058; e-mail b.verma@cqu.edu.au).



Fig. 1. Left craniocaudal view of a mammogram with a tumor bounded by a rectangle. This tumor is readily identifiable however others are not so easily seen. The similar contrast to surrounding tissue should be noted. Image sourced from Varela et al. 2007 [16].

The techniques that have been employed to perform classification include artificial neural networks [14], support vector machines [15], statistical classifiers, rule based approaches, radial basis hybrid networks [16], genetic algorithms [16] and hybrid style networks.

Varela [17] utilized the gray level and morphological features to train a neural network to achieve an Area under the Receiver Operating Curve (Az or AUC) of 0.90 ± 0.02 . While Sahiner [14] utilized a genetic algorithm to determine the most discriminating textual and morphological features which were fed into a neural network and linear classifier (LDA). Chang [19] however used a likelihood function to compare the gray level and shape characteristics of a Regions of Interest (ROI) against a database of classified This was then used to compute a likelihood masses. measure. Tourassi [19] undertook a similar mechanism to classify ROI which measured the difference between the ROI being classified and all the ROIs in the database. The ROI being classified is then labeled as belonging to the class with the closest similarity measurement. A problem associated with the approach of Tourassi however is that it needs to be recalculated for each new element that is added. Meanwhile Georgiou et al. [15] utilized shape features together with the C-Support Vector Machine (C-SVM) to obtain 91.54% classification accuracy on masses. A custom dataset was utilized for their research where they were determining the shape features that resulted in the highest classification accuracies. Verma [20] however added an additional neuron for benign and malignant classes to the hidden layer of a Multi Layer Perceptron (MLP) style neural network to achieve 100% training accuracy and 94% classification accuracy on masses from the DDSM. Verma also replaced the traditional gradient descent mechanism of the network with the incorporation of least squares to avoid network paralysis. Mazurowski et al. [21] achieved an Az of 0.907 ± 0.024 with the use of a genetic algorithm to produce

an optimization framework for determining the importance of each image as it relates to the diagnostic process in a case based system. Meanwhile Padayachee [22] compared one mammographic view to another (cranio-caudal and mediolateral oblique) in order to confirm the presence of a suspicious feature. This was achieved by a method which evaluated the similarity of the views. The measures took into account AUC and the contrast between the area of the matched region and the background of the similarity map. The results of this process were quite successful in that an accuracy of AUC of 0.96 ± 0.05 was achieved. However only 68 mammograms were utilized in this study and the algorithms are very time consuming to come to a classification.

III. PROPOSED TECHNIQUE

Medical image classification has typically been a binary classification dilemma. In Digital Mammography an anomaly is determined to be either benign or malignant.

However such a black and white classification scheme is a model that has been utilized to simplify the problem domain



Fig. 2. Proposed research methodology, where the clustering step and target class assignment (shaded areas), i.e. assigning output subclasses as the output of the classifier are additional steps to the traditional methodology.

and is one that may explain why the rate of incorrect classifications can occur with a classifier. The similarities that exist between benign and malignant anomalies is a consequence of too large a grouping as the physiological / textual features between the two groups can be very similar. An expansion of the classification scheme to recognize that sub-populations exist for these two groups in theory can lead to improved classification accuracy. The logic is that by expanding our groupings we are minimizing the variance within each group and maximizing the differences between others, thus facilitating successful classification. In this research the determination of these sub classes is undertaken through clustering. Figure 2 provides a general format that is utilized in mammographic research. The shaded boxes in Figure 2 show the proposed steps that modify this general methodology. Details about the processes involved in the research methodology are detailed below.

A. Mammograms

The mammographic images that have been utilized for this research comprise a dataset of 200 ROI made up of mass type anomalies with equal numbers of benign and malignant samples that have been obtained from the Digital Database of Screening Mammography [23]. Being a publicly available database (http://marathon.csee.usf.edu/Mammography/Database.html) that contains a large number (2600 plus) of high quality images together with pertinent case information it is a premier resource for those researchers who wish to utilize a dataset that can facilitate comparison in their research. For this research 10 fold cross validation was utilized.

B. Image Segmentation

The processing of large high quality images is a computationally expensive process which could lead to inaccurate results as interference from surrounding regions could occur. By isolating the anomalous tissues we can process only those regions which are relevant. The process of extracting only the relevant region is known as image segmentation or Area extraction. Images from the DDSM [23] come with an ".overlay" file that provides a chain code which delineates the boundary of the anomaly together with a surrounding area that facilitates anomaly extraction.

C. Feature Extraction

Traditionally morphological (e.g. Shape) and gray level spatial features have been utilized by radiologists in order to classify anomalies as either malignant or benign [24]. It has also been demonstrated that these types of features are better discriminates for CAD systems [25-27]. These findings are utilized in choosing six features with four descriptor features from the BI-RADS reporting lexicon [28] together with patient age and a subtlety value [23] for performing the classification function. The four BI-RADS features are:

- Mass Density
- Mass Shape

- Mass Margin
- Abnormality Assessment Rank

D. Clustering

Clustering partitions a set of data by grouping data in ddimension feature space with a view to maximizing the similarity within a cluster compared to the dissimilar data that forms other clusters. This means that each cluster should be relatively homogenous. By assuming that only benign and malignant classes exist in medical image processing an abstraction (model) has been assumed which leads to only two classes which are more than likely not homogenous which produces sub-optimal training of a classifier. This can be seen in Figure 3 where grouping one represents unclassified anomalies. Such a grouping is unsatisfactory for diagnostic and treatment purposes. Grouping two represents the normal level of abstraction where two classes are represented. The split between Benign and Malignant facilitates classification and subsequent treatment. The establishment of only two classes however does help to facilitate a conceptual understanding of the problem domain. Establishing a more representative model which encapsulates the richness of the images acknowledges the existence of more than two classes. Taking the level of abstraction further we can recognize subclasses within our benigna and malignant major classes. This process is already undertaken by Doctors and Radiologists in their classification process (i.e. Ductal Cancer in Situ, Lobular Carcinoma in Situ, Metaplastic Carcinoma etc) however this process has not been extended widely to intelligent classifiers. By clustering the benign classes separately we arrive at a number of target sub-classes (e.g. B_1 , B_2 ,... B_n). The process is repeated for our malignant class (e.g. $M_1, M_2, \dots M_n$) which are our target subclasses for our training and classification processes. The same major classes exist for the purposes of our classification goal but we recognize the existence of subclasses in our abstraction for the purpose of training the The recognition of subclasses is shown in classifier. grouping three in Figure 3.

The nature of a benign or malignant class to fall into subclasses when the benign and malignant classes are clustered separately substantiates the existence of these subclasses. These subclasses represent the target output class for our classifier during the training and testing processes. For the purpose of this research the k-means clustering algorithm has been utilized due to its simplicity as well as the fact that it has been widely used in this field. However other clustering algorithms such as self organising map can be substituted.



Fig. 3. Breast Masses at different abstraction levels. Grouping one shows a collection of both benign and malignant masses without being categorized. Moving to grouping two we see the normal level of abstraction that is applied by differentiating between benign and malignant classes. Grouping three represents a higher level of abstraction that encompasses the richness of the condition and facilitates a cleaner mapping between features and classification.

E. Classifier

The proposed methodology has been designed to work with intelligent classifiers that can support multiple output classes. This is necessary since the number of output target classes has increased due to the introduction of target subclasses. However this should provide the advantage that the input feature set should map more appropriately to the output class as either a malignant or benign sub class. In theory this should provide improved training to the classifier resulting in improved classification accuracy (compared to a binary only classification mechanism). This mechanism for improving classification accuracy has been chosen as it is a mechanism that provides the following advantages:

- Classifier independent and transferrable to other classifiers.
- Can be transferred to other problem domains and datasets.
- Is conceptually easy to understand and model.
- Generally does not add greatly to the computational, memory requirements or complexity of the classifier.

Thus such a mechanism can potentially have a profound effect on the classification accuracy and is a potentially transferrable mechanism for classifier optimization.

Since historically the classification results in CAD systems have been binary outputs (benign or malignant) our output test results have been expressed likewise. Although it is possible to express classification accuracy for each subclass, this has not been done as the number of sub-classes varies according to the number of clusters. Potentially this would also impact on the comparability of results without adding any benefit as the composition of each cluster is different depending on the dataset and number of clusters involved in creating the sub-classes.

In this research a Radial Basis Function Network, Support Vector Machine, Multi Layer Perceptron and a Simple Logistics classifiers have been utilized.

IV. RESULTS

Different rounds of experiments were conducted in order to evaluate the efficacy of the proposed technique. All experiments were performed on a dataset consisting of 200 ROIs from the DDSM [23] with an equal number of benign and malignant abnormalities. Experiments were performed using ten fold cross validation. For the Radial Basis Function Network, Support Vector Machine, Multi Layer Perceptron experiments and the Simple Logistics classifiers we used the Weka Software (version 3.7) [29].

The results of the experiments for the Radial Basis Network are tabulated in Table I. Table II provides details on the Multi Layer Perceptron and Table III provides details on Support Vector Machine performance. Table IV provides details on the performance of the Simple Logistics classifier.

TABLE I Performance of radial basis function network					
Clusters ^a	Accuracy	Sensitivity	Specificity		
1	86.0%	87.5%	84.6%		
2	85.0%	85.7%	84.3%		
3	89.0%	88.2%	89.8%		
4	86.5%	87.6%	85.4%		
5	84.0%	84.0%	84.0%		
6	83.0%	83.0%	83.0%		
10	83.5%	83.8%	83.2%		
20	83.5%	83.8%	83.2%		

^aClusters refer to the number of clusters per class (benign) or malignant). Thus a standard binary classification is 1 cluster per class.

 TABLE II

 PERFORMANCE OF MULTI LAYER PERCEPTRON CLASSIFIER

Clusters ^a	Accuracy	Sensitivity	Specificity
1	85.0%	83.7%	86.5%
2	83.5%	80.7%	86.8%
3	87.0%	83.6%	91.1%
5	87.0%	84.9%	89.4%
6	85.5%	85.1%	85.9%
10	85.0%	83.7%	86.5%
20	88.5%	88.1%	88.9%

^aClusters refer to the number of clusters per class (benign) or malignant). Thus a standard binary classification is 1 cluster per class.

TABLE III Performance of SVM Classifier

Clusters ^a	filter	Accuracy	Sensitivity	Specificity	
		RBF K	ernel		
1	normalize	83.5%	83.8%	83.2%	
1	standardize	87.0%	87.0%	87.0%	
2	normalize	84.0%	86.2%	82.1%	
2	standardize	85.5%	86.6%	84.5%	
3	normalize	84.0%	85.4%	82.7%	
3	standardize	87.5%	89.5%	85.7%	
5	normalize	85.5%	85.9%	85.1%	
5	standardize	87.5%	87.9%	87.1%	
PolyKernel					
1	normalize	85.5%	85.9%	85.1%	
1	standardize	86.5%	87.6%	85.4%	
2	normalize	84.5%	86.3%	82.9%	
2	standardize	82.0%	83.3%	80.8%	
3	normalize	85.5%	88.2%	83.2%	
3	standardize	89.5%	89.9%	89.1%	
5	normalize	87.0%	87.8%	86.3%	
5	standardize	88.5%	88.1%	88.9%	

^aClusters refer to the number of clusters per class (benign) or malignant). Thus a standard binary classification is 1 cluster per class.

TABLE IV Performance of simple logistics Network

Clusters ^a	Accuracy	Sensitivity	Specificity
1	84.5%	85.6%	83.5%
2	84.0%	84.7%	83.3%
3	87.5%	87.1%	87.9%
4	87.0%	89.4%	84.9%
5	86.0%	90.0%	82.7%
6	85.0%	84.3%	85.7%
10	86.5%	86.9%	86.1%
20	86.0%	86.0%	86.0%

^aClusters refer to the number of clusters per class (benign) or malignant). Thus a standard binary classification is 1 cluster per class.

Details of the True Positive (TP). True Negative (TN), False Positive (FP) and False Negative (FN) used to calculate the sensitivity and specificity values are tabulated in tables V to VIII.

TABLE V True Positive, Negative and False Positive, Negative values of Radial Basis Function Network

111	010 01 1	CIDINE DI 1010 I	enemon	THET WORK	
Clusters ^a	ТР	TN	FN	FP	
1	84	88	12	16	
2	84	86	14	16	
3	90	88	12	10	
4	85	88	12	15	
5	84	84	16	16	
6	83	83	17	17	
10	83	84	16	17	
20	83	84	16	17	

^aClusters refer to the number of clusters per class (benign) or malignant). Thus a standard binary classification is 1 cluster per class.

TABLE VI True Positive, Negative and False Positive, Negative values of Multi layer perceptron network

Clusters ^a	ТР	TN	FN	FP	
1	87	83	17	13	
2	88	79	21	12	
3	92	82	18	8	
5	90	84	16	10	
6	86	85	15	14	
10	87	83	17	13	
20	89	88	12	11	

^aClusters refer to the number of clusters per class (benign) or malignant). Thus a standard binary classification is 1 cluster per class.

TABLE VII TRUE POSITIVE, NEGATIVE AND FALSE POSITIVE, NEGATIVE VALUES OF SUPPORT VECTOR MACHINE CLASSIFIER

Clusters ^a	Filter	TP	TN	FN	FP			
	RBF Kernel							
1	normalize	83	84	16	17			
1	standardize	87	87	13	13			
2	normalize	81	87	13	19			
2	standardize	84	87	13	16			
3	normalize	82	86	14	18			
3	standardize	85	90	10	15			
5	normalize	85	86	14	15			
5	standardize	87	88	12	13			
		PolyK	lernel					
1	normalize	82	89	11	18			
1	standardize	85	88	12	15			
2	normalize	82	87	13	18			
2	standardize	80	84	16	20			
3	normalize	82	89	11	18			
3	standardize	89	90	10	11			
5	normalize	86	88	12	14			
5	standardize	89	88	12	11			

^aClusters refer to the number of clusters per class (benign) or malignant). Thus a standard binary classification is 1 cluster per class.

TABLE VIII TRUE POSITIVE, NEGATIVE AND FALSE POSITIVE, NEGATIVE VALUES OF SIMPLE LOGISTICS NETWORK

Clusters ^a	ТР	TN	FN	FP	
1	83	86	14	17	
2	83	85	15	17	
3	88	87	13	12	
4	84	90	10	16	
5	81	91	9	19	
6	86	84	16	14	
10	86	87	13	14	
20	86	86	14	14	

^aClusters refer to the number of clusters per class (benign) or malignant). Thus a standard binary classification is 1 cluster per class.

In addition to the results in tables V to VIII a graph of Sensitivity versus the False Positive Rate (ROC) was generated for each classifier. The ROC curves are presented below in figures IV to VIII below.



Fig. 4. ROC Curve for Radial Basis Function Network (RBF).



Fig. 5. ROC Curve Multi Layer Perceptron Network (MLP).



Fig. 6. ROC Curve for Support Vector Machine (RBFKernel).



Fig. 7. ROC Curve for Support Vector Machine (PolyKernel).



Fig. 8. ROC Curve for Simple Logistics Classifier.

V. DISCUSSION

The experimental results contrast the performance of the proposed research technique against a base line binary classification. As discussed in the background various researchers have obtained high areas under the ROC curve (Az), or high sensitivity values. However comparing the performance of one CAD system to another where different databases and or different test cases are used is not an effective measure of algorithmic performance [30]. Unfortunately this means that a full quantitative comparison with other researchers is not possible due to differences between one dataset and another. In this research however the aim is to provide a mechanism for improving the classification accuracy of an existing classifier. Accordingly a more appropriate measure of performance is the degree to which the base line classification performance compares with the performance of the classifier where sub-classes through clustering have been introduced.

Due to the differences in making an incorrect diagnosis

(diagnosing a malignant tumor as benign can be fatal whereas diagnosing a benign as malignant may result in patient stress and an unnecessary operation) several ratios are calculated. Sensitivity is the True Positives / (True Positives + False Negatives) and represents the probability of correctly detecting cancer. Specificity is the True Negatives / (True Negatives + False Positives) and represents the likelihood of correctly not being diagnosed as having cancer. Accordingly we have reported the sensitivity and specificity results in tables one to four. In this case the performance of the clustered classifier for all classifiers has increased in comparison to the base line (binary) classifier. The improvement for each classifier is listed below in table IX.

TABLE IX Performance improvement of target sub-class Classifier over baseline (binary) classifier

Classifier	Accuracy	Sensitivity	Specificity
RBF ^a	+3.0%	+0.7%	+5.2%
SVM ^b	+2.0%	+2.1%	+1.9%
SVM ^c	+0.5%	+2.5%	-1.3%
SVM ^d	+1.5%	+1.9%	+1.2%
SVM ^e	+3.0%	+2.3%	+3.7%
MLP ^f	+3.5	+4.4	+2.4
${ m SL}^{ m g}$	+3.0	+1.5	+4.4

^a3 Clusters.

^b5 Clusters. RBF Kernel, normalize filter

°3 Clusters. RBF Kernel, standardize filter

^d5 Clusters. PolyKernel, normalize filter

^e3 Clusters. PolyKernel, standardize filter

^f20 Clusters ^g3 Clusters

The ability of the sub class targets generated through clustering to improve the accuracy of all the classifiers demonstrates the effectiveness of this technique as a boosting mechanism for improving the performance of a multi-class capable classifier. The improvement in performance of the classifier substantiates the existence of sub-classes for the benign and malignant classes. Although this technique has not been (to our knowledge) tried in other problem domains the technique should be effective in other problem areas. This technique however does rely on the determination of an optimal number of sub-classes to provide optimal performance. Since the optimal number of clusters can vary between classifiers this needs to be determined through experimentation.

Accordingly the incorporation of sub-classes provides a mechanism for improving the operation of an intelligent classifier.

VI. CONCLUSIONS

The research presented in this paper has put forward an

approach for improving the classification of suspicious areas in digital mammograms by recognizing that the standard binary classification solution is a potential abstraction that hinders the efficacy of classifiers resulting in sub-optimal training and classification. The approach was evaluated on a benchmark database by incorporating a number of classifiers with clustered sub classes. In our experiments, the incorporation of sub-class targets as outputs for the classifiers resulted in a performance improvement for all tested classifiers. The overall classification accuracy has significantly improved and the false positive rate has dropped when compared to the base line approach. In our future research we would like to conduct more experimental analysis to find an optimal number of sub-classes and evaluate the impact of clustering and learning techniques used in the proposed approach.

REFERENCES

- J. R. Benson, I. Jatoi, M. Keisch, F. J. Esteva, A. Makris, and V. C. Jordan, "Early breast cancer," *The Lancet*, April 2009, vol. 373, pp. 1463-1479.
- [2] X. Llado, A. Oliver, J. Freixenet, R. Marti, and J. Marti, "A textural approach for mass false positive reduction in mammography," *Computerized Medical Imaging and Graphics*, vol. 33, pp. 415-422, 2009.
- [3] American Cancer Society: Breast Cancer: facts and figures, 2009-2010, Atlanta, *American Cancer Society, Inc.*, 2009.
- [4] D. Roder, N. Houssami, G. Farshid, G. Gill, C. Luke, P. Downey, K. Beckmann, P. Iosfidis, L. Grieve, and L. Williamson, "Population screening and intensity of screening are associated with reduced breast cancer mortality: evidence of efficacy of mammography screening in Australia," *Breast Cancer Res Treat*, vol. 108, pp. 409-416, 2008, doi: 10.1007/s10549-007-9609-5.
- [5] R. Brem, "Clinical Versus research Approach to Breast Cancer Detection with CAD: Where Are We Now?," *American Journal of Roentology*, vol. 188, pp. 234-235, 2007.
- [6] T. Freer, and M. Ulissey, "Screening mammography with computeraided detection: prospective study of 12,860 patients in a community breast center," *Radiology*, vol. 220, pp.781-786, 2001.
- [7] R. Birdwell, P. Bandodkar, and D. Ikeda,"Computer-aided detection with screening mammography in a University hospital setting" *Radiology*, vol. 236, pp. 451-457, 2005.
- [8] J. Dean, and C. Ilvento, "Improved cancer detection with diagnostic and screening mammography: prospective study of 104 cancers." American Journal of Roentology, vol. 187, pp. 20-28, 2006.
- [9] M. Morton, D. Whaley, K. Brandt, and K. Amrami, "Screening mammograms: interpretation with computer-aided detection – prospective evaluation," *Radiology*, vol. 239, pp. 204-212, 2006.
- [10] A. Lauria, R. Palmiero, G. Forni, P. Cerello, B. Golosio, F. Fauci, R. Magro, G. Raso, S. Tangaro, and P. Indovina, "The CALMA system: an artificial neural network method for detecting masses and microcalcifications in digitized mammograms," *Nuclear Instruments and Methods in Physics Research*, vol. 518, 99. 391-393, 2004.
- [11] J. J. Fenton, S. H. Taplin, P. A. Carney, L. Abraham, E. A. Sickles, C. D'Orsi, E. A. Berns, G. Cutter, R. E. Hendrick, W. E. Barlow, and J. G. Elmore, "Influence of computer-aided detection on performance of screening mammography," *New England Journal of Medicine*, vol. 356, no. 14, pp. 1399-1409, 2007.
- [12] R. Nishikawa, M. Kallergi, "Computer-aided detection, in its present form, is not an effective aid for screening mammography," *Medical Physics*, vol. 33, pp. 811-814, 2006.
- [13] M. Giger, "Update on the potential of computer-aided diagnosis for breast cancer", *Future Oncology*, vol. 6, no. 1, pp. 1-4, 2010. doi: 10.2217/FON.09.154.
- [14] B. Sahiner, H. P. Chan, D. Wei, N. Petrick, M. A. Helvie, D. D. Adler, "Image feature selection by a genetic algorithm: Application to

classification of mass and normal breast tissue," *Medical Physics*, vol. 23, pp. 1671-1684, 1996.

- [15] H. Georgiou, M. Mavrofarakis, N. Dimitropoulos, D. Cavouras, S. Theodordis, "Multiscaled morphological features for the characterization of mammographic masses using statistical classification schemes," *Artificial Intelligence in Medicine*, vol. 41, pp. 39-55, 2007.
- [16] D. Manrique, and J. Rios, A. Rodriguez-Paton, "Evolutionary system for automatically constructing and adapting radial basis function networks," *Neurocomputing*, vol. 69, pp. 2268-2283, 2006.
- [17] C. Varela, P. G. Tahoces, A. J. Mendez, M. Souto, and J. J. Vidall, "Computerized detection of breast masses in digitized mammograms," *Comput Biol Med*, vol. 37, no. 2, pp. 214-226, 2007.
- [18] Y. H. Chang, L. A. Hardesty, C. M. Hakim, T. S. Chang, B. Zheng, and W. F. Good, "Knowledge-based computer-aided detection of masses on digitized mammograms: a preliminary assessment," *Med Phys*, vol. 28, no. 4, pp.455-461, 2001.
- [19] G. D. Tourassi, B. Harrawood, S. Singh, J. Y. Lo, and C. E. Floyd, "Evaluation of information-theoretic similarity measures for contentbased retrieval and detection of masses in mammograms," IEEE Trans Med Imag, vol. 34, no. 1 pp. 140-150, 2007.
- [20] B. Verma, "Novel network architecture and learning algorithm for the classification of mass abnormalities in digitized mammograms," Artificial Intelligence in Medicine, vol. 42, no. 1, pp. 67-79, 2008.
- [21] M. Mazurowski, M. P. Habas, J. Zurada and G. Tourassi, "Decision optimization of case-based computer-aided decision systems using genetic algorithms with application to mammography", *Physics in Medicine and Biology*, vol. 53, pp. 895-908, 2008.
- [22] J. Padayachee, M. Alport and W. Rae, "Mammographic CAD: Correlation of regions in ipsilateral views – a pilot study", *South African Journal of Radiology*, August 2009, pp. 48-54.
- [23] M. Heath, K. Bowyer, D. Kopans, R. Moore and P. Kegelmeyer, "The digital database for screening mammography," IWDM-2000, *Medical Physics Publishing*, 2001.
- [24] M. M. Eberl, C. H. Fox, S. B. Edge, C. A. Carter and M. C. Mahoney, "BI-RADS Classification for Management of Abnormal Mammograms," The Journal of the American Board of Family Medicine, vol. 19, pp. 161-164, 2006.
- [25] I. Christoyianni, A. Kotras, E. Dermatus, and G. Kokkinakis, "Computer aided diagnosis of breast cancer in digitized mammograms," Computerized Medical Imaging and Graphics, vol. 26. pp. 309-319, 2002.
- [26] H. Alto, R. Rangayyan, and J. Desautels, "Content-based retrieval and analysis of mammographic masses," Journal of Electronic Imaging, vol. 14, 023016, pp, 1-17, 2005.
- [27] T. Andre, and R. Rangayyan, "Classification of breast masses in mammograms using neural networks with shape, edge sharpness and texture features," Journal of Electronic Imaging, vol. 15, no. 1, 2006, viewed 12 January 2008, doi:10.1117/1.2178271.
- [28] American College of Radiology, "Breast imaging reporting and data system," *breast imaging atlas*, 4th edn, Reston, Va, American College of Radiology, 2003.
- [29] I. Witten, and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2nd edn., Morgan Faufmann, 2005.
- [30] R. M. Nishikawa, M. L. Giger, K. Doi, C. E. Metz, F. F. Yin, C. J. Vyborny, and R. A. Schmidt, "Effect of case selection on the performance of computer aided detection schemes," *Med Physics*, vol. 21. pp. 265-269, 1994.