Performance evaluation of clustering algorithms on microcalcifications as mammography findings

Emmanouil K. Ikonomakis, George M. Spyrou, Panos A. Ligomenides and Michael N. Vrahatis

***Abstract*— Breast cancer can be prevented with regular mammography screening. Yet, the incorporation of Computational Intelligence relies on training classifiers on a set of predefined Regions of Interest (ROIs). Data Clustering has been applied to address the problem of ROI detection, yet no extensive research has been carried out on which algorithm to utilize. This contribution focuses on microcalcification clustering as a Data Clustering application, giving insights concerning the performance of three main clustering algorithms.**

I. INTRODUCTION

Digital mammogram screening remains the most efficient way of early breast cancer detection. Radiologists utilize computer systems in an attempt to reduce the number of misdiagnosed cases. Such systems are divided into two categories. Computer Aided Detection Systems that aid the radiologist through the use of image processing tools (CADe systems) and Computer Aided Diagnosis Systems that provide a provisional diagnosis to the radiologist (CADx systems) [1]. Both system types employ image processing tools. However, microcalcifications and noise appear indistinguishable on an image. The reason is that microcalcifications, as well as noise, appear on a mammogram as small bright spots. Hence, the step of data preprocessing is essential for the success of a CADx system.

Current CADx systems employ Computational and Artificial Intelligence methods such as Artificial Neural Networks (ANN) [2] and Support Vector Machines (SVM) [3] to determine whether a given microcalcifications grouping depicts an actual cluster of them or simply noise signals. Some other approaches involve the usage of Fuzzy *C*-means or Possibilistic Fuzzy *C*-means (FCM and PFCM respectively). Finally, the usage of population based optimization is not new to this problem, since Genetic Algorithms (GA) have been already applied [4].

Yet these algorithms are trained on ROIs depicting clusters located either by processing the image using trivial clustering rules, or by segmenting the image into distinct or overlapping regions. In this contribution, probable microcalcifications are clustered together by applying more sophisticated clustering algorithms such as DBSCAN, Affinity Propagation (AP) as well as an in house developed clustering algorithm called Intelligent Unsupervised Clustering (IUC). Applying clustering algorithms to create useful clusters of probable microcalcifications is not new to this problem [5], [6]. However, in most cases it is combined with a classification step.

This contribution analyzes the preprocessing step in Section III, before describing the clustering algorithms used in Section IV as well as the evaluation measures in Section V. Thereafter, the experimental setup is described in Section VI as well as the results obtained. Finally, some concluding remarks are made in Section VII.

E.K. Ikonomakis and M.N. Vrahatis are with the Computational Intelligence Laboratory (CILab), Department of Mathematics, University of Patras, GR-26110 Patras, Greece, eki@math.upatras.gr, vrahatis@math.upatras.gr

G.M. Spyrou and P.A. Ligomenides are with the Biomedical Informatics Unit, Biomedical Research Foundation of the Academy of Athens, 4 Soranou Ephessiou, 11527 Athens, Greece gspyrou@bioacademy.gr, pligo@academyofathens.gr

II. DATASETS

The Digital Database for Screening Mammography (DDSM) [7], [8] was used. DDSM consists of 2620 cases. Each case consists from up to 4 mammograms. Resolution of these images varies from $42\mu m$ to $50\mu m$. Malignant cases as well as cases with benign findings are accompanied by a list of ROIs as defined by radiologists. Out of these cases only images with microcalcification findings were considered, since the aim of this contribution is to investigate some algorithmic schemes that are able to identify and cluster together, microcalcifications belonging to the same group.

Under the assumption that single microcalcifications or small groups of microcalcifications may be noise, only regions with density in a pre-given threshold are kept. More precisely, for each image the maximum microcalcification object density is evaluated, and points residing in areas with at least $p\%$ of the maximum density are preserved. p's values used are $40, 50, 60, 70, 80$ as well as 90. Out of these files, only files that contain microcalcifications are considered. Thus, 6 instances of each processed image were created. Each instance corresponds to a microcalcifications dataset. Hereafter, the dataset created for $p = 40$ will be addressed as p40 and so on. For the purposes of the study, the available number of images suitable for analysis was sufficiently large to allow for a hold-out validation method. Thus, we created a train set and an independent test set in order to achieve stronger validation. The sizes of the datasets are summarized in Table I.

$p\%$ values:	40	50	60	70	80	90
train	181	175	150	127	93	55
test	180	174	149	127	92	54

TABLE I

NUMBER OF IMAGES FOR TRAIN AND TEST DATASETS

978-1-4799-3163-7/13/$31.00 ©2013 IEEE

III. PREPROCESSING

At the preprocessing step each image is transformed from a grayscale image to a dataset represented in the Vector Space Model (VSM). In order to achieve this, the image is initially segmented by a 50×50 lattice. Each dimension is divided into 50 sections and hence the image is divided into a total of 2500 segments. For each image the breast boundary is determined and a filtering process is applied on the image region lying inside the breast boundary.

Each segment is determined whether it is part of the breast or depicts part of the remainder of the image and hence contains solely noise. This is achieved by using breast boundaries according to segmentation of the breast region as introduced by Ojala *et al.* [9]. The breast region segmentation method is a five step process, hereby shortly described.

The process of breast region segmentation starts by analyzing the histogram of a given image. Based on areas of high pixel intensity, a rough breast boundary is created. As a next step morphological filtering based on the common dilation and erosion operations is applied to the image. Hence, a more accurate boundary can be formed which still contains a large part of the pectoralis muscle tissue. By locating the narrowest line above and underneath the breast, the breast boundary is restrained to the actual breast. Finally, a Fast Fourier Transform is applied to smoothen the image. Once the breast boundary is determined, areas describing a part of the breast, are filtered by the process proposed in [10]. In short, the algorithm performs unsharp masking and block averaging, evaluating the resulting image's variance under a Noise model. The pixels that differentiate from noise pass from an adaptive filter as introduced in [11] by Lorenz to reveal the microcalcifications.

As a result of the aforementioned process, a binary image is created. Out of this image, a set of candidate microcalcifications is transformed to a $n \times 2$ matrix where each line contains the coordinates of the center of each potential microcalcification. In order to achieve this, the image is segmented into group of 8-connected neighborhoods of pixels. Such neighborhoods are dropped if they consist of less than 4 pixels as too small to represent a microcalcification. The remaining neighborhoods are considered as potential microcalcifications. Their center is calculated as the average x-coordinate and y-coordinate of each pixel belonging to a potential microcalcification.

IV. CLUSTERING ALGORITHMS

Three clustering algorithms, have been compared in order to cluster microcalcifications in groups that will enhance the performance of a CADx system.

A. Intelligent Unsupervised Clustering (IUC)

This algorithm is an evolutionary clustering algorithm introduced by Antzoulatos *et al.* [12]. Its concept is the combination of an evolutionary algorithm with a windowing technique and aims to discover the clusters of the dataset in an iterative process. It should be noted that, the proposed methodology can utilize any evolutionary algorithm as well as any computational swarm intelligence scheme. As in [12], the Differential Evolution (DE) is applied to optimize the Window Density Function (WDF), in order to find a region of high density.

WDF expresses the density of the region (orthogonal range hereafter called window) around a point. Its size is defined as α and since WDF is evaluated as the cardinality of a set, it holds that $\mathrm{WDF}_\alpha(z) \geqslant 0$. The discovery of high density regions of the datasets through the WDF is a maximization problem, hence $-$WDF is utilized by DE as the fitness function. The points that are included in this region can be effectively estimated using computational geometry methods [13]. Initially, DE locates the position z_0 in the image with the highest density. A window w of size α centered at z_0 is constructed. This window, which represents the core of the cluster, is enlarged as much as possible over one coordinate and, afterwards the next coordinate is considered. The points residing inside it are removed from the dataset and DE is reapplied to locate an additional cluster center. Once only noise remains in the dataset, this iterative process terminates. Finally, all detected clusters are merged.

In this contribution, IUC has been modified to rerun if less than 3 or more than 10 clusters are located. Each time IUC repeats the clustering process, it slightly modifies its parameters to adjust itself into finding more or less clusters than expected. No more than 10 reruns are allowed in order to avoid infinite loops, regardless of the number of clusters located.

B. DBSCAN

DBSCAN (Density Based Spatial Clustering of Applications with Noise) [14] relies on a density-based notion of clusters and it can deal with arbitrary shaped clusters in a single-scan mode. DBSCAN aims to group adjacent points into clusters based on local density criterion. This algorithm is closer to the notion of cluster as implied in microcalcification clustering. It is based on two parameters, $MinPts$ and eps. An eps-neighborhood is defined as the set $N_{eps}(p) = \{q \in X | d(p,q) \leqslant eps\}$. $MinPts$ is the lower boundary of the number of points in an eps-neighborhood.

DBSCAN starts by randomly selecting a point of the database. If this point q has less than $MinPts$ points in its eps-neighborhood, it is marked as noise and another point is visited. Else, all the density reachable points from the selected point are retrieved and clustered together. This process terminates when all points have been examined. The points that cannot be assigned to a cluster are considered as noise.

C. Affinity Propagation

Affinity Propagation (AP), introduced by Frey and Dueck [15], is an unsupervised iterative process of data clustering. Each iteration consists of a set of two "message" updates. Availability $a(i,j)$ is sent from point j to point i and represents point's j potential of becoming i's exemplar. On the other hand, Responsibility $r(i,j)$, which is sent from point i to point j, is the suitability of point j as i's exemplar.

A damping factor λ is used for the update of availabilities and responsibilities so that oscilations are avoided [15]. Once these values are updated, the exemplars are redefined based on each point's availability and responsibility as well as a predefined preference of choosing this point as an exemplar. The process terminates when changes in the messages fall below a threshold for a given number of iterations.

V. EVALUATION MEASURES

Before proceeding to describe the evaluation measures used in this contribution, some auxiliary notions must be introduced. Let R be a ROI or cluster considered as ground truth. Since a subset of the DDSM is used, these ROIs are provided by physicians. C is a cluster as resulted by the clustering algorithm. O_{ij} is the overlap of R_i and C_j. A set of parentheses is used to denote the respective area. E.g. (C_j) is the area of cluster j. Finally, a mammogram is considered as TP if at least one cluster in the mammogram is annotated as TP.

A. Standardized Evaluation Rule (SER)

This is one of the most common rules of identifying a cluster as an actual microcalcification cluster (TP). As mentioned in [16], "A group of detected signals is considered a TP cluster if a specified minimum number of the signals is found inside the area containing the true cluster". The usual choice for this minimum number of signals, e.g. probable microcalcifications, ranges from 1 to 5. In this contribution a cluster is considered as a TP cluster if at least 3 probable microcalcifications are located in O_{ij}.

B. Window Coverage

Window Coverage (WC) is the ratio of the area of the intersection between the ROI as denoted by the physician and the cluster resulted by the algorithm over the area of the aforementioned cluster. If a cluster has WC more than 0.5, then the cluster is denoted as True Positive (TP), otherwise it is denoted as False Positive (FP). A cluster with WC value more that 0.5 is a cluster with more that 50% of its area inside the ROI denoted by the physician.

This can be summarized in the following expression:

$$ImageClassification = \begin{cases} TP, & \text{if } \max\limits_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}} \left\{ \dfrac{(O_{ij})}{(C_j)} \right\} \geqslant 0.5 \\ FP, & \text{otherwise} \end{cases} \tag{1}$$

where n is the number of clusters in a mammogram and m is the number of ROIs (or ground truth clusters) in the given mammogram.

VI. EXPERIMENTAL RESULTS

A. Experimental Setup

Every algorithm ran on the images of the training sets in order to determine its parameters that maximize its performance given the evaluation described in Section V. Once these parameters where determined, the algorithms where applied on the test sets for their final evaluation. Since the algorithms showed an increased performance for the p50, p60 and p70 configurations, p40, p80 and p90 where not used for testing.

AP ran on every image of the train dataset, given a specific set of parameters. Specifically, the algorithm ran for a maximum of 1000 iterations and $\lambda = 0.9$. The exemplars' preference is based on the distribution of the similarity matrix. A set of runs was executed for a quantile with threshold q of the preference, where $0.1,\ldots,0.9$ quantiles where used. The highest performance in training was achieved for $q = 0.7, 0.8$ and 0.9. Therefore, these values where used for testing.

DBSCAN ran for every image in each train dataset, with $MinPts = 1,2,\ldots,20$ and $Eps = 20,30,\ldots,150$. Yet DBSCAN did not provide clusters for a number of combinations of parameter values. For IUC, DE's parameters as well as secondary IUC parameters where configured through the training phase. Therefore, population size was set to 1000 and epochs to 100, while the weighting factor and crossover constant where 0.6 and 0.8 respectively. α values where limited to $\{50,60,\ldots,250\}$.

B. Results

In table II the number of clusters found by the respective algorithms is summarized. AP for $q = 0.7$ had the lowest number of clusters. Since a mammogram rarely has many clusters, a small number of clusters ensures less false positive findings, while at the same time provides a more balanced dataset for cluster classification since it does not contain many false positives. DBSCAN achieves its values for radii (Eps values) of 140 for p50 and p70 and 150 for p60 and $MinPts$ values of 7 and 8.

In table III True Positive ratios are given based on SER and WC respectively. Ratios are evaluated on the number of converged images. As expected, the values for SER are higher as for WC, since SER enhances the performance of the algorithms [16]. AP's best values for WC is achieved for 0.9 regardless of the test set's p-value, while for SER, AP performance with $q = 0.9$ was slightly inferior than with q values 0.7 and 0.8. DBSCAN provided its best results for $Eps = 150$ and $MinPts = 1$ for p70, while $Eps \in \{110,120,\ldots,150\}$ combined with $MinPts \in \{1,2\}$ provided similar or equal to the best result for p50 and p60. Finally, IUC performed best when $\alpha \in \{200,\ldots,250\}$.

The results of the conducted experiments can be summarized in the following comments. AP clusters candidate microcalcifications in a mammogram consistently better than DBSCAN and IUC. Although DBSCAN can achieve under circumstances comparable results to those of AP, IUC lacks significantly in performance. AP's major drawback in comparison to DBSCAN and IUC is that AP clusters all candidate microcalcifications depicted in a mammogram. This contradicts with the observation that many of the candidate microcalcifications represent noise instead of actual microcalcifications. In addition, the entire mammography is segmented in regions to be considered for further investigation by radiologists. Noise can be left unclustered and be

Dataset		AP	DBSCAN	IUC
p50	min	6	0	1
	mean	22.01	2	3.313
	std	7.4655	0	1.542
	max	48	2	10
p60	min	4	0	1
	mean	14.13	2	2.86
	std	5.3064	0	1.133
	max	31	2	7
p70	min	4	0	1
	mean	9.46	2	2.28
	std	3.1008	0	0.5727
	max	17	2	5

TABLE II
NUMBER OF CLUSTERS

Dataset		AP	DBSCAN	IUC
p50	**SER**	**159**	147	93
	(%)	**(91.379%)**	(84.483%)	(53.45%)
	WC	*156*	140	70
	(%)	*(89.655%)*	(80.46%)	(40.23%)
p60	**SER**	**120**	101	62
	(%)	**(80.537%)**	(68.707%)	(44.3%)
	WC	*115*	97	57
	(%)	*(77.181%)*	(65.986%)	(42.3%)
p70	**SER**	*71*	56	44
	(%)	*(55.906%)*	(44.444%)	(34.65%)
	WC	**68**	54	38
	(%)	**(53.543%)**	(42.857%)	(29.92%)

TABLE III
TP IMAGES BASED ON WINDOW COVERAGE (WC – RULE 1) AS WELL AS SER.

removed from further consideration by using DBSCAN or IUC. Furthermore, AP results in a large number of clusters per image, while the number of clusters resulted by using DBSCAN varies depending on it's parameters values. IUC managed to locate at least one cluster in each image and overall managed to provide a low number of clusters, while providing a relatively constant performance.

VII. CONCLUSIONS AND FUTURE WORKS

This contribution focuses on clustering microcalcifications in a mammogram with more sophisticated methods beside trivial clustering rules as is accustomed. More precisely, AP is a state of the art distance-based clustering algorithm, DBSCAN is a well known density-based algorithm and IUC is an emerging evolutionary clustering algorithm addressing both issues arising in using evolutionary computation to cluster a dataset.

The results from the experiments conducted show that clustering microcalcifications to capture ROIs in the DDSM mammography database is still a open problem due to the relatively low number of images where ROIs where successfully identified. Therefore, clustering microcalcifications in the context of this work, is a problem still to be addressed. Evolutionary Clustering has proven its performance in various applications. The results of these contribution indicate that in this application, there is still much room for improvement. The modification of these algorithms will be addressed so that clustering may increase its performance in the context of the problem at hand.

More importantly, it must be investigated whether such clustering results are enough to increase the performance of a CADe system. In order to answer this question, a classifier must be trained by using the results of a clustering process and compared with a classifier trained on a training set created by conventional clustering rules.

REFERENCES

[1] R. M. Nishikawa, "Current status and future directions of computer-aided diagnosis in mammography," *Computerized medical imaging and graphics: the official journal of the Computerized Medical Imaging Society*, vol. 31, no. 4, pp. 224–235, 2007.

[2] L. Bocchi, G. Coppini, J. Nori, and G. Valli, "Detection of single and clustered microcalcifications in mammograms using fractals models and neural networks," *Medical Engineering & Physics*, vol. 26, pp. 303–312, 2004.

[3] A. Bazzani, A. Bevilacqua, D.Bollini, R. Brancaccio, R. Campanini, N. Lanconelli, A. Riccardi, and D. Romani, "An SVM classifier to separate false signals from microcalcifications in digital mammograms," *Physics in Medicine and Biology*, vol. 46, pp. 1651–1663, 2001.

[4] H. Soltanian-Zadeh, F. Rafiee-Rad, and S. Pourabdollah-Nejad D, "Comparison of multiwavelet, wavelet, haralik and shape features for microcalcification classification in mammograms," *Pattern Recognition*, vol. 37, pp. 1973–1986, 2004.

[5] C. Marrocco, M. Molinara, and F. Tortorella, "Algorithms for detecting clusters of microcalcifications in mammograms," in *ICIAP 2005, LNCS 3617* (F. Roli and S. Vitulano, eds.), pp. 884–891, Springer-Verlag Berlin Heidelberg, 2005.

[6] L. Zhang, R. Sankar, and W. Qian, "Advances in micro-calcification clusters detection in mammography," *Computers in Biology and Medicine*, vol. 32, no. 6, pp. 515–528, 2002.

[7] M. Health, K. Bowyer, D. Kopans, R. Moore, and W. P. Kegelmeyer, "The digital database for screening mammography," in *Proceedings of the Fifth International Workshop on Digital Mammography* (M. Yaffe, ed.), pp. 212–218, Medical Physics Publishing, 2001.

[8] M. Heath, K. Bowyer, D. Kopans, W. P. Kegelmeyer, R. Moore, K. Chang, and S. MunishKumaran, "Current status of the digital database for screening mammography," in *Digital Mammography, Proceedings of the Fourth International Workshop on Digital Mammography*, pp. 457–460, Kluwer Academic Publishers, 1998.

[9] T. Ojala, J. Näppi, and O. Nevalainen, "Accurate segmentation of the breast region from digitized mammograms," *Computerized Medical Imaging and Graphics*, vol. 25, no. 1, pp. 47–59, 2001.

[10] G. Spyrou, S. Kapsimalakou, A. Frigas, K. Koufopoulos, S. Vassilaros, and P. Ligomenides, ""Hippocrates-mst": a prototype for computer-aided microcalcification analysis and risk assessment for breast cancer," *Medical and Biological Engineering and Computing*, vol. 44, no. 11, pp. 1007–1015, 2006.

[11] H. Lorenz, G. M. Richter, M. Capaccioli, and G. Longo, "Adaptive filtering in astronomical image processing – part one – basic considerations and examples," *Astronomy and Astrophysics*, vol. 277, no. 1, pp. 321–330, 1993.

[12] G. S. Antzoulatos, E. K. Ikonomakis, and M. N. Vrahatis, "Efficient unsupervised clustering through intelligent optimization," in *IASTED 2009, Artificial Intelligence and Soft Computing (ASC 2009)*, 2009.

[13] P. Alevizos, B. Boutsinas, D. K. Tasoulis, and M. N. Vrahatis, "Improving the orthogonal range search *k*-windows algorithms," in *14th IEEE International Conference on Tools and Artificial Intelligence*, pp. 239–245, 2002.

[14] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining*, pp. 226–231, 1996.

[15] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, pp. 972–976, 2007.

[16] M. Kallergi, G. M. Carney, and J. Gaviria, "Evaluating the performance of detection algorithms in digital mammography," *Medical Physics*, vol. 26, no. 2, pp. 267–275, 1999.