ISSN:0975-9646

Arun kumar M.N et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (6), 2012,5346-5350

Building Accurate Classifier for the Classification of Microcalcification

Arun kumar M.N, H.S. Sheshadri

PET Research Center,

Karnataka,India

***Abstract*— The most common life threatening type of cancer affecting woman is breast cancer. Mammography is an effective screening tool for breast cancer. For mammogram the CAD system is like a spell checker. CAD systems use digital image processing techniques to improve the detection performance and efficiency of mammography screening. The two most common features that are associated with cancers are clusters of microcalcifications and masses. There are some reasons such as small size of microcalcification, less brightness of microcalcification compared to the background, and superimposition of microcalcification on textures make the detection of microcalcification difficult. This paper proposes a methodology for the classification of microcalcification in mammograms. An improved classifier that introduces balanced learning for the accurate classification for the classification of microcalcification is proposed as one of the main steps in the methodology. The experiments are conducted on the samples collected from well known MIAS database and outperforms other methods in the classification of microcalcification.**

***Keywords*— CAD System, Classification, Imbalanced data, Mammography, Microcalcification**

I. INTRODUCTION

Breast cancer is the second leading cause of cancer death among women in the 40-55 age groups [1-4]. Mammography is a screening tool for breast abnormalities detection. Mammography allows identification of tumour before being palpable. Mammography [5-10, 33-36] is a transmission planar X-ray image formed by diverging X-ray beam. Poor image quality and oversight are some of the factors that lead to missed detections of breast cancer. 15-25% of biopsy proven cancerous is not detected for various reasons such as technical problems and abnormalities that are not observable.

The diagnosis errors actually form the foundation of CAD [4, 8, 9, 12, 13] in Digital Mammography. CAD integrates diagnostic imaging with image processing, pattern recognition, computer science, and artificial intelligence technologies. Radiologist uses the output from the CAD as a "second opinion" in detecting abnormalities and makes the diagnostic decisions with much less probable errors. A mammogram image analyzer program [4, 8, 9, 11] would decrease the possibility of false detection of breast cancer by a radiologist Clusters of calcifications and masses are two most common features associated with cancer. Micro calcifications(MC) [16,20] are tiny deposits of calcium that appear as small bright spots in the mammogram and have higher inherent attenuation properties. Microcalcifications cannot be distinguished from the high-frequency noise because of their small size. For MC detection a large number of MC detection algorithms are developed. In recent years, several CAD systems that support MC detection have been deployed for clinical use.

Imbalanced data classification often arises in many practical applications. Many classification methods are developed but they assume evenly distribution of underlying training set. When the training set is a highly imbalanced distribution many approaches are faced with severe bias problem. There are many real-world problems faced with severe problem of learning for imbalanced class. The imbalanced data cause classifiers to perform poorly on the minority class. Due to the highly imbalanced nature of the data many existing classifier tend to misclassify the minority class. Classification of potential microcalcification can be thought of as an example for class imbalance problem as it is shown in the next coming sections.

The proposed scheme of the CAD system for microcalcifications, consists of many steps. A preliminary step deals with preprocessing. It removes the artifacts and noises. It also removes the pectoral muscle that appears as a bright smooth triangle in the left/right highest part of the image. As a second step, suspicious regions are isolated by a detection algorithm. Potential microcalcifications are detected in this step. Next step deals with feature extraction and a selection procedure is implemented to classify the potential microcalcifications to the real microcalcification. The proposed methodology is carried on mammographic images taken from MIAS.

The remaining part of this paper is organized as follows. Section 2 contains works related to the abnormality detection and classification in digital mammograms. In Section 3, methodology is explained. Experimental results are discussed in Section 4. Last section discusses the conclusion and future works.

II. PREVIOUS WORKS

There have been various approaches to the task of isolating the breast region and/or pectoral muscle segmentation in mammograms. Many researchers have developed the CAD for the automatic detection of masses and micro-calcification in a Digitized Mammograms.

Verma and Zakos [14] presented a system based on fuzzy-neural and feature extraction techniques for detecting and diagnosing micro calcification patterns in digital mammograms. A hierarchical learning approach which consists of a cascade of a binary classifier and a regression module is proposed by El Naqa et al [15] to optimize retrieval effectiveness and efficiency. K.Thangavel et. al. [17] proposed a Markov Random Field method hybrid with Ant Colony System, Genetic Algorithm and Back propagation Network. The Spatial Gray Level Dependency Method, Surrounding Region Dependence Method, Gray-

Level Run-Length Method and Gray Level Difference Method are used to extract the features from the segmented image. A three-layer Back propagation Neural Network classifier is used to classify the extracted features into benign or malignant. Oporto et.al. [18] used successfully sequences of Difference of Gaussian (DOG) filters to detect and classify microcalcification. DoG filter extracts all available points from the digital mammograms and use three selection methods to select the potential microcalcification.

Bocchi et. al. [19] used a fractal model to describe the mammographic image and analyzed the spatial arrangement to detect and classify the MC. Matched filter enhances microcalcifications and region growing algorithm, combined with a neural network classify the MC. De Santo et.al. proposed an approach [21] for the classification system based on a multiple expert systems. They argued that multiple expert systems significantly outperform single expert system. Nakayama et.al. [22] developed a filter bank based on the concept of the Hessian matrix for classification. In each Region Of Interest, eight features were extracted and employed Bayes discriminant function. Proposed method evaluated the detection performance using 600 mammograms. Liyang et.al. [23] investigated several machine learning methods for the classification of clustered microcalcification. They employed methods like support vector machine (SVM), kernel Fisher discriminant (KFD), relevance vector machine (RVM), and committee machines.

Many techniques for handling imbalanced data for binary class and multi class problems are developed.

Xue-wen et.al. developed a method based on SVM and backward pruning technique [24] for handling the imbalanced data classification. Proposed method used 1484 samples collected from SWISS-PROT and Hyperspectral/Polarimetric target detection. Proposed method outperforms SVM-based weighting method. Linear embedding algorithm incorporated on conventional SMOTE algorithm [25] was proposed by Juanjuan Wang. Three datasets are collected from several chest x-ray image databases. Sample size of three data sets 1180, 1530, and 1164 with minority classes of .038%, 0.029%, and 0.042% respectively collected for experiment. They compared the traditional SMOTE algorithm with the proposed method and it outperforms conventional SMOTE algorithm. The proposed algorithm uses Naïve Bayesian, K-NN and support vector Machine classifiers for classification. Chao Chen et.al. adopted two methods [26] and experimented on many imbalanced data sets like Oil, Mammography, Satimage, Hypothyroid, Euthyroid, KDD thrombin of size 937, 11183, 6435, 2520, 2640, and 2543 respectively. Percentage of Minority class is 4.4, 2.3, 9.7, 4.8, 9.1, and 7.6 respectively. Proposed two methods outperform SHRINK, SMOTE, AND SMOTEBoost. Kernel-Based Two-Class Classifier [27] using orthogonal forward selection is developed by Xia Hong. Developed method applied on Synthetic data, Pima Indian diabetes data, Haberman data, ADI data, Satimage data. Performance comparison with 1-NN, 3-NN, SVM, SUPANOVA, RIPPER, SMOTE, SMOTE-Boost, BRF, WRF shows effectiveness of their work. Yuchun et.al. [28] developed SVM-WEIGHT, GSVM-RU, SVM-SMOTE, and SVM-RANDU. Many imbalanced datasets like Oil, Mammography, Satimage, Abalone (19 vs. other), Abalone (9 vs. 18), Yeast (ME2 vs.other), Yeast (CYT vs. POX) are used for analysis. The proposed method compared with WRF, AdaCost, KBA, SMOTE-Boost. It is found that GSVM-RU outperforms all other classifier. Many other methods address the class imbalance problem effectively [29-32].

III. METHODOLOGY

The proposed methodology consists of preprocessing, segmentation, feature extraction and classification. The method does not treat the detection of microcalcification as simple blob detection.

A. Preprocessing

Preprocessing step removes the artifacts and noises. It extracts the pectoral muscle to reduce the processing area for the remaining stages in the detection and classification of microcalcification. This phase accepts a mammogram, opens the mammogram using a structuring element, and reconstructs the image. The difference image is thresholded with a suitable value which is experimentally obtained. Morphological operators and Sobel edge detector are applied to smooth irregularities and to detect the edges respectively. Breast contour detected is superimposed on original image. Pectoral muscle segmentation uses a wavelet decomposition of fourth level and edge detection using Canny filter. Refer our work [9] for more preprocessing details.

B. Segmentation

Segmentation is a three-step entropy thresholding algorithm and is the modified version of [37]. First step filter mammogram with a morphological white top hat for detecting the small bright points. This step enhances the visibility and detectability of microcalcifications. Second step accepts the filtered image and uses the ranked thresholds which is explained later to segment all the maximum number of points inside the mammogram. These points are transformed to the potential microcalcification based on some selection methods.

Detection of Potential microcalcification*: This step removes the background without reducing the microcalcifications by a top hat filtering, fixes optimal thresholds for segmenting the filtered image and thus separates the potential microcalcification. This step is explained below.*

Let i(x,y) be the digitized mammogram image IM of size NxN with L grey levels. Perform the following steps.

1. Filter the image IM by a top hat filter.
2. For all the grey levels of g from 1 to L of IM compute the image Entropy as explained in [37]. Store image entropy and the g used to compute it in the arrays IE and G respectively.
3. Rank the array IE (i.e., obtain the descending order of IE) and then rank the corresponding G array. Array G stores the optimal thresholds (grey levels) in the descending order.

To detect greater possible quantity of points the top 8 optimal thresholds are selected and used to segment the filtered image. When each of these thresholds is applied the newly obtained points are added and the repeated points are deleted. Two selection procedures are used to transform a point to potential microcalcification. Area is used as the first selection method to transform the detected points to potential microcalcification. Second selection method uses grey gradient to transform the detected points to potential microcalcification. After these selection methods many potential microcalcifications are obtained. With the available truth information in MIAS database and with the support of 3 radiologists these potential microcalcification were classified as microcalcification and not-microcalcification. From these potential microcalcification only 5% were microcalcifications and others not. These percentages indicate a class imbalance problem and we address this issue by using an improved classifier that introduces balanced learning for the accurate classification of potential microcalcifications into microcalcifications.

1) Feature extraction and Classification of potential microcalcification

In order to accurately classify the potential microcalcification to microcalcification a set of features are extracted. The following features are extracted. Average grey level, standard deviation of grey level, edge strength, background grey level, Foreground background ratio and difference, compactness, elongation, shape and invariant moment, second order histogram related features, Five features and three features related to contrast and related contrast respectively. These features are passed through a feature selection process and features that present high correlation with other features are removed. After the feature selection procedure the following features are selected. Absolute contrast, standard deviation of grey levels, Difference ratio, Area, compactness, entropy, angular second moment, inverse different moment, correlation, sum entropy.

We build the classifier using Class Confidence Weighted (CCW) kNN [38] as the traditional kNN uses only the prior information to estimate class labels, and when the data set is highly imbalanced it has suboptimal classification performance especially on the minority class. The merits of CCW kNN is proved in [38] and refer the same for more details.

Let (ai, bi) (i = 1, n) represent training data, ai $\in$Rm are feature vectors, m is the number of features bi $\in$ {class1, class2}. kNN algorithm finds group of prototypes from the training set that are the closest to a test instance at by a certain distance measure. When the k neighbours vary widely in their distances and nearest neighbours are reliable, the neighbours are weighted by the multiplicative-inverse (MI) or the additive-inverse (AI) of their distances [38].

MI: $b't = \text{argmax}_{c\in\{c1,c2\}\ ai\in\phi\,(at)} \sum T(bi = c) \cdot (1/dist(\mathbf{a}_t, \mathbf{a}_i))$

AI: $b't = \text{argmax}_{c\in\{c1,c2\}\ ai\in\phi\,(at)} \sum T(bi = c) \cdot (1 - (dist(\mathbf{a}_t, \mathbf{a}_i)/dist_{maximum})$

where $b't$ is a predicted label, $T(\cdot)$ is an indicator function that returns 1 if its condition is true and 0 otherwise, $\phi(\mathbf{a}_t)$ denotes the set of k training instances closest to $\mathbf{a}_t$, $dist(\mathbf{a}_t, \mathbf{a}_i)$ represents the distance between the test point $\mathbf{a}_t$ and a prototype $\mathbf{a}_i$, and $dist_{maximum}$ is the maximum possible distance between two training instances in the feature space which normalizes $dist(\mathbf{a}_t, \mathbf{a}_i)/dist_{maximum}$ to the range of [0,1]. CCW-kNN capture the probability of attributes values given a class label. CCW on a training instance i is defined as follows:

$wi^{CCW} = p(\mathbf{a}i|bi)$

where ai and bi represent the attribute vector and the class label of instances i. Now the classification rule integrated with CCW is: $b't = \text{argmax}_{c\in\{c1,c2\}\ ai\in\phi\,(at)} \sum T(bi = c) \cdot wi^{CCW}$

It was proved that CCW changes the bases of kNN rule from using priors to posteriors and handle effectively the class imbalance problem.

IV EXPERIMENTAL RESULTS

We used mammogram provided by the Mammographic Image Analysis Society (MIAS). The Mammographic Image Analysis Society database contains 322 images. All these images have resolutions of 50 microns/pixel and 200 microns/pixel. The truth data consists of the location of the abnormality and the radius of a circle which encloses it. Calcifications, circumscribed masses, spiculated masses, ill-defined masses, architectural distortions and asymmetry are the abnormalities represented in the MIAS database. Normal cases are also represented in the MIAS database. Out of 322 images, only 25 contain microcalcifications and out of these 25 images, 13 cases are reported as malignant and 12 as benign.

Our methodology receives a digital mammogram and processes it through the four stages: - preprocessing, segmentation, feature extraction, and classification. Preprocessing extracts the breast contour and removes the pectoral muscle. Figure 1 and 2 shows the successful result of preprocessing. Refer our work [9] for more results of contour extraction and intermediate results generated in pectoral muscle segmentation.

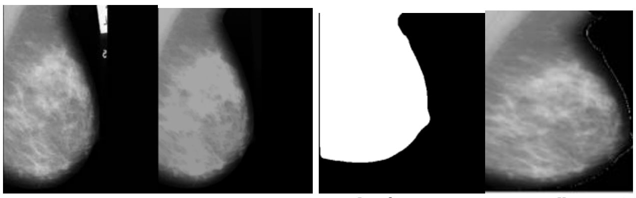
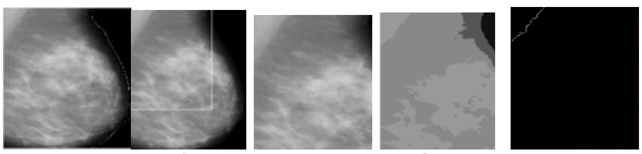


Fig. 1 Mammogram preprocessing results for MIAS image mdb016. (a). Original Mammogram; (b). Noise & Artifacts removal after filtering and morphological operation. (c). Binary Image; (d). Contour superimposed on original image



(a) (b) (c) (d) (e)

Fig. 2. Pectoral muscle identification results for MIAS image mdb016. (a).Breast contour superimposed on original image; (b). The region of interest containing the pectoral muscle; (c). Segmented area containing the pectoral muscle; (d). Wavelet decomposed image; (e). Pectoral muscle edge identified

As the first step in the segmentation phase to detect the potential microcalcification a top hat morphological filtering is used to remove the slow rate of variation of image intensity values. It enhances the image contrast. We used structuring element as square kernels of size 21x21 to get the best results. From all the grey levels from 1 to L we computed Image Entropy, stored these values in the array and ranked the array. Also top 8 optimal thresholds are computed as explained earlier. Top 8 optimal thresholds are used to segment the filtered image. Each of these thresholds is used to obtain all the possible points. Figure 3 shows an example of the application of top 8 thresholds to select the points in a ROI.

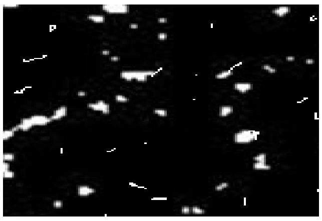


Fig. 3. Example of the application of top 8 thresholds to select the points in an image (ROI).

All the points obtained are passed through two selection methods. In our work pixel area of 1 to 99 pixel numbers are used as the first selection methods and all the points obtained by this method are passed through the second selection method. Grey gradient is used in second selection method to transform the detected points to potential microcalcification. For this method grey gradient is set in the range of 5 to 60 which is obtained after analyzing the mean grey gradient of point surrounding the real microcalcification. These potential microcalcifications obtained after the two selection methods are classified as microcalcification and not-microcalcification.

But the percentage of true microcalcification was only 5%. Since the dataset is imbalanced CCW-kNN (class confidence weights) kNN is used. The advantage of CCW is that it corrects the inherent bias to the majority class in kNN algorithm. To identify whether the detected obtained potential microcalcification corresponds to an individual microcalcification or not a set of features are extracted from the potential microcalcification. We extracted 10 features of potential microcalcification by removing the features that present high correlation with other features. We build a CCW- kNN for the accurate classification of potential microcalcification into microcalcification as this approach is more efficient. We have used the metric area under precision-recall curve (AUC-PR) to evaluate classifier performance. Figure 4 shows the image with microcalcifications detected.

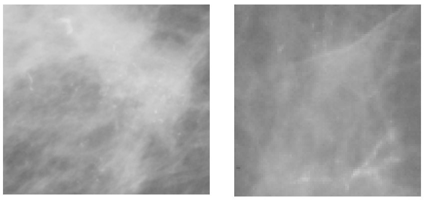


Fig. 4. A cluster of microcalcifications

V CONCLUSION AND FUTURE WORK

The most common life threatening type of cancer affecting woman is breast cancer. Mammography is an effective screening for breast cancer. Microcalcification clusters is a feature associated with cancer. Microcalcifications are tiny objects and the detection and classification of microcalcification is a challenging task. In this approach we made an attempt to develop a method for the detection and classification of microcalcification. For the detection of microcalcification our approach used ranked thresholds and built a CCW-kNN classifier for the accurate classification of potential microcalcification into microcalcification. This classifier effectively classifies the potential microcalcifications as it handles the imbalanced data effectively. We implemented our work and applied it to MIAS dataset. The detection step is time consuming but our approach increases the classifier performance.

REFERENCES

[1] Wei D, Chan H.P, Helvie M.A, Shahiner B, Petrick N, and Alder D.D, “Classification of Mass and Normal Breast Tissue on Digital Mammograms”, Med Physics, 22(9), 1501-13, Sept. 1995.

[2] Undrill P.E, Gupta R, and Henry S, “The Use of Texture Analysis and Boundary Refinement to Delineate Suspicious Masses in Mammography”, SPIE, Medical Imaging: Image processing, SPIE 2710, 301-310, 1996.

[3] American Cancer Society, “Cancer Facts and Figures”, GA: American Cancer Society, 2003.

[4] Mansees B S, “Evaluation of Breast Microcalcifications”, The Radiologic Clinics of North America, Breast Imaging, Vol. 33, 6, 1109-1121, Jan 1995 .

[5] B.Surendiran, A.Vadivel, Henry Selvaraj, “A Soft-Decision Approach for Microcalcification, Mass Identification from Digital Mammogram”, Proceedings of World Academy of Science, Engineering and Technology, volume 36 December, 2008.

[6] Roshan Dharshana Yapa and Koichi Harada, “Connected Component Labeling Algorithms for Gray-Scale Images and Evaluation of Performance using Digital Mammograms”, IJCSNS International Journal of Computer Science and Network Security, Vol.8 No.6, June 2008.

[7] Grim, Petr Somol, Michal Haindl, and Jan Dane, “Computer-Aided Evaluation of Screening Mammograms Based on Local Texture Models”, IEEE Transactions On Image Processing, Vol. 18, No. 4, April 2010.

[8] Jinshan Tang, Rangaraj M. Rangayyan, Jun Xu, Issam El Naqa, and Yongyi Yang, ”Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances, IEEE Transaction On Information Technology In Biomedicine, Vol. 13, No. 2, March 2010.

[9] Arun Kumar M.N , and H.S. Sheshadri “Breast Contour Extraction and Pectoral Muscle Segmentation in Digital Mammograms”, IJCSIS Vol. 9 No. 2, February 2011, (pp. 53-59).

[10] ICGST International Journal on Graphics, Vision and Image Processing (GVIP) GVIP Special Issue on Mammograms, 2007

[11] Stelios Halkiotis, John Mantas & Taxiarchis Botsis. University of Athens, Faculty of Nursing- Health Informatics Laboratory. “Computer-aided detection of clustered microcalcifications in digital mammograms”. 2000.

[12] B. Pataki, L. Lasztovicza “Extending Mammographic Microcalcification Detection Method to Cluster Characterization”, Department of Measurement and Information Systems, Budapest University of Technology and Economics, Budapest, Hungary, 2008.

[13] Ranadhir Ghosh, Moumita Ghosh, John Yearwood, “A Modular Framework for Multi category feature selection in Digital mammography”, ESANN'2004 proceedings - European Symposium on Artificial Neural Networks Bruges (Belgium), 28-30 April 2004. pp. 175-180.

[14] B. Verma, J. Zakos, “A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques”, IEEE Transaction on Biomedical. 5 (1) (2001) 46–54.

[15] I. El-Naqa, Y. Yang, M.N. Wernick, et al., “A support vector machine approach for detection of microcalcifications”, IEEE Transaction on Medical Imaging 21 (12) (2002) 1552–1563.

[16] R.N. Strickland and H.I Hahn, " Wavelet transforms for detecting microcalcifications in mammograms", IEEE Transactions on Medical Imaging, vol.15, no.2, pp.218-229, April 1993

[17] K.Thangavel and M.Karnan and R.Sivakumar and A. Kaja Mohideen, "Ant Colony System for Segmentation and Classification of Microcalcification in Mammograms" GVIP Special Issue on Mammograms, 2007, pp.1-12

[18] S. Oporto-Dıaz, R. R. Hernandez-Cisneros and H. Terashima-Marn, "Detection of microcalcification clusters in mammograms using a difference of optimized Gaussian filters", in Proceedings of the Second International Conference on Image Analysis and Recognition, ICIAR 2005, Toronto, pp. 998–1005, 2005.

[19] L. Bocchi, G. Coppini, J. Nori, G. Valli, "Detection of single and clustered microcalcifications in mammograms using fractals models and neural networks", Medical Engineering Physics. 26 (4) (2004) 303–312.

[20] Cancer Research UK: Key Facts on Breast Cancer, 2009. http://info.cancerresearchuk.org/cancerstats/types/breast/.

[21] M. De Santo, M. Molinara, F. Tortorella, M. Vento, "Automatic classification of clustered microcalcifications by a multiple expert system", Pattern Recogn. 36(7) (2003) 1467–1477.

[22] R. Nakayama, Y. Uchiyama, K. Yamamoto, et al., "Computer-aided diagnosis scheme using a filter bank for detection of microcalcification clusters in mammograms", IEEE Transaction on Medical Imaging, 53 (2) (2006) 273–283.

[23] Liyang Wei, Yongyi Yang, Robert M. Nishikawa, and Yulei Jiang, "A Study on Several Machine-Learning Methods for Classification of Malignant and Benign Clustered Microcalcifications", IEEE Transactions On Medical Imaging, Vol. 24, No. 3, March 2005.

[24] Xue-wen Chen, Byron Gerlach, and David Casasent, " Pruning support vectors for imbalanced data classification", Proceedings of International Joint Conference on Neural Networks, Montreal, Canada, July 31 - August 4, 2005.

[25] Juanjuan Wang, Mantao Xu, Hui Wang, Jiwu Zhang, "Classification of imbalanced data by using the SMOTE algorithm and locally linear embedding", ICSP2006 Proceedings.

[26] Chao Chen, Andy Liaw , and Leo Breiman, "Using random forest to learn imbalanced data", www.stat.berkeley.edu/tech-reports/666.pdf

[27] Xia Hong, "A kernel-based two-class classifier for imbalanced data sets", IEEE Transactions on Neural Networks, Vol. 18, NO. 1, January2007

[28] Yuchun Tang, Nitesh V. Chawla,, and Sven Krasser, "SVMs modeling for highly Imbalanced classification ", IEEE Transactions On Systems, MAN, and Cybernetics, Vol. 39, No. 1, FEBRUARY 2009

[29] David P. Williams, Vincent Myers, and Miranda Schatten Silvious, "Mine classification with imbalanced data ", IEEE Geosciences and Remote Sensing Letters, Vol. 6, No. 3, July 2009.

[30] Alberto Fernández, María José del Jesus, Francisco Herrera, "Hierarchical fuzzy rule based classification systems with genetic rule selection for imbalanced data-sets ", International Journal of Approximate Reasoning 50 (2009) 561–577.

[31] Chao-Ton Su and Yu-Hsiang Hsiao, "An evaluation of the robustness of MTS for imbalanced data ", IEEE Transactions on knowledge and data engineering, Vol. 19, No. 10, October 2007.

[32] Mu-Chen Chen a, Long-Sheng Chen, Chun-Chin Hsu, Wei-Rong Zeng , "An information granulation based data mining approach for classifying imbalanced data ", Elsevier, Information Sciences 178 (2008) 3214–3227.

[33] Sheshadri H.S, and Kandaswamy A, ―Detection of breast cancer tumor based on morphological watershed algorithm‖, GVIP, 2005, pp. 17-21.

[34] Sheshadri H.S, and Kandaswamy A, ―Experimental investigation on mammogram segmentation for early detection of breast cancer‖, Journal of Computerized Medical Imaging and Graphics, Elsevier science Vol. 31, 2005, 46-48.

[35] Sheshadri H.S. and Kandaswamy A, ―Mammogram image analysis using recursive watershed algorithm‖, National Journal of Technology, Vol. 1, No. 1, 2004, pp. 73-77.

[36] Sheshadri H.S, and Kandaswamy A, ―Computer aided decision system for early detection of breast cancer‖, Indian Journal of Medical research, Vol. 124, 2006, pp. 149-154.

[37] Moti Melloul and Leo Joskowicz, School of Computer Science and Engineering, The Hebrew University of Jerusalem, "Segmentation of microcalcification in X-ray mammograms using entropy thresholding " CARS 2002

[38] Wei Liu and Sanjay Chawla: Class Confidence Weighted kNN Algorithms for Imbalanced Data Sets. In proceedings of the 15th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2011), Part II, page 345–356.