

Smart Digital Mammographic Screening System for Bulk Image Processing

Dr. M. Duraipandian,

Head of Department, Department of Computer Science and Engineering,
Vivekanandha College of Technology for Women,
Namakal, India.
Email: svsduraipandian@gmail.com

Mr. R. Vinothkanna,

Department of ECE, Vivekanandha College of Technology for Women,
Namakal, India.
Email: rvinothkannaphd@gmail.com

Abstract: Treating breast cancer is easier at early stages. However, proper diagnosis is essential for this purpose. Mammography helps in early detection of cancer cells. Existence of masses, calcification and mammogram are the evidences that help radiologists in early cancer identification. This paper proposes a smart digital mammographic screening system for processing images in large volumes irrespective of the nature of images. Watershed segmentation is performed based on appropriate selection of internal and external markers using multiple threshold extended maxima transformations in this technique. Distinguishing between healthy breast tissue and masses can be performed efficiently using a two-stage classifier. Extreme Learning Machine based single layer feed forward network along with Bayesian classifier is used for reducing false positive areas. Feature vector with features like texture and contrast are calculated using these approaches. Digital Mammography Screening database (DMS) is created with 100 mammographic images for the purpose of evaluation. Further, online databases like Breast Cancer Database (BCDB) and BreakHis are also used for analysis. Overall sensitivity of the datasets using the Bayesian classifier and Extreme Learning Machine is found to be 85% and 90% respectively.

Keywords: Mammography, image processing, Bayesian classifier, extreme learning machine, watershed segmentation;

1. Introduction

In most developing and developed countries, the female population is largely affected by breast cancer. Death rate due to this issue is also considerably high [1]. According to a survey, over 10% of women in USA have a possibility of being diagnosed with breast cancer during their lifetime. The chances of curing breast cancer are high in asymptomatic patients when it is diagnosed at an early stage. A cost effective and quick acquisition scheme based on X-ray technology termed as mammography is an effective and reliable mechanism for early stage breast cancer diagnosis. For proper investigation, it is essential that the radiologists must identify the existence of masses, calcification, and visual lesions using the mammographic image [2]. Extensive experience and knowledge is essential for accurate evaluation of these mammographic images. The low contrast of image segments, variation in shape and sizes makes it a challenging process for mass detection. Masses may be with improper or well-defined borders, thorny, irregular or oval shaped. Identification of masses may be further complicated when glandular elements of the breast are involved. Mammograms contain similar imaging for both masses and healthy breast tissue. This may cause incorrect estimation in initial stages [3].

Computer-aided diagnostic systems (CAD) is used for is commonly used for analysis of breast images since recent decades [4]. Improved patient diagnosis is performed by radiologists with this decision support tool through visualization of patient condition and providing subjective metrics. Sensitivity of this technique can be further improved though obtaining opinion from a second physician. The specificity and sensitivity of these systems can be further improved despite the extended research. With CAD systems, acceptable number of false positives is involved while reducing oversight of small micro calcifications that can be easily missed [5]. Invasive cancer characterization and detection rate may be increased even for less clearer masses using CAD systems.

2. Related Work

Level set based and watershed based segmentation algorithms are used for developing a fully automatic mass segmentation algorithm [6]. The computational time is reduced by set levels while improving the

segmentation efficiency and segmentation processing time on combining the two segmentation schemes. Detection and mass segmentation can be performed using automated systems. The mammograms are analyzed for detection and segmentation of the masses using local statistical texture analysis and constraint region growing technique schemes for improving the distinct structure of breast tissues with the help of contrast enhancement [7]. Simple image processing schemes are used for analyzing mammograms and detecting cancerous tissues. Tumor detection is performed using least-variance and max-mean techniques. The boundary of region of interest (ROI) is extracted with image gradient technique on defining the tumor patch with morphological closing operation [8].

Several approaches on deep learning and machine learning schemes are published over the last few years [9]. These schemes offer improved results when tested on similar datasets. MIAS public database is used for testing the automated breast mass characterization scheme. Mass classification is performed from the grey-level co-occurrence matrix (GLCM) by extracting four texture features namely homogeneity, energy, correlation and contrast for segmenting the mass region using fuzzy C-means algorithm. Sparse representation classifier (SRC), back-propagation neural network (BPNN) and AdaBoost classifiers are used for classification [10]. Comparison is performed based on Matthews's correlation coefficient (MCC), negative predictive value (NPV), positive predictive value (PPV), specificity, sensitivity, and accuracy [11]. The accuracy of SRC is high when compared to BPNN and AdaBoost. MIAS and DSSM database are used in the automated classification and segmentation scheme. The internal structure is enhanced using a rough set and chain codes are used in the extraction of ROIs [12]. 32 features of ROI are extracted using a vector field convolution snake. Decision trees, support vector machine (SVM), random forest classifier, particle swarm optimization (PSO)-SVM and genetic algorithm SVM classifiers are compared for their performance in the classification stage. Random forest classifier offered the most efficient performance [4].

A comprehensive set of features that are manually designed are compared using mammography CAD system and conventional neural network (CNN) on a head-to-head basis. At low sensitivity, CNN offers better performance than the traditional CAD system rather than at high sensitivity [10]. The performance comparable to certified screening radiologists were obtained when the training of CNN model is executed on patch level. For full-field digital mammogram, automated mass detection using patch-based CNN scheme is proposed. In order to adapt specific domain, transfer learning is investigated [13]. CBIS-DDSM public database is used for training the CNN and INbreast dataset is used for testing the model. InceptionV3, ResNet50 and VGG16 deep CNN architectures are evaluated using a fully automated framework in terms of classification of non-mass and mass breast regions. An improved testing accuracy was obtained on CBIS-DDSM dataset for non-mass and mass breast region classification using InceptionV3 CNN [14].

3. Proposed Methodology

This paper proposes a smart digital mammographic screening system for processing images in large volumes irrespective of the nature of images. Watershed segmentation is performed based on appropriate selection of internal and external markers using multiple threshold extended maxima transformations in this technique. Extreme Learning Machine based single layer feed forward network along with Bayesian classifier is used for reducing false positive areas. Feature vector with features like texture and contrast are calculated using these approaches. Digital Mammography Screening database (DMS) is created with 100 mammographic images for the purpose of evaluation. Further, online databases like Breast Cancer Database (BCDB) and BreakHis are also used for analysis. Images in these datasets are available in several old formats like lossless JPEG (LJPEG) and so on. In such images, it is challenging to locate suspicious areas. Digital imaging and communications in medicine (DICOM) format is used as a standard and images are converted on decompression from the datasets. The DICOM format is a standard for training data for pathological diagnosis and other medical images. Bounding boxes and other annotations are included in these images.

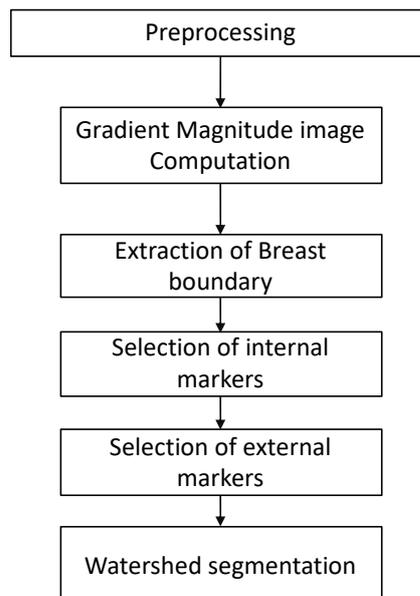


Figure 1: Flow Diagram of Image Segmentation

A Digital Mammography Screening database (DMS) is created with 100 mammographic images for the purpose of evaluation. Of this, 25 mediolateral oblique view and 50 craniocaudal view images are considered. All 100 mammographic images contain masses. Figure 1 represents the basic flow diagram of image segmentation. Initially, preprocessing is performed. Further, the gradient magnitude image is computed and the best boundary is extracted. The internal and external markers are selected and watershed segmentation is performed. Figure 2 provides the stages of image processing where multiple threshold values are used for extracting features from the image. Further, skeletonization is performed to obtain specific information from the processed image. Finally, external markers are selected and the image is segmented and ROIs are extracted.

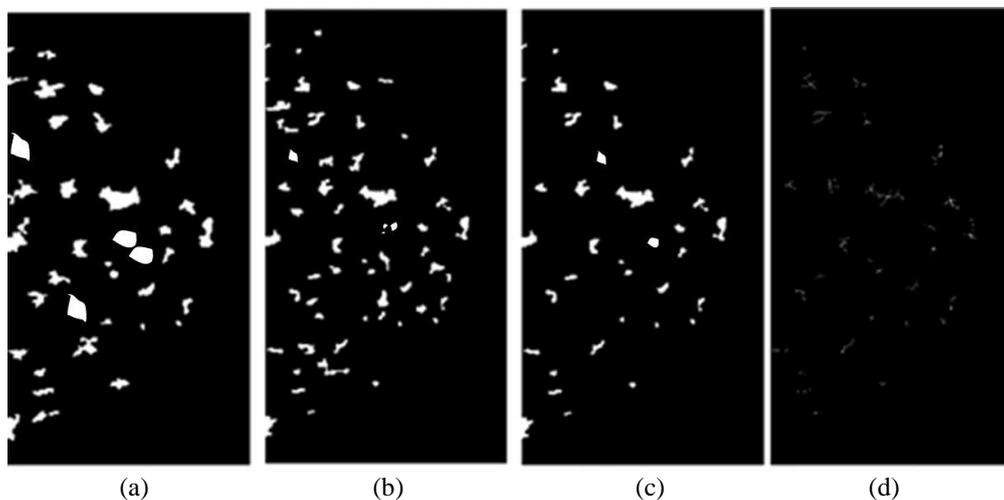


Figure 2: (a) Image with threshold value of 25, (b) Image with threshold value of 15, (c) Crossing of two threshold images, (d) Skeletonization image

4. Results and Discussion

15 features are available in the feature vector estimated. Two different classifiers are tested independently in second stage of classification. The false positive areas are reduced further by implementing two alternative schemes. Extreme Learning Machine (ELM) based single layer feed forward network along with Bayesian classifier is used for this purpose. These supervised classification schemes are used extensively in medical diagnosis. The ELM based network offers improved performance with images that are semi-structured and unstructured. Bayesian classifiers requires lesser training data, they are insensitive to irrelevant features and are quick in classification. Leave one out scheme is used while implementing Bayesian classifier for training

and evaluation. During the first classifier application, the suspect areas that remain from the mammogram are used as test samples. The classifier training set is derived from the suspect areas obtained from the rest of the mammograms. This process is repeated for multiple iterations equivalent to the total number of images so as to use all suspicious entities from every mammogram as test samples.

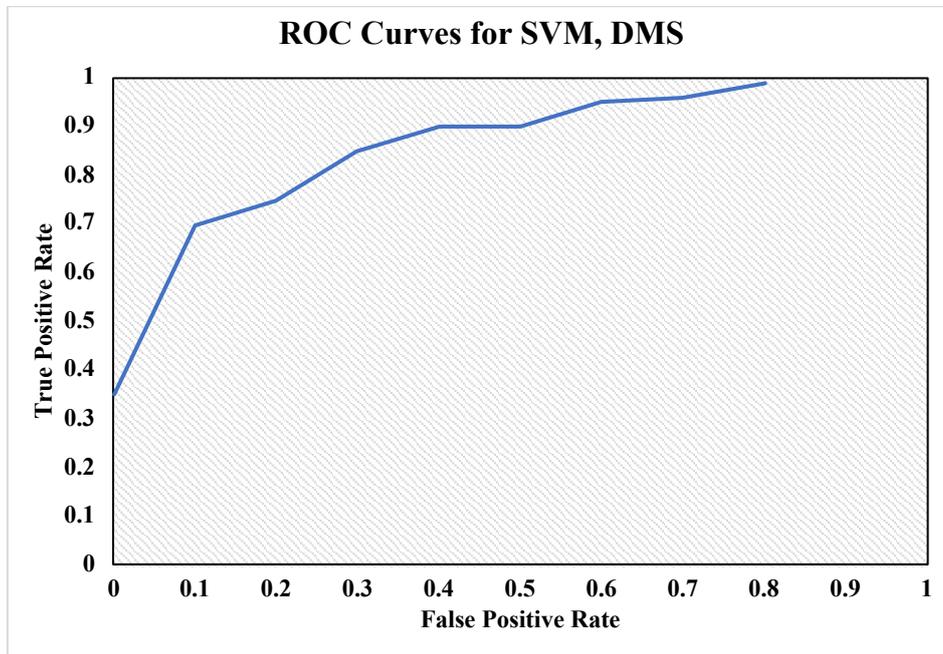


Figure 3: DMS - Overall false positive rate

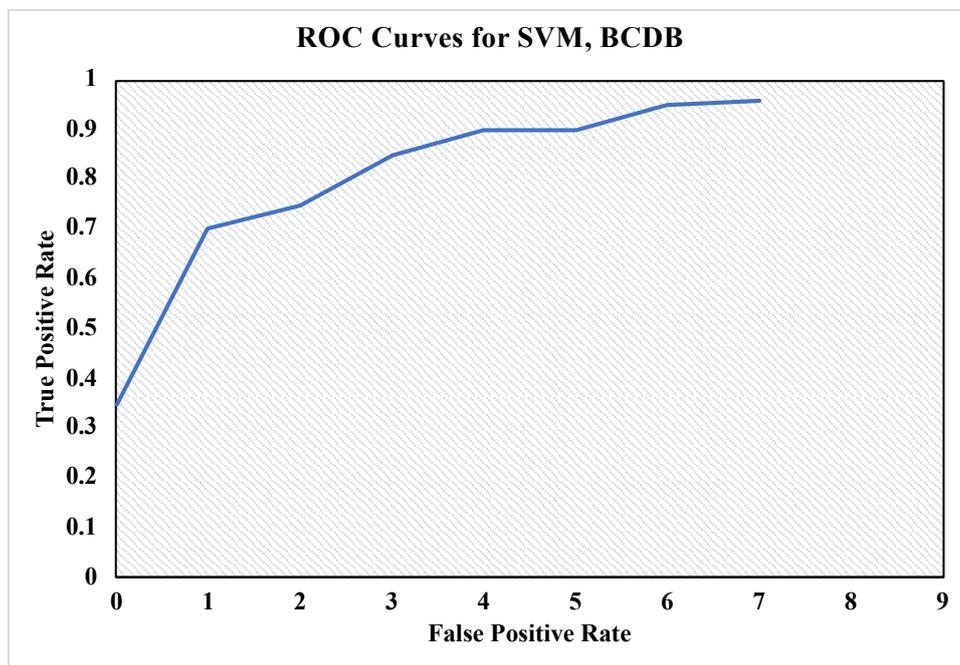


Figure 4: DCBD - overall false rate

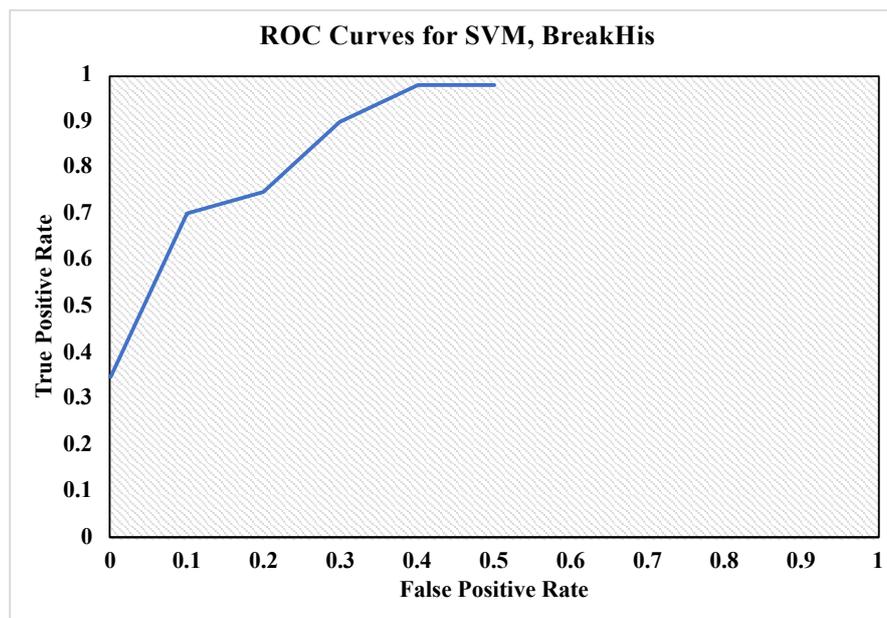


Figure 5: BreakHis - Overall false positive rate

The feature vector is calculated and target values of the suspect areas are estimated based on the feature values. Synthetic minority oversampling scheme is used to set balance in the system. Figure 3-5 provides the overall false positive rate and ROC values for DMS, DCBS and BreakHis datasets.

5. Conclusion

A smart digital mammographic screening system for processing large volume of images for recognition and characterization of masses is proposed in this paper. The system consists of three stages. At the initial stage, morphological operation is performed for detection of suspect objects in the mammographic image. The false positive objects are reduced on applying the two-stage classifier after applying watershed segmentation. The healthy breast tissues and masses are distinguished in this stage. The drawbacks of retrospective processes and computational complexity do not exist in watershed transformation as well as initial detection. This increases the speed of operation of the overall system. Over-segmentation issue is also avoided as marker-based watershed segmentation cooperates with extended maxima transformation to a large extent. The background breast tissues and masses are discriminated using the feature set for computation. This provides good quality segmentation. The masses proven by biopsy vary from the boundaries in certain cases during initial detection of suspect areas despite the remarkable performance of the proposed scheme. However, rather than accuracy of the limits, mass detection is crucial. The biopsy proven masses are not rejected while the false positive areas are reduced efficiently. The flexibility of two-stage classifier is also extremely high.

False positive results may be obtained and certain masses may not be detected by the CAD system in the mammograms. These systems are used by the radiologists as a second opinion for identification of the misinterpreted lesions. Machine learning technique helps in automatic selection of threshold value based on the type of image to perform better image processing. Future work is directed towards enhancing the watershed segmented areas with addition or deletion of pixels. This would help in better identification of masses from mammographic images. Gabor filters and wavelet transformation can further be applied to improve the classification efficiency and computation of feature vector.

References:

- [1] Al-antari, M. A., Al-masni, M. A., Park, S. U., Park, J., Metwally, M. K., Kadah, Y. M., ... & Kim, T. S. (2018). An automatic computer-aided diagnosis system for breast cancer in digital mammograms via deep belief network. *Journal of Medical and Biological Engineering*, 38(3), 443-456.
- [2] Kim, E. K., Kim, H. E., Han, K., Kang, B. J., Sohn, Y. M., Woo, O. H., & Lee, C. W. (2018). Applying data-driven imaging biomarker in mammography for breast cancer screening: preliminary study. *Scientific reports*, 8(1), 1-8.

- [3] Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., & Sieh, W. (2019). Deep learning to improve breast cancer detection on screening mammography. *Scientific reports*, 9(1), 1-12.
- [4] Salvi, M., Acharya, U. R., Molinari, F., & Meiburger, K. M. (2020). The impact of pre-and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis. *Computers in Biology and Medicine*, 104129.
- [5] Aboutalib, S. S., Mohamed, A. A., Berg, W. A., Zuley, M. L., Sumkin, J. H., & Wu, S. (2018). Deep learning to distinguish recalled but benign mammography images in breast cancer screening. *Clinical Cancer Research*, 24(23), 5902-5909.
- [6] Xie, L., Zhang, L., Hu, T., Huang, H., & Yi, Z. (2020). Neural networks model based on an automated multi-scale method for mammogram classification. *Knowledge-Based Systems*, 208, 106465.
- [7] Ahmed, L., Iqbal, M. M., Aldabbas, H., Khalid, S., Saleem, Y., & Saeed, S. (2020). Images data practices for semantic segmentation of breast cancer using deep neural network. *Journal of Ambient Intelligence and Humanized Computing*, 1-17.
- [8] Shi, P., Wu, C., Zhong, J., & Wang, H. (2019, August). Deep Learning from Small Dataset for BI-RADS Density Classification of Mammography Images. In 2019 10th International Conference on Information Technology in Medicine and Education (ITME) (pp. 102-109). IEEE.
- [9] Tan, Y. J., Sim, K. S., & Ting, F. F. (2017, November). Breast cancer detection using convolutional neural networks for mammogram imaging system. In 2017 International Conference on Robotics, Automation and Sciences (ICORAS) (pp. 1-5). IEEE.
- [10] Thakur, R. S., Yadav, R. N., & Gupta, L. (2019). State-of-art analysis of image denoising methods using convolutional neural networks. *IET Image Processing*, 13(13), 2367-2380.
- [11] Vijayakumar, T. (2019). Neural network analysis for tumor investigation and cancer prediction. *Journal of Electronics*, 1(02), 89-98.
- [12] Pandian, A. P. (2019). Identification and classification of cancer cells using capsule network with pathological images. *Journal of Artificial Intelligence*, 1(01), 37-44.
- [13] Shakya, S. (2020). Analysis of artificial intelligence based image classification techniques. *Journal of Innovative Image Processing (JIIP)*, 2(01), 44-54.
- [14] Shakya, S. (2020). Analysis of artificial intelligence based image classification techniques. *Journal of Innovative Image Processing (JIIP)*, 2(01), 44-54.