ISSN 1870-4069

Machine Learning Techniques for Diagnosis of Breast Cancer

Alfonso Rojas Domínguez

Tecnológico Nacional de México, Campus León, Mexico

alfonso.rojas@gmail.com

Abstract. Breast cancer is the most common form of cancer in the female population. As with any form of cancer, early detection of breast cancer is one of the most important factors affecting the possibility of recovery from the disease. Early detection of breast cancer can be achieved through mammography screening programs. Studies have shown that double reading of mammograms improves the detection of breast abnormalities. Unfortunately, the enormous amount of mammograms to be examined prohibits the practice of double reading by human experts. Computerized systems for automated detection and classification of breast abnormalities in mammograms have been developed as a possibility to alleviate this problem. These systems have the objective of accompanying human experts and eventually replace one of them in double or triple reading of mammograms. Some of the signs associated with breast cancer that can be observed in mammograms include: masses, calcifications, distortion of the parenchimal tissue, asymmetry of breast tissue between the breasts of a patient, etc. Beginning from about three decades ago, computerized systems and algorithms specialized in the detection and diagnosis of each type of these signs have been proposed by researchers and described in the scientific literature. Interest in the problem persists to this day; the development of automated systems for detection/diagnosis of breast cancer is currently a very active research area. In this work, an overview of the techniques developed for diagnosis of breast cancer is presented from a perspective of the work that has been preciously carried out personally in this area.

Keywords: machine learning, mammography analysis, breast cancer, CAD.

1 Introduction

Breast cancer is the most common form of cancer among women, and the second deadliest type of cancer (after lung cancer) in high-risk countries such as the United Kingdom and the United States [20]. Most medical experts agree that successful treatment of breast cancer is often linked to early diagnosis. Over 80% of patients diagnosed at the earliest stage of the disease are cured by current therapies. Thus, while breast cancer in general cannot be prevented, it is clear that its detection at an early stage is one of the most important factors contributing to a positive prognosis. Physically, breast cancer manifests itself by one or more of the following breast abnormalities: micro calcifications, breast tumors (i.e. malignant breast masses), distortion of the breast tissue (known as architectural distortion), asymmetry of breast tissue between breasts of the same patient, etc.

All of these abnormalities can remain undetected for years even if monthly self-examination is performed, and many cases are completely asymptomatic under routine physical examination. Screening methods for breast cancer currently available or under development include: breast self-examination; mammography (and Digital Breast Tomosynthesis, DBT); magnetic resonance imaging (MRI); and breast ultrasound. Among these, mammography is the preferred screening method. This is due to its sensitivity (much higher than self-examination, although lower than MRI), specificity (higher than MRI), ability to reduce mortality (which self-examination does not possess) and other advantages such as relative cost (less expensive than MRI, DBT and ultrasound), relative non-invasiveness (MRI requires the injection of a chemical agent), etc.

A true-positive (TP) detection occurs when the screening indicates a positive prediction for a case that is later confirmed as cancer, and a false-positive (FP) detection occurs when the screening returns a positive prediction for a case that is later found to be normal. True-negatives (TN) and a false-negatives (FN) are similartly defined. The number of TPs, FPs, TNs and FNs in a test population can be combined into two measures that are used to summarize the efficiency of a screening method: sensitivity is defined as the number of TPs divided by the total amount of cancer cases in the test population; specificity is defined as the number of TNs divided by the total number of cases in the test population with no breast cancer. For example, a recent study on screening mammography reports an overall sensitivity of 86.6% and specificity of 96.8%, with a 0.6% incidence of breast cancer in the population of the study [3]. In other words, the particular screening program is almost always right when it returns a negative prediction (high specificity of $\approx 97\%$), but misses about 13% of the cancer cases (sensitivity is $\approx 87\%$).

Mammography is a specific type of imaging: the radiographic method of imaging the human breast using a low dose of x-ray radiation passed through the breast to form an image of its internal tissue. Nowadays there is a distinction between two slightly different methods of mammography: in conventional mammography the radiation exposes a photographic film. The images thus formed are known as mammograms, and can either be viewed directly on the film by means of a view box or on a work station with previous digitization of the image by electronic scanning. On the other hand, modern mammography systems make use of a digital image receptor instead of a photographic film to produce digital images directly; the method that use these systems is known as digital mammography, and the images that result from this method may sometimes be referred to as digital mammograms. Mammography can aid in the early detection of breast cancers through the screening of asymptomatic women to detect occult abnormalities in the breast internal tissue. This modality is known as Screening Mammography. The first controlled trial of screening for breast cancer was performed in the 1960's [20,12]. A large amount of data is now available regarding screening mammography, however, many questions remain to be answered and some aspects of screening mammography (such as the particular efficacy in different age groups, or its cost-effectiveness), are still debated. It is generally agreed that mammographic screening reduces breast cancer mortality, and a number of leading health care organizations recommend screening mammography on an annual or two-year basis for all women over a certain age (some indicate a two-year mammogram from age 40 and annual screening from age 50, while others recommend annual screening from age 40), since the risk increases throughout a woman's lifetime.

Screening mammography is currently the most effective tool for early detection of breast cancer. However, the visual examination of a mammogram by a radiologist expert in search for abnormalities is a hard and time-consuming task because the images must be examined with great detail and attention. Furthermore, only approximately five out of a thousand cases examined will return findings related to breast cancer. As a result, radiologists fail to detect 10% to 30% of cancers [24]. On the other hand, the cancer detection rate increases about 5% if double reading of mammograms is employed in the screening process. Thus, even if the number of sufficiently trained radiologists were enough to perform human-based examination of the mammograms, the rate of false diagnosis would be predictively large given the difficulty of the task and the results obtained by current studies on the accuracy of diagnosis (positive predictive value is less than 35% [24]). Not surprisingly, the debate as to the best means for analyzing the large volume of screening mammograms has been focused on the possible use of computer technology, both to reduce the analysis time, as well as to increase the accuracy of diagnosis.

Since computer technology can be employed to aid in the detection and the diagnosis of breast cancer with mammography, two different types of systems are identified by researchers and developers: computer-aided detection (CADe) systems, and computer-aided diagnosis (CADx) systems. CADe systems have been developed to help radiologists in detecting lesions that may indicate the presence of breast cancer. Currently, the objective of these systems is to act as a second reader in double-reading of mammograms, where the final decision is made by the radiologist. The objective of CADx systems is to help radiologists in making a recommendation for patient management. CADx systems are used after a positive detection of a breast abnormality has occurred. If the abnormality is suspected to be malignant, a biopsy must be performed to confirm or reject this suspicion. Fig 1 shows a diagram of the typical components in a CADe/CADx system.

From the point of view of researchers, automated detection of micro calcifications is generally considered a well studied problem. This is not the case with other breast abnormalities such as masses and architectural distortions, where

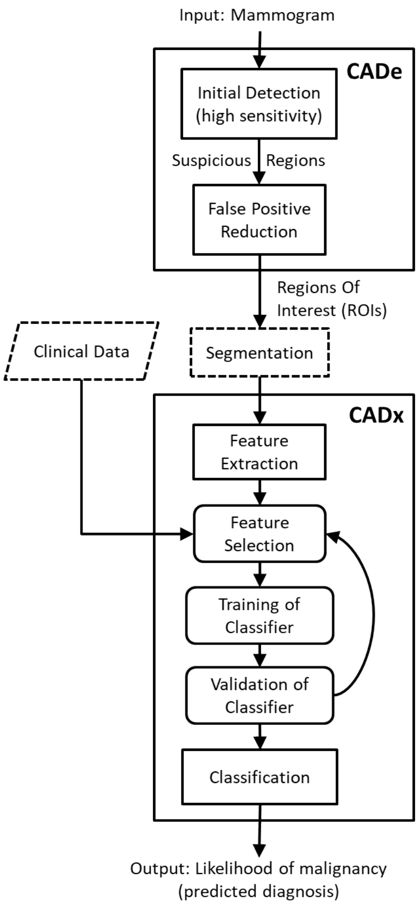

Fig. 1. Typical CADe and CADx systems. Rounded boxes represent stages that are exclusive to the design/development mode of the system. Solid-line square boxes represent stages corresponding to the operational mode of the system. Dashed-line boxes represent stages that may appear in the operational mode of the system or may not be present in that mode.

automated detection and diagnosis methods are still under development. The large variability in the appearance of breast masses, added to the significant overlap in the appearance of malignant and benign masses, and the fact that abnormalities (masses and other) are often occluded or hidden in dense breast tissue, make both their detection and diagnosis very difficult.

2 Diagnosis of Breast Cancer as a ML Task

Both of the computer-aided types of systems (CAD and CADx) for breast cancer with mammography are incarnations of the general model for a pattern recognition system, with the incorporation of ad-hoc image processing and image analysis techniques.

Detection of breast masses involves the scanning of mammograms with the intention of identifying suspicious image regions that may correspond to breast masses. Automated detection methods aim to identify as many abnormalities as possible without labelling normal regions as suspicious. The variability of the appearance of abnormalities, coupled with the similarity between abnormalities and normal breast tissue, make detection a very difficult task.

The most common approach taken for the solution of this problem consists in three steps: 1) application of a mammogram-enhancement routine, 2) segmentation of all potentially suspicious regions, and 3) selection of the suspicious regions. The output of mass detection methods are all the regions labelled as suspicious, or Regions Of Interest (ROIs).

The segmentation of masses consists of separating the masses from the surrounding normal tissue in regions of interest. In this case, automated methods aim to produce a very precise segmentation, so that features that will be later used for the classification of abnormalities can be extracted with reliability.

One of the main obstacles associated with the automated segmentation of masses is the difficulty to determine whether a segmentation is correct or not. This is because in many cases the shape of breast masses is very complex, and the boundaries between mass and background tissue can appear obscured or undefined. Segmentation algorithms may be used as a pre-processing stage for classification of masses.

The ultimate objective of automated methods for classification of masses is to provide a tentative diagnosis (the final decision is produced by a human expert) of individual masses, based on their physical attributes. These methods are incarnations of a generic model for supervised pattern classification systems. According to this model, a classifier is presented with features obtained from a selection of the objects that are to be classified, in a process known as training. The trained classifier can later label objects which were not used in its training, an ability known as generalization. The performance of classification methods depends on the type and quality of the features employed to train the classifier. Features that possess some degree of robustness against segmentation errors are preferred.

The goal of CADx system is to perform the analysis of the ROIs returned by CADe systems. This analysis can be a pixel-based analysis or a region-based analysis. In a pixel-based analysis the features for the classifier are computed directly from every pixel that is part of the lesion in a particular ROI. In contrast, in a region-based analysis the features for the classifier are the result of an extra level of abstraction introduced by an intermediate representation between pixels and features.

3 Mass Detection Techniques

Although the detection of breast masses can be seen as an end on itself, in the context of Computer-Aided mammography-analysis it is often the first of a series of steps required to achieve the classification of abnormalities. Detection algorithms return the location of the breast masses in the mammograms; additionally, these can also return an approximate segmentation of the detected abnormalities, which can be used to perform the extraction of features required for automated classification.

In our previous work [21] we described a method for mammogram enhancement and abnormality detection. The proposed method is divided into three main stages. The first stage is a mammogram enhancement procedure that has the objective of improving the segmentation of the distinct structures in the mammogram when performed via simple conversion to binary images at multiple threshold levels. This enhancement algorithm is different from others in the literature in that it computes the parameter of the enhancing function in an adaptive manner, based on local statistics of the pixels in the mammographic image, and it considers multiple scales. The second stage consists of segmentation and feature extraction steps. In this stage, regions are segmented and several shape and gray-level characteristics of the regions are computed. Finally, in the third stage, a ranking system is employed to select suspicious regions. This is a novel approach to the problem of region selection (elimination of FPs) that does not require training, and implements a type of on-the-fly feature selection.

Our multi-thresholding segmentation method has been cited in different studies as a means to enable the extraction of classification features [6,7,2].

For instance, Al-Najdawi et al. published their own method for enhancement of mammograms and segmentation of breast masses, similar to what we described previously.

The authors also presented their method for classification of masses. Unfortunately, their classification method does not employ a robust classifier, but instead reaches a decision based on a set of heuristic rules and no validation or independent (from the training data) test results were reported. In contrast, Liu & Zeng [17] trained an Support Vector Machine (SVM) on texture features and geometry features, achieving a sensitivity of 76.9% at 1.43 FPs per image. In another example, Pak et al. published a CAD method in which the Non-Subsampled Contourlet Transform and Super Resolution are used for image enhancement, shape analysis features are extracted from thresholded regions and classified with AdaBoost (ANN, SVM and K-NN were also tested). They report a mean sensitivity of 87.15% with 93.58% specificity [19].

Finally, Choi et al. described their proposal of a method for the generation of a classifier ensemble for CADe/CADx systems on mammography [7]. The ensemble contained a number of base classifiers, each of which was associated with a particular feature set previously used in the literature. A total of twelve feature sets were considered, including our own proposal of features for characterization of spiculation and fuzziness of mass margin [22]. For the base classifiers, the authors compared experimentally between SVMs with radial basis function as kernel and Neural Networks trained with backpropagation. Importantly, their study proved that an ensemble of classifiers can achieve a higher performance (for both the detection and the classification tasks) than single classifiers even if feature selection is used to boost the performance of the individual classifiers. Ensemble systems often overcome other strategies in many applications.

4 Robust Features for Breast Mass Classification

In our previous work [22,23] four new features for the analysis of breast masses were presented. These features were designed to be insensitive to the exact shape of the contour of the masses, so that an approximate contour, such as one extracted via an automated segmentation algorithm, can be employed in their computation. Two of the features, Sp_{SI} and Sp_{GO}, measure the degree of spiculation of a mass and its likelihood of being spiculated. The Sp_{GO} feature is a measure of the relative gradient orientation of pixels that correspond to possible spicules based on a feature known as Phase Congruence [16].

The Sp_{SI} feature is based on a comparison of mutual information measures between selected components of the mammographic images, which are obtained by means of Gabor filter banks. The last two features, Fz_1 and Fz_2, measure the local fuzziness of the mass margins based on points defined automatically. The features were tested for characterization (i.e. discrimination between circumscribed and spiculated masses) and diagnosis (i.e. discrimination between benign and malignant masses) of breast masses using a set of 319 masses and three different classifiers. In the characterization experiments the features produced a result of approximately 90% correct classification. In the diagnosis experiments, the performance achieved was approximately 76% of correct classification.

Similar ideas to the classification features described above have been published and have received high numbers of citations [18,1,14]. For instance, Casti et al. presented a multistage system for detection and classification of breast abnormalities [5].

The detection stage included analysis of a Gradient Vector Field, while the classification features were based on the response of multidirectional Gabor filters, clustering and differential feature extraction based on the clusters created. The system detected nearly 80% of the malignant tumors from the DDSM and MIAS datasets at 3.47 and 2.92 FPs per image, respectively.

Khan et al. also investigated the performance of different approaches for directional feature extraction for mass classification based on banks of Gabor filters [15]. The authors compared six approaches and concluded that a Windows based Statistical Magnitude Gabor Response method was significantly superior to the other approaches for the classification of masses vs. normal tissue and for the classification of benign vs. malignant masses. The classification was performed by means of an SVM adapted to deal with unbalanced datasets (since in this problem, normal instances are much more frequent than abnormal instances). Khan et al. also worked on finding optimized Gabor filter banks through an incremental clustering algorithm (for filter selection) and Particle

Swarm Optimization (to optimize the parameters of the filters). Their results indicate that optimization of the parameters of the banks produces significantly better results and comparable to those in the state-of-the-art.

In 2016, Jiao et al. presented a deep-feture based framework for breast masses classification [14]. Their framework employed a Convolutional Neural Network (CNN) to extract hierarchical features from different layers of the CNN. Then the authors introduced a decision mechanism with two classifiers and a similarity measure to produce a final result. The CNN was pretrained on the LSVRC dataset [8] and fine tuned using the DDSM. The classification results of this method were superior to those in the state-of-the-art for the DDSM. Similarly, other deep-based frameworks have been quite successful in solving this problem and are preferred nowadays over schemes based on the more "traditional" feature extraction methodologies [9,11,10]. Jiao et al. also described a parameter updating strategy for improving performance of pre-trained CNNs models (through a metric-learning network trained jointly with the CNN), further improving their classification results on the DDSM.

More recently, Al-antari et al. described a CADx system based on texture (statistical) features extrated from the Gray Level Co-occurrence Matrix (GLCM) [13] of breast abnormalities previously detected, together with invariant moments, and first order statistics for a total set of 347 features [1]. The authors compared the results of classification by means of LDA, QDA, NN and a Deep Belief Network (DBN), concluding that the DBN achieves significantly superior performance, although their experimentation employed a limited number of instances (220 masses, equally split between benign and malignant classes and 116 normal tissue samples).

5 Conclusion

A brief overview of techniques for computer-aided detection and diagnosis or breast abnormalities has been presented. The most important observations that can be made are the following:

1. There is a large variety of frameworks for CAD/CADx of breast cancer based on mammography analysis. Although many different features for breast masses detection and classification have been proposed over the years, the most popular and effective features are those based on robust texture (statistical) analysis and multidirectional/multiscale filtering for identification of spiculated masses.
2. Traditionally, researchers in this area have preferred the use of complex features followed by simple classifiers. Some of the classifiers that appear most often are LDA and (linear) SVMs. In some cases, to boost performance, ensembles of classifiers have been employed. Although the results can be significantly better than those of individual classifiers, the complexity of the system and the number of system parameters to be adjusted also increases. In some cases, a degree of over-fitting to the training data cannot be excluded.

3. From approximately 2015-to date, the deep learning models that have dominated in many pattern recognition applications, have been adopted by researchers working on CAD of breast cancer. However, due to the particular characteristics of the problem (such as unbalanced data, lack of sufficient amounts of labelled training data, and complexity of the problem), rather than using typical deep learning models, the preferred solutions include transfer learning from large generic datasets or from large amounts of hand-crafted features, and hybrid/augmented systems that employ optimization methods and incorporate additional knowledge to improve the performance of regular deep models.

Ultimately, the goal of researchers should be to produce CAD systems that are as robust and directly applicable to different problem domains as possible. In our present case, this would mean to design CAD systems that could be used for different types of cancers, on different imaging modalities, not just mammogram analysis for breast cancer detection and diagnosis. Current research trends are moving in that direction, but at the same time, some new problems arise. For instance, although well-known methods for hyper-parameter optimization of learning models exist (i.e., evolutionary algorithms), training of deep models of sufficient scale is still very expensive computationally, so that the need for accelerated training algorithms exist. Similarly, although nowadays the techniques of dropout and data augmentation are successfully employed to ensure the generalization capabilities of deep models, steps for their systematic application are not yet well established. Clearly, these challenges are not exclusive of CAD systems, but the problems of automated breast cancer detection and diagnosis, because of their intrinsic characteristics, lend themselves to the study of interesting solutions to said challenges.

Finally, regarding the clinical applicability of CAD systems, a recent study indicates that the implementation of CAD as part of the clinical practice requires initial training and involves a learning curve at the beginning of which (first two months) the recall rates may increase dramatically [4]. The same study also reports that some radiologists find the marks produced by CAD systems annoying (there are too many FPs) or threatening (because the use of CAD may be interpreted as if the radiologists were incompetent). Nevertheless, most of the current evidence supports the conclusion that CAD for screening mammography increases the detection sensitivity with a reasonable increase in the recall rate.

List of Acronyms and Abbreviations.

- ANN: Artificial Neural Network.
- CAD(e)/CADx: Computer Aided Detection/Diagnosis.
- CNN: Convolutional Neural Network.
- DBN: Deep Belief Network.
- DBT: Digital Breast Tomosynthesis.
- DDSM: Digital Database for Screening Mammography.
- FN/FP: False Negative/Positive.
- GLCM: Gray Level Co-occurrence Matrix.

– K-NN: K-Nearest Neighbors.
– LDA: Linear Discriminant Analysis.
– LSVRC: Large Scale Visual Recognition Competition.
– MIAS: Mammographic Image Analysis Society.
– ML: Machine Learning.
– MRI: Magnetic Resonance Imaging.
– QDA: Quadratic Discriminant Analysis.
– ROI: Region of Interest.
– SVM: Support Vector Machine.
– TN/TP: True Negative/Positive.

Acknowledgments. This work was supported by the National Council of Science and Technology of Mexico (CONACYT) under Research Grant CÁTEDRAS-2598 (A. Rojas).

References

1. Al-antari, M.A., Al-masni, M.A., Park, S.U., Park, J., Metwally, M.K., Kadah, Y.M., Han, S.M., Kim, T.S.: 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 (2018)
2. Al-Najdawi, N., Biltawi, M., Tedmori, S.: Mammogram image visual enhancement, mass segmentation and classification. Applied Soft Computing 35, 175–185 (2015)
3. Banks, E., Reeves, G., Beral, V., Bull, D., Crossley, B., Simmonds, M., Hilton, E., Bailey, S., Barrett, N., Briers, P., et al.: Influence of personal characteristics of individual women on sensitivity and specificity of mammography in the million women study: cohort study. BMJ 329(7464), 477 (2004)
4. Birdwell, R.L.: The preponderance of evidence supports computer-aided detection for screening mammography. Radiology 253(1), 9–16 (2009)
5. Casti, P., Mencattini, A., Salmeri, M., Ancona, A., Mangeri, F., Pepe, M.L., Rangayyan, R.M.: Contour-independent detection and classification of mammographic lesions. Biomedical Signal Processing and Control 25, 165–177 (2016)
6. Chaddad, A.: Automated feature extraction in brain tumor by magnetic resonance imaging using gaussian mixture models. Journal of Biomedical Imaging 2015, 8 (2015)
7. Choi, J.Y., Kim, D.H., Plataniotis, K.N., Ro, Y.M.: Classifier ensemble generation and selection with multiple feature representations for classification applications in computer-aided detection and diagnosis on mammography. Expert Systems with Applications 46, 106–121 (2016)
8. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255 (2009)
9. Dhungel, N., Carneiro, G., Bradley, A.P.: Automated mass detection in mammograms using cascaded deep learning and random forests. In: International Conference on Digital Image Computing: Techniques and Applications (DICTA). pp. 1–8 (2015)

10. Dhungel, N., Carneiro, G., Bradley, A.P.: A deep learning approach for the analysis of masses in mammograms with minimal user intervention. Medical Image Analysis 37, 114–128 (2017)
11. Dhungel, N., Carneiro, G., Bradley, A.P.: Fully automated classification of mammograms using deep residual neural networks. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). pp. 310–314. IEEE (2017)
12. Dixon, J.M.: Breast cancer: diagnosis and management. Elsevier Science Health Science div (2000)
13. Haralick, R.M., Shanmugam, K., et al.: Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics 6, 610–621 (1973)
14. Jiao, Z., Gao, X., Wang, Y., Li, J.: A deep feature based framework for breast masses classification. Neurocomputing 197, 221–231 (2016)
15. Khan, S., Hussain, M., Aboalsamh, H., Bebis, G.: A comparison of different gabor feature extraction approaches for mass classification in mammography. Multimedia Tools and Applications 76(1), 33–57 (2017)
16. Kovesi, P., et al.: Image features from phase congruency. Videre: Journal of Computer Vision Research 1(3), 1–26 (1999)
17. Liu, X., Zeng, Z.: A new automatic mass detection method for breast cancer with false positive reduction. Neurocomputing 152, 388–402 (2015)
18. de Oliveira, F.S.S., de Carvalho Filho, A.O., Silva, A.C., de Paiva, A.C., Gattass, M.: Classification of breast regions as mass and non-mass based on digital mammograms using taxonomic indexes and svm. Computers in Biology and Medicine 57, 42–53 (2015)
19. Pak, F., Kanan, H.R., Alikhassi, A.: Breast cancer detection and classification in digital mammography based on non-subsampled contourlet transform (nsct) and super resolution. Computer Methods and Programs in Biomedicine 122(2), 89–107 (2015)
20. Patlak, M., Nass, S.J., Henderson, I.C., Lashof, J.: Mammography and beyond: developing technologies for the early detection of breast cancer, vol. 2. National Academy Press (2001)
21. Rojas-Domínguez, A., Nandi, A.K.: Detection of masses in mammograms via statistically based enhancement, multilevel-thresholding segmentation, and region selection. Computerized Medical Imaging and Graphics 32(4), 304–315 (2008)
22. Rojas-Domínguez, A., Nandi, A.K.: Development of tolerant features for characterization of masses in mammograms. Computers in Biology and Medicine 39(8), 678–688 (2009)
23. Rojas-Domínguez, A., Nandi, A.K.: Toward breast cancer diagnosis based on automated segmentation of masses in mammograms. Pattern Recognition 42(6), 1138–1148 (2009)
24. Sampat, M.P., Markey, M.K., Bovik, A.C., et al.: Computer-aided detection and diagnosis in mammography. Handbook of Image and Video Processing 2(1), 1195–1217 (2005)