ISSN 1870-4069

CAD of Breast Cancer: A Decade-Long Review of Techniques for Mammography Analysis

Alfonso Rojas Domínguez, Héctor Puga,
Manuel Ornelas Rodríguez, Itzel Guerrero Gasca

Tecnológico Nacional de México,
Campus León,
Mexico

{alfonso.rojas, lezti10.ig}@gmail.com

Abstract. Breast cancer is the most common type of cancer among women. Early detection is essential to reduce mortality by breast cancer, and mammography is the main tool for detection and first diagnosis of breast cancer. However, mammogram interpretation is a difficult task, and up to 30% of visible cancers in mammograms are missed by human readers. For this reason, Computer-Aided Detection/Diagnosis (CAD) systems have been developed since the 1990's to ameliorate this problem. On the other hand, the popularity of some Machine Learning (ML)/Artificial Intelligence techniques has increased dramatically in the last decade due to technology development leading to improved performance. Naturally, these techniques have also been introduced into CAD systems for mammography analysis. In this work, a review of the techniques employed in CAD systems is presented; this review is focused on the paradigm shift observed within the ML research community and how this has been reflected in the design of CAD systems for breast cancer based on mammography analysis. Through this work, the reader may gain valuable insights into this field.

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

1 Introduction

Breast cancer is the most common[1] cancer in women worldwide with over 2 million new cases in 2018, contributing 25.4% of the total number of new cases diagnosed in that year.[2] Early detection of breast cancer can help to increase the 5-year survival rate. The primary technology for breast cancer screening is mammography. Double-reading from independent expert radiologists is recommended to minimize detection misses; however, this practice represents additional costs and workload. Although there are ethical and technical difficulties that prevent Computer-Aided Diagnosis (CADx) systems from completely replacing human

[1] Excluding non-melanoma skin cancer, which is extremely common but often curable.
[2] https://www.wcrf.org/dietandcancer/cancer-trends/worldwide-cancer-data

experts, these systems can provide second-, or third-opinions to support the decisions of radiologists based on mammography [8].

In this work we present a review of the most pertinent research about Machine Learning (ML, a subset of Artificial Intelligence) based CADx of breast cancer that has been published in the past ten years. Our review is limited to the techniques that have been developed for the analysis of mammograms and is presented from the perspective of ML practitioners, rather than from that of a radiologist or a health care expert. A search of the literature in this topic produces thousands of results. Necessarily, only the most cited papers and from the publications with the highest impact factors were included in this review. Also, only studies that experiment on mammography, either as digital/digitized mammograms or as Digital Breast Thomosynthesis (DBT), were considered. A number of recent surveys exist on the topic of ML/AI in medical imaging [22,31,23], which were also examined for contributions relevant to the topic of mammography analysis for CADx of breast cancer.

Our review is organized chronologically into two periods: the *Feature Engineering* period (prior to 2015) and the *Feature Learning* period (2015 to date). Also topically, according to the main tasks tackled by CAD systems: detection of abnormalities (masses or microcalcifications), and classification of abnormalities. Thus, our work is presented in two main sections, with subsections corresponding to the mentioned tasks.

2 Background

Before Deep Learning's (DL) near hegemony in almost every area of practical application of ML [27], the research on medical image analysis in general, and on mammography analysis in particular, was organized by tasks rather than by techniques. Thus, different techniques from ML / image processing were used for the detection of breast masses, while other (sometimes similar but often quite different) techniques were developed for their classification into benign/malignant classes, etc. The advent of DL with the AlexNet of Krizhevsky et al. in 2012 [26,2], and its further impressive results in many of the areas of application of ML, circa 2015, brought with it a radically different paradigm in the way that medical image analysis can be addressed. Thus, in most surveys published after 2016, two clearly delineated approaches for the use of ML in CAD are present.

The two paradigms are illustrated in Fig. 1. In the *Feature Engineering* approach, a human expert performs feature design and selection, based on problem data from which specific problem knowledge is extracted, and on his/her own engineering expertise. Afterwards, to be classified, a data instance needs to be preprocessed to extract the engineered features, and these can be fed to a simple classifier[3] such as an Artificial Neural Network (ANN) a Support Vector Machine (SVM) [12], etc.

[3] Said classifier must be trained beforehand on a dataset of extracted features; this process has been disregarded for the sake of simplicity in our exposition.

In the *Feature Learning* paradigm, the problem data is used to train a Deep Neural Network that includes a Feature Extraction Stage and a so-called Classification Head (a classifier embedded within the deep network). Afterwards, a data instance can be input directly into the trained network, to be classified. Notice that substantially more data is required to perform automated feature learning than what is usually required for feature engineering[15]. The reason for this is that humans inherently possess generalization capabilities that allow us to discover patterns from, and to transfer prior knowledge into, new problems; on the other hand, a network must learn everything from scratch, which can only do by examining large amounts of training data [27,7].

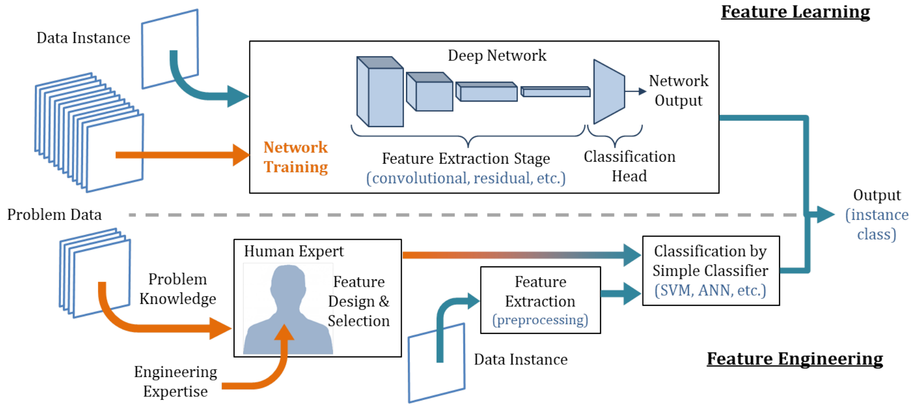

Fig. 1. Two paradigms of ML-CAD: in Feature Engineering the orange arrows represent the data flow for feature design & selection, while in aqua is the classification of unseen instances. In Feature Learning, a Deep Network is trained on large amounts of data (orange), afterwards the network will perform feature extraction & classification of new instances (aqua).

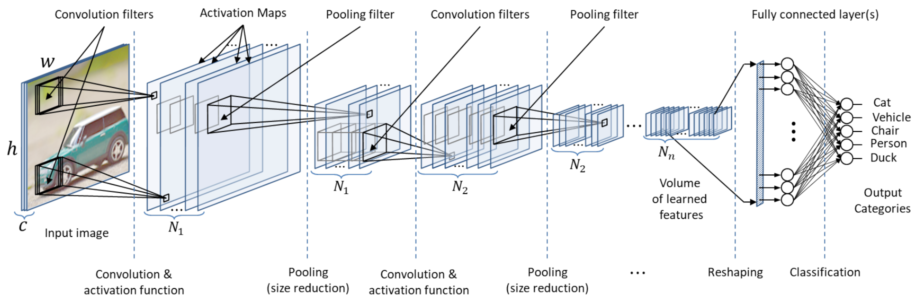

Fig. 2. Convolutional Neural Network for classification.

With few exceptions, the features that are manually engineered by human experts cannot be applied to different problems. In other words, features engineered for breast-mass detection will not be useful for mass segmentation, nor for mass characterization, etc.

On the other hand, the features ultimately learned by a Deep Network, or high-level features, are based on other lower-level features, hierarchically generated by different layers of the network. The features in the first layers describe abstract shapes (lines, edges, gradients, etc.); the features in the last layers correspond to the specific structures of interest, such as spiculated masses or fuzzy boundaries indicative of malignancy. Thus, low-level features possess a degree of generality, so that to some extent the knowledge acquired by the network can be *transferred* to other networks designed for different problems. This is known as Transfer Learning (TL) [6], and is an important advantage of DL over other ML techniques.

Regarding the type of networks that can be employed at the core of the Feature Learning paradigm, we can list a variety of architectures [7]: Convolutional Neural Networks (CNNs), Residual Networks, Recurrent Networks, Deep Autoencoders [21], Deep Belief Networks [20], etc. The most commonly employed type of network for mammography analysis is the CNN [1] (Fig. 2). A CNN is characterized by being formed by several *convolutional* layers, that contain a number of trainable convolution filters that produce so-called activation maps.

3 Feature Engineering Period (2009-2015)

Detection of breast masses by means of 2D mammography is hindered by dense parenchyma, which can also generate false positive (FP) detection because the overlapping tissue may mimic lesions. One alternative is the use Digital Breast Tomosynthesis, patented in 1999[4], although initial experiences and comparisons between conventional mammography and DBT were being performed circa 2007 [19,35]. DBT produces a set of 2D images or 'slices' from which a 3D voulume of the breast can be reconstructed.

Overlapping of the parenchyma is reduced and this may offer higher sensitivity than regular mammograms (and reduce recall rate [17]), but with increase in workload. In the Feature Engineering period, several studies describe the developing of early CADe systems to automate the analysis of DBT data [11,9].

Detection of Masses.– In 2008, Chan et al. presented a comparison of three approaches for CADe of masses in DBT [10]. The comparison was carried out on a dataset of 100 DBT cases. The approaches differ in the information that their system employs (i.e. the 3D DBT reconstruction; the backprojected 2D images that form a DBT case; or the combination of those). Detection was based on three-dimensional gradient field analysis and several features that describe a candidate mass. Features included morphological features (volume (3D), area (2D), perimeter, diameter and compactness); statistics of the gray level; run length texture features[14] obtained via the Rubber Band Straghtening Transform [39] and from spatial gray level dependence matrices [18], and a Hessian feature in 2D. The authors reported that the combined approach was superior to the other approaches compared. In a previous work (Rojas-Domínguez & Nandi, 2009) we explored the techniques for segmentation of masses, including the standard features and other proposed features [38].

[4] https://www.massgeneral.org/imaging/services/digital-breast-tomosynthesis.aspx

Detection of Microcalcifications.– Typical techniques employed for detection of microcalcifications vary from simple thresholding to sophisticated methods, such as: the Wavelet Transform and similar techniques, like the Contourlet Transform, that also enable the generation of mammograms in super-resolution [33,47], as well as image denoising and enhancement [16]; the Laplacian of Gaussian (or Difference of Gaussians) filter, used as a *blob* detector; morphological operators; and other ad-hoc methods such as fuzzy-logic, Gabor filters, Zernike moments, etc. A previous work (Rojas-Domínguez & Nandi, 2008) provides a review of the most important of those methods [36]. The review by Elter& Horsch [13] includes a table comparing these techniques in terms of their strengths and weaknesses, while Pak et al. provide a performance-based comparison of these techniques (up to 2015) [33].

Elter & Horsch stated that "half of all missed cancers seem to be missed due to missclasification rather than due to oversight"[13]. This means that CADx systems to characterize lesions are as important as CADe systems that detect those lesions.

Classification of Abnormalities.– The most crucial features for classification of breast masses are their shape and appearance of their margins. In simple terms, benign masses possess clearly defined, smooth margins and round shapes, while malignant masses appear as irregular shapes with diffuse/fuzzy or *spiculated* margins (a spicule is a needle-like structure). Irregular borders indicate that the mass has invaded surrounding tissue, a sign of a metastatic process. Typical results achieve a sensitivity of 86% with 3 FP per image [41]. See our previous work (Rojas-Domínguez & Nandi, 2009) for a discussion of this topic and our proposed robust features for characterization of breast masses [37].

4 Feature Learning Period (2015-to date)

A comprehensive technical review of DL in medical image analysis is provided by Shen et al. that includes details about the most common DL models and an overview of the different applications [42]. A more general overview of DL in medical imaging is [45]. One of the first works that tested the use of DL techniques for mammography analysis is [4]. In 2016 Arevalo et al. studied the automatic classification of breast lesions on a proposed biopsy-proven benchmarking dataset from 344 cases with a total of 736 film[5] mammography views, and manually segmented lesions (426 benign and 310 malignant) [5]. In contrast to previous works where DL models were trained in an unsupervised way [34,24], the authors trained a CNN for feature learning in a supervised fashion, with the peculiarity that instead of performing the classification by means of embedded fully-connected layers, they used an independently trained SVM (nowadays this

[5] That project has been updated to digital mammograms: https://bcdr.ceta-ciemat.es/

practice is far less common). They also used a preprocessing stage with local and global image normalization and data augmentation.[6]

The authors report that their method exhibits improved performance (from 0.787 to 0.822 AUC[7]) when compared to state-of-the-art image descriptors, such as Histogram of Oriented Gradients and Histogram of the Gradient Divergence [32]. Their model also outperformed a set of 17 hand-crafted features that take advantage of additional information from segmentation by the radiologist (notice that these are generic image descriptors):

- Intensity: mean, median, maximum, minimum, standard deviation, skewness, kurtosis.
- Shape: area, perimeter, circularity, elongation, mass centroid, form.
- Texture: contrast, correlation, entropy.

In recent studies, Giger [15] states that both handcrafted and deep-learned features are important, underlining that systems combining both types of features perform the best [3].

Detection of Abnormalities.– In 2016, Samala et al. employed TL on a CNN-based CAD system to improve detection of breast masses in DBT [40]. They pre-trained the more generic layers (first 3 of 4), on a large training set of mammograms and then trained the more specific layer on DBT data. The CNN-based CAD system was compared against feature-based CAD including morphological (volume, area, perimeter, longest diameter and compactness), grey level (statistics, contrast and histogram features) and texture features (run-length statistics on the rubber-band [39] of the objects margin). Based on said features, the authors applied linear discriminant analysis-based classifiers to perform FP reduction. They obtained statistically significant improvements of their CNN-based CAD system over their feature-based system. More importantly, they showed that TL can preserve the low level similarities and capture the high-level differences between representations of the masses.

In 2017 Kooi et al. compared a CNN-based system for detection of mammographic lesions against a state-of-the-art, feature-based system relying on a set of 74 manually engineered features [25]. Among these features, those related to Location, Context and Patient features were also provided to the CNN-based system because they were considered complementary to the information that the CNN could extract from the training mammograms. A total of 40,090 mammograms were used for training, and 18,182 for evaluation. The CNN was a scaled-down VGG [43] with 5 convolutional layers plus 2 fully connected layers. The comparison showed no statistically significant differences ($p = 0.2$) using the AUC. However, at high specificity, statistically significant differences were found between the CNN+manual features against their reference state-of-the-art system. Finally, in a comparison against human readers, the performance of the CNN was lower than the average of 3 human readers.

[6] Data augmentation produces variations of the original input data by performing simple geometric transformations such as translations, rotations and mirroring.

[7] Area Under the ROC Curve, a detection performance measure.

Classification of Abnormalities.– The very recent review of Abdelhafiz et al. identifies major challenges in training CNNs for mammography analysis and the most popular solutions that have been employed to overcome them [1]. A major challenge for the training of deep networks is the lack of sufficient labelled data.

Sun et al. presented a graph based scheme of semi-supervised learning (SSL) for CNNs [44]. While CNN training usually requires a large amount of labelled data, their proposed scheme allowed the training of the CNN with only (before data augmentation) 100 labeled ROIs plus 2400 originally unlabeled ROIs that were labeled automatically by a co-training algorithm [17]. Their CNN followed the design of LeCun [29,28] and was formed by 3 convolutional layers. The CNN trained by SSL achieved the best performance (AUC=0.88), followed by an SVM also trained by SSL (AUC=0.85).

Ting et al. [46] recently proposed to employ an interactive detection-based lesion locator by means of an auxiliary network of the type Single Shot Multibox Detector [30], which is a deep network developed for multiple object detection. Their proposal achieves an AUC=0.90. Ting et al. also performed a sophisticated preprocessing, including patch-based data augmentation by rotating and flipping the lesion patches.

5 Conclusion

The effects of the paradigm shift observed in the field of ML extend to those scientific fields in which ML techniques have been successfully applied. The ML-based CAD systems for mammography analysis developed in the last decade, clearly reflect these effects. However, the dominance of DL in this field is not as extensive as in other ML applications. Based on the present review, we can conclude that the *Feature Engineering* (favouring generic image descriptors) and the *Feature Learning* approaches, currently coexist. The best performance is obtained by fusion strategies that combine both. Large high-quality mammography/DBT datasets could promote the maturity of DL techniques in the near future, but these are not easy to compile. In the meantime, the inclusion of complementary information not available in the mammograms, plus such strategies as TL and data augmentation, are essential to the implementation of the most competitive CAD systems.

Acknowledgements. This work was supported by the National Council of Science and Technology of Mexico (CONACYT) under Research Grant: CÁTEDRAS-2598 (A. Rojas) and Postgraduate Scholarship 748219 (Itzel Guerrero).

References

1. Abdelhafiz, D., Yang, C., Ammar, R., Nabavi, S.: Deep convolutional neural networks for mammography: advances, challenges and applications. BMC Bioinformatics 20(11), 281 (2019)

2. Alom, M.Z., Taha, T.M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M.S., Van Esesn, B.C., Awwal, A.A.S., Asari, V.K.: The history began from alexnet: A comprehensive survey on deep learning approaches. arXiv:1803.01164 (2018)
3. Antropova, N., Huynh, B., Giger, M.: Deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. Med. Phys. 44(10), 5162–5171 (2017)
4. Arevalo, J., González, F.A., Ramos-Pollán, R., Oliveira, J.L., Lopez, M.A.G.: Convolutional neural networks for mammography mass lesion classification. In: 37th Annual International Conference of the IEEE Eng. in Medicine and Biology Society (EMBC). pp. 797–800 (2015)
5. Arevalo, J., González, F.A., Ramos-Pollán, R., Oliveira, J.L., Lopez, M.A.G.: Representation learning for mammography mass lesion classification with convolutional neural networks. Computer Methods and Programs in Biomedicine 127, 248–257 (2016)
6. Bengio, Y.: Deep learning of representations for unsupervised and transfer learning. In: Proceedings of ICML Workshop on Unsupervised and Transfer Learning. pp. 17–36 (2012)
7. Bengio, Y., et al.: Learning deep architectures for AI. Foundations and Trends® in Machine Learning 2(1), 1–127 (2009)
8. Birdwell, R.L.: The preponderance of evidence supports computer-aided detection for screening mammography. Radiology 253(1), 9–16 (2009)
9. Chan, H.P., Wei, J., Sahiner, B., Rafferty, E., Wu, T., Roubidoux, M., Moore, R., Kopans, D., Hadjiiski, L., Helvie, M.: Computer-aided detection system for breast masses on digital tomosynthesis mammograms: preliminary experience. Radiology 237(3), 1075–1080 (2005)
10. Chan, H.P., Wei, J., Zhang, Y., Helvie, M.A., Moore, R.H., Sahiner, B., Hadjiiski, L., Kopans, D.B.: Computer-aided detection of masses in digital tomosynthesis mammography: Comparison of three approaches. Medical Physics 35(9), 4087–4095 (2008)
11. Chan, H.P., Wei, J., Zhang, Y., Moore, R., Kopans, D., Hadjiiski, L., Sahiner, B., Roubidoux, M., Helvie, M.: Computer-aided detection of masses in digital tomosynthesis mammography: combination of 3d and 2d detection information. In: Medical Imaging 2007: Computer-Aided Diagnosis. vol. 6514. International Society for Optics and Photonics (2007)
12. Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning 20(3), 273–297 (1995)
13. Elter, M., Horsch, A.: Cadx of mammographic masses and clustered microcalcifications: a review. Medical Physics 36(6Part1), 2052–2068 (2009)
14. Galloway, M.M.: Texture analysis using grey level run lengths. NASA STI/Recon Technical Report N 75 (1974)
15. Giger, M.L.: Machine learning in medical imaging. Journal of the American College of Radiology 15(3), 512–520 (2018)
16. Gorgel, P., Sertbas, A., Ucan, O.N.: A wavelet-based mammographic image denoising and enhancement with homomorphic filtering. Journal of Medical Systems 34(6), 993–1002 (2010)
17. Haas, B.M., Kalra, V., Geisel, J., Raghu, M., Durand, M., Philpotts, L.E.: Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. Radiology 269(3), 694–700 (2013)
18. Haralick, R.M., Shanmugam, K., et al.: Textural features for image classification. pp. 610–621. No. 6, IEEE (1973)

19. Helvie, M., Roubidoux, M., Hadjiiski, L., Zhang, Y., Carson, P., Chan, H.: Tomosynthesis mammography versus conventional mammography: comparison of breast masses detection and characterization. In: Radiological Society of North America 93rd Scientific Assembly, Nov (2007)
20. Hinton, G.E.: Deep belief networks. Scholarpedia 4(5), 5947 (2009)
21. Hinton, G.E., Osindero, S., Teh, Y.W.: A fast learning algorithm for deep belief nets. Neural Computation 18(7), 1527–1554 (2006)
22. Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L.H., Aerts, H.J.: Artificial intelligence in radiology. Nature Reviews Cancer 18(8), 500 (2018)
23. Hu, Z., Tang, J., Wang, Z., Zhang, K., Zhang, L., Sun, Q.: Deep learning for image-based cancer detection and diagnosis- a survey. Pattern Recognition 83, 134–149 (2018)
24. Jamieson, A.R., Drukker, K., Giger, M.L.: Breast image feature learning with adaptive deconvolutional networks. In: Medical Imaging 2012: Computer-Aided Diagnosis. vol. 8315, p. 831506. International Society for Optics and Photonics (2012)
25. Kooi, T., Litjens, G., Van Ginneken, B., Gubern-Mérida, A., Sánchez, C.I., Mann, R., den Heeten, A., Karssemeijer, N.: Large scale deep learning for computer aided detection of mammographic lesions. Medical Image Analysis 35, 303–312 (2017)
26. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. pp. 1097–1105 (2012)
27. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 521(7553), 436 (2015)
28. LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., Jackel, L.: Backpropagation applied to handwritten zip code recognition. Neural Comp. 1(4), 541–551 (1989)
29. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., et al.: Gradient-based learning applied to document recognition. vol. 86, pp. 2278–2324. Taipei, Taiwan (1998)
30. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: Ssd: Single shot multibox detector. In: European Conference on Computer Vision. pp. 21–37. Springer (2016)
31. Lu, L., Zheng, Y., Carneiro, G., Yang, L.: Deep learning and convolutional neural networks for medical image computing. Advances in Computer Vision and Pattern Recognition; Springer (2017)
32. Moura, D.C., López, M.A.G.: An evaluation of image descriptors combined with clinical data for breast cancer diagnosis. International Journal of Computer Assisted Radiology and Surgery 8(4), 561–574 (2013)
33. 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)
34. Petersen, K., Nielsen, M., Diao, P., Karssemeijer, N., Lillholm, M.: Breast tissue segmentation and mammographic risk scoring using deep learning. In: International Workshop on Digital Mammography. pp. 88–94. Springer (2014)
35. Poplack, S.P., Tosteson, T.D., Kogel, C.A., Nagy, H.M.: Digital breast tomosynthesis: initial experience in 98 women with abnormal digital screening mammography. American Journal of Roentgenology 189(3), 616–623 (2007)
36. 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)

37. 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)
38. 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)
39. Sahiner, B., Chan, H.P., Petrick, N., Helvie, M.A., Goodsitt, M.M.: Computerized characterization of masses on mammograms: The rubber band straightening transform and texture analysis. Medical Physics 25(4), 516–526 (1998)
40. Samala, R.K., Chan, H.P., Hadjiiski, L., Helvie, M.A., Wei, J., Cha, K.: Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography. Medical Physics 43(12), 6654–6666 (2016)
41. Sampat, M.P., Bovik, A.C., Whitman, G.J., Markey, M.K.: A model-based framework for the detection of spiculated masses on mammography a. Medical Physics 35(5), 2110–2123 (2008)
42. Shen, D., Wu, G., Suk, H.I.: Deep learning in medical image analysis. Annual Review of Biomedical Engineering 19, 221–248 (2017)
43. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 (2014)
44. Sun, W., Tseng, T.L.B., Zhang, J., Qian, W.: Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data. Computerized Medical Imaging and Graphics 57, 4–9 (2017)
45. Suzuki, K.: Survey of deep learning applications to medical image analysis. Med Imaging Technol 35(4), 212–226 (2017)
46. Ting, F.F., Tan, Y.J., Sim, K.S.: Convolutional neural network improvement for breast cancer classification. Expert Systems with Applications 120, 103–115 (2019)
47. Yanez-Vargas, I., Gonzalez-Reyna, S., Gonzalez-Ramirez, A., Guerrero-Gasca, I., Astudillo-Montenegro, F.: Super-resolution of mammograms based on analysis of wavelet family and iterative scales. In: In Proceeding International Conference on Electronics, Communications and Computers CONIELECOMP. pp. 1–5. IEEE (2017)