Deep Learning in Breast Cancer Detection and Classification

Chapter · March 2020

DOI: 10.1007/978-3-030-44289-7_30

4 authors, including:

Ghada Hamed
Ain Shams University
11 PUBLICATIONS **44** CITATIONS
SEE PROFILE

Mohammed Abd El Rahman Marey
Ain Shams University
26 PUBLICATIONS **250** CITATIONS
SEE PROFILE

Safaa Amin
Ain Shams University
3 PUBLICATIONS **5** CITATIONS
SEE PROFILE

Deep Learning in Breast Cancer Detection and Classification

Ghada Hamed(✉), Mohammed Abd El-Rahman Marey, Safaa El-Sayed Amin, and Mohamed Fahmy Tolba

Faculty of Computer and Information Sciences,
Ain Shams University, Cairo, Egypt
ghadahamed@cis.asu.edu.eg

Abstract. Breast cancer is considered one of the primary causes of mortality among women aged 20–59 worldwide. Early detection and treatment can allow patients to have proper treatment and consequently reduce rate of morbidity of breast cancer. Research indicates that most experienced physicians can diagnose cancer with 79% accuracy while 91% correct diagnosis is achieved using machine learning techniques. In this paper, we present the most recent breast cancer detection and classification models that are machine learning based models by analyzing them in the form of comparative study. Also, in this paper, the datasets that are public for use and popular as well are listed in the recent work to facilitate any new experiments and comparisons. The comparative analysis shows that the recent highest accuracy models based on simple detection and the classification architectures are You Only Look Once (YOLO) and RetinaNet.

Keywords: Breast cancer · Mortality · Detection · Classification · YOLO · RetinaNet

1 Introduction

The cancer tumor is developed through the abnormal growth of cells, which invades the surrounding tissues in the human body. There are two classes of tumors, benign and malignant, and if there is no tumor in the breast it is considered normal. The benign tumor cells are non–cancerous cells and grow only locally and cannot spread by invasion. While malignant tumors are cancerous cells and they have the ability to multiple uncontrollably, to spread to various parts of the body and invade surrounding tissue. Breast cancer is one of the most common cancers and a severe public health issue for women all over the world. According to the report of International Agency for Research on Cancer (IARC) presented by the World Health Organization (WHO), in the year 2012, approximately 8.2 million deaths are caused by cancer, which further states that in 2030 the death ratio from cancer is expected to increase up to 27 million [1]. So, on-time and accurate detection, early diagnosis, and active prevention are critical requirements for reducing mortality rate among women.

A.-E. Hassanien et al. (Eds.): AICV 2020, AISC 1153, pp. 322–333, 2020.
https://doi.org/10.1007/978-3-030-44289-7_30

Abnormalities such as masses, micro-calcifications, and areas of asymmetry and distortion within the breast may indicate the existence of breast cancer. Among these abnormalities, masses are the most representative and common lesion type. However, masses can be easily hidden by overlapping breast tissues, making it difficult to detect them. Moreover, some breast tissues are morphologically similar to masses, and thus misidentified as masses. An undetected mass is a false negative, which delays a patient diagnosis until the next screening. A misidentified mass is a false positive, which leads to additional tests including re-screening and biopsy, causing unnecessary anxiety and pain to patients. These problems limit the effectiveness and utility of mammography. So, this task is considered a daily challenge for radiologists since there is a huge number of mammograms that consumes time and effort to examine each view of a mammogram [2–4]. Consequently, there is a tradeoff between sensitivity and specificity that has been realized during the diagnosis process of the existing masses. Through the use of a second reading, either by other experts or by a Computer-Aided Diagnosis (CAD) systems, the overall accuracy and specificity of the mass detection, segmentation, and classification could be improved [5,6] and false positive and negative cases reduced.

There are few studies involving a completely integrated system that include all phases of detection, segmentation, and classification. In fact, the variation of the masses within the surrounding tissues in terms of texture, shape, size, as well as the location in mammograms, makes the detection task challenging [7,8]. Besides this, improving overall accuracy in addition to reducing false positive and negative rates by mass segmentation is a big challenge due to the strong association between the presence of masses and their irregularities in shape, size, and location with low contrast and ambiguous boundaries [9–12]. Unfortunately, these methods still lack performance in handling mass segmentation automatically, because the simple hand-crafted or semi-automatic features based on prior knowledge cannot deal with complex shape variations, as well as the different density distribution of the masses and their surrounding tissues [9]. Besides that the manual feature extraction and selection are extremely time consuming. The manual feature extraction and selection make it difficult to define as well meaningful features. Recently, a few studies based on deep learning models have offered a good alternative to other conventional segmentation methods, by automatically extracting deep high-level hierarchy features for mass segmentation directly from input raw data to avoid the problems of hand-crafting features [13].

A Convolutional Neural Network (CNNs) consists of a number of convolutional layers which can extract features that represent the various contexts of images without feature engineering. That is why CNN has become the most widely used method for image interpretation tasks in many domains such as the breast cancer detection and classification. After the success of CNNs in standard object detection tasks [14], several studies have exploited the advantages of deep CNNs to overcome the drawbacks of conventional mass detection models [15–22].

The main purpose of mammography is to detect early signs of cancer and to diagnose breast masses from the images [23]. Mammograms are considered as the most important tool doctors have not only to screen for breast cancer, but also to diagnose, evaluate, and follow people who have had breast cancer. Screening with mammography is performed in two steps. First, the breast is compressed between two small flat plates. Then, a low X-ray dose is applied directly through the breast and acquired by a two-dimensional (2D) panel detector.

The rest of the paper is organized as follows. In the dataset section, the most common used and public datasets are given and the main differences between each are discussed. Then, the dataset section is followed by the comparative study and analysis section which begins with the detection and classification accuracy metrics used to detect how accurate the proposed model is. The metrics are listed with their definitions and how they are calculated as well. Then, the recent published work in the breast cancer detection and classification from 2017 to 2019 are compared with respective to different aspects including the used datasets, data preprocessing and data augmentation in section three followed by the main building blocks they build on the cancer classification and detection. Finally the conclusion section is presented in the last section.

2 Dataset

This section lists the most common datasets containing mammograms used for the breast cancer detection and classification. There are five datasets presented which are: DDSM, CBIS-DDSM, MIAS, INbreast and BCDR datasets.

DDSM Dataset: The Digital Database for Screening Mammography (DDSM) is an available online dataset [24] that is used for the research purposes in the breast cancer detection and classification systems. This dataset is collected by South Florida University [25]. It is collected to represent real breast data with an average size of 3000×4800 pixels, resolution of 42 microns, and 16 bits.

The DDSM database consists of 2,620 scanned film mammography studies that are categorized into 43 volumes. Each case involves four breast images, two of them are Mediolateral Oblique (MLO) views and the others are Cranio-Caudal (CC) views of each breast. Benign and malignant masses in all mammograms are recognized and annotated by expert radiologists.

CBIS-DDSM Dataset: Curated Breast Imaging Subset of DDSM (CBIS-DDSM) is an updated and standardized version of the Digital Database for Screening Mammography (DDSM) [26]. The CBIS-DDSM collection is available as well [27] which includes a subset of the DDSM data selected and curated by a trained mammographer. The CBIS-DDSM contains 6775 studies. The images have been decompressed and converted to DICOM format. Updated ROI segmentation and bounding boxes, and pathologic diagnosis for training data are also included.

MIAS Dataset: The Mammographic Image Analysis Society (MIAS) is an organisation of UK research groups interested in the understanding of mammograms and has generated a database of digital mammograms [28]. Mammographic images are available via the Pilot European Image Processing Archive (PEIPA) at the University of Essex. The database contains 322 digitized films and is available on 2.3 GB 8 mm (ExaByte) tape. The database has been reduced to a 200 micron pixel edge and padded/clipped so that all the images are 1024×1024.

INbreast Dataset: The INbreast was obtained from the S. João Hospital Centre in Porto [29]. It consists of 410 full digital mammograms. All lesions including masses were assigned a standardized Breast Imaging-Reporting and Data System (BI-RADS) category [30] by a radiologist after interpreting a mammogram. INbreast does not exist publicly on the web but it can be obtained by a request from [31].

BCDR Dataset: The Breast Cancer Digital Repository (BCDR) is a compilation of Breast Cancer anonymized patients' cases annotated by expert radiologists containing clinical data (detected anomalies, breast density, BIRADS classification, etc.), lesions outlines, and image-based features computed from CC and MLO mammography image views.

Currently, two repositories (examples) are available for public domain: one containing digitalized Film mammography (FM) and other one containing Full Field Digital (DM) mammography and related ultrasound images. Also, four benchmarking datasets (two masses-based and two microcalcifications/calcifications-based) representatives of benign and malignant lesions (biopsy-proven) comprising instances of clinical and image-based features are available for free download to registered users [32].

Finally, Table 1 gives the main characteristics of the mentioned common used and public datasets for breast mammograms.

Table 1. The common breast mammographic datasets (N: Normal cases, B: Benign cases, M: Malignant cases)

Dataset	Number of cases	Number of images	Available classes	Image format	Publicly available
DDSM	2620	10,480	N, B & M	JPEG	Yes
CBIS-DDSM	6775	10,239	N, B & M	DICOM	Yes
MIAS	–	322	N, B & M	PGM	Yes
INbreast	115	410	N, B & M	DICOM	No
BCDR	1734	3703 FM - 3612 DM	N, B & M	TIFF	Yes

3 Comparative Study for the Used Evaluation Metrics and Datasets

There have been active developments in the research in each specific area of abnormalities detection, segmentation, and classification. The Mass detection step is considered an important preprocessing stage to detect potential cancerous regions for further analysis by a Computer-Aided Diagnosis system (CAD) system. Breast mass segmentation also plays a crucial role in accurately extracting discriminative shape features of specific mass regions, while excluding surrounding tissues.

This section presents first the metrics used to evaluate a proposed breast mass detection and classification system and how to deduce each metric with its intent. Then, a comparison between the selected recent proposed approaches in the breast cancer detection and classification from the perspective of the data preparation done in each. Finally, a comparative analysis is proposed to show the detection and classification models used and their performance.

3.1 Detection and Classification Accuracy Metrics

This section presents the accuracy metrics used to evaluate the breast cancer detection and classification work. They are used as well to indicate the accuracy measurement process. They are:

- True Positives (TP): They are the cases when the actual class of the data point was 1 (True) and the predicted is also 1 (True).
- True Negatives (TN): They are the cases when the actual class of the data point was 0 (False) and the predicted is also 0 (False).
- False Positives (FP): They are the cases when the actual class of the data point was 0 (False) and the predicted is 1 (True).
- False Negatives (FN): They are the cases when the actual class of the data point was 1 (True) and the predicted is 0 (False).
- Confusion Matrix: It is used for finding the correctness and accuracy of the model. It derives TP, TN, FP and FN in one matrix.
- Accuracy: It is the number of correct predictions made by the model over all kind of predictions made as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

- Precision: It is a measure that tells us the proportion of patients that we diagnosed as having cancer or actually had cancer. The predicted positives (People predicted as cancerous are TP and FP) and the people actually having a cancer are TP as follows:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

– Recall (Sensitivity): It is a measure that tells us the proportion of patients that actually had cancer and were diagnosed by the model as having cancer. The actual positives (People having cancer are TP and FN) and the people diagnosed by the model having a cancer are TP as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{3}$$

– Specificity: It is a measure that tells us what proportion of patients that did NOT have cancer and were predicted by the model as non-cancerous. The actual negatives and the people diagnosed by us not having cancer are TN as follows:

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

– Dice similarity coefficient (F1-score): It represents a harmonic average of the precision and sensitivity and measures the similarity rate between predicted and ground-truth regions as follows:

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{5}$$

– Matthews Correlation Coefficient (MCC): It measures the classification quality even if the classes are in different sizes, counting true and false positives and negatives as follows:

$$MCC = \frac{TP * TN - FP * FN}{[(TP + FP) * (FN + TN) * (FP + TN) * (TP + FN)]^{1/2}} \tag{6}$$

– Intersection over union (IoU): IoU measures the overlap between 2 boundaries to measure how much the predicted boundary overlaps with the ground truth (the real object boundary). In some datasets, an IoU threshold is predefined (say 0.5) in classifying whether the prediction is a true positive or a false positive.

$$IOU = \frac{AreaofOverlap}{AreaofUnion} \tag{7}$$

– mean Average Precision (mAP): AP (Average precision) is a popular metric in measuring the accuracy of object detectors which quantifies how good our model is at performing a specific query. Given Q which is the number of queries in the set and AveP(q) which is the average precision (AP) for a given query, q, the mAP is calculated as follows:

$$mAP = \frac{\sum_{q=1}^{Q} AveP(q)}{Q} \tag{8}$$

3.2 Datasets Preparation

Seven recent proposed ideas for breast cancer detection and classification are selected to be compared in this section which are recent and considered end to

Table 2. Comparative study for the dataset preparation for the recent work in the breast cancer detection and classification

Ref.	Dataset	Size	Views	Augmentation	Preprocessing
[33]	INbreast	107 M	CC/MLO	Augment mammograms eight times by rotating them with the angles of 45 = ° (i.e., 0°, 45°, 90°, 135°, 180°, 225°, 270° and 315°)	A contrast-limited adaptive histogram equalization (CLAHE) method is utilized as a preprocessing step for all detected masses
[4]	DDSM	600 M	CC/MLO	Rotating mammograms with 90, 180, 270°	Otsu thresholding technique to generate breast region. Then, multi-threshold peripheral equalization enhanced peripherals thickness and remove irrelevant information
[2]	DDSM	150 M	CC/MLO	No data augmentation	Multi-threshold peripheral equalization to enhance peripherals thickness and remove irrelevant information
[34]	BCDR-F03	736 M	CC/MLO	To each refined breast lesion, another seven new samples are generated using a combination of flipping and rotation ($\pi/2$, Π and $3\pi/2$) transformations	No preprocessing
[35]	DDSM + NP	2242 M	CC/MLO	Each ROI was flipped and rotated four times to obtain eight augmented samples	No preprocessing
[36]	DDSM	1820 M	CC/MLO	The selected mammograms augmented by 5 random rotations and sampled 5 random crops for each rotation, thus effectively multiplying the training set size by 25	First extract the mass from the full mammogram by taking a bounding box around the pixel-level mask applied to the original image. Followed by a fixed padding of 50 pixels all around the mass
[37]	INbreast + NP	329 M	CC/MLO	Mammograms were flipped, randomly cropped, and rotated up to 90°, 180°, and 270°	Original mammograms are divided into small sub-sections which do not require the resizing method

end systems as well. They are studied with respect to different aspects including the dataset they used, if they made preprocessing on this data or not, or data augmentation exist and the mammograms views they trained with.

In Table 2: Ref. column is used as the abbreviation for the paper Reference, Dataset column indicates the used dataset, size column is the used dataset size, views column is the views included in the training process which may be Mediolateral Oblique (MLO) views or Cranio-Caudal (CC) views or both, augmentation columns determines if augmentation is done by the proposed method and what was done to augment data and finally preprocessing column is the steps used for preprocessing. Inside the table, M is used instead of mammograms and NP is used instead of Non Public Dataset.

Table 3. Comparative study for the recent proposed work in the breast cancer detection and classification

Ref.	Detection & segmentation	Classification	Classes	Fine tuned	Accuracy
[33]	YOLO is used for mass detection. Then, they proposed a new full resolution convolutional network (FrCN) deep learning model for pixel-to-pixel mass segmentation	AlexNet	B/M	They used the transfer learning method to initialize the parameters of all deep learning models	– Detection accuracy = 98.96% – Classification accuracy = 95.64%
[4]	YOLO is used to detect the mass	YOLO	B/M	Fine-tuned with the pre-trained weights with a large computer vision ImageNet dataset	– Detection accuracy = 99.7% – Classification accuracy = 97%
[2]	Adaptive threshold & morphological operations	Deep Belief Network (DBN)	B/M/N	Not fine tuned	– Overall accuracy = 92.86%
[34]	ROIs are first extracted and then Otsu segmentation algorithm followed by morphological operation to refine the mass	AlexNet, GoogLeNet and Shallow CNN	B/M	Pre-trained on a large-scale visual database	GoogLeNet AUC = 88% & AlexNet AUC = 83%
[35]	No detection and segmentation	DCNN	B/M	Fine-tuned with the pre-trained weights with a large computer vision ImageNet dataset	The multi-task transfer learning DCNN was found to have significantly ($p = 0.007$) higher performance generalization compared to the single-task transfer learning DCNN
[36]	No detection and segmentation	AlexNet, GoogLeNet and Shallow CNN	B/M	Fine-tuned with the pre-trained weights with a large computer vision ImageNet dataset	The GoogleNet with Deeper Training outperforms the other models with a recall of 0.934 compared to at most 0.901 for the other models
[37]	RetinaNet	No classification	No classification	They used ResNet50 that is pre-trained on the ImageNet dataset for the backbone netw	They did a lot of experiments and they achieved higher true positive rate than the complex model which is average 97%

4 Comparative Study for the Applied Detection and Classification Models

The same seven approaches are also compared based on the used detection model if exist, the used classification model if exist, the classes the model trained on if benign (B), malignant (M) or normal (N) and the achieved accuracy as shown in Table 3. In this table: Ref. column is the paper Reference, detection and classification column shows the applied model for mass detection or segmentation

or both, classification columns shows the mass classification model followed by the classes column then the fine tuned columns to check if the model is fine tuned or not and finally the performance in the last accuracy column.

From this study, it is found that the recent and common proposed models for the mass detection and classification are based on AlexNet [37], GoogleNet [38], YOLO [4] and RetinaNet [39]. The most recent and best achieved accuracy work in the breast detection and classification area is [4] and [39].

4.1 Limitation of the Used Detection and Classification Models

Besides the high accuracy achieved using the above models, they have some limitations. AlexNet model causes duplication of data due to the overlapping blocks of pixels which leads to more memory to be used. While GoogleNet used sparse connections (not all the output feature maps are connected to all the input feature maps), to overcome the problem of redundant information and reduced cost by omitting feature maps that were not relevant. However, the main drawback of the GoogleNet was its heterogeneous topology that needs to be customized from one module to another. Besides that its representation bottleneck that drastically reduces the feature space in the next layer and thus sometimes may lead to loss of useful information. On the other-hand the YOLO model is customizable but it does have some limitations on how close objects can be, i.e. it couldn't get good accuracy with the small sized objects and tends to do a few mistakes in localization. This results in introducing the RetinaNet to fill in for the imbalances and inconsistencies of the single shot object detectors like YOLO in a fast and efficient manner.

5 Conclusion

In this paper, we compare between the recent proposed models in the breast cancer detection and classification in order to reduce the death rate caused by the breast cancer. We integrated from the recent published work, the various datasets that are available public such as DDSM, CBIS-DDSM and INbreast. The differences are discussed between these datasets from many points as the size, the views included, the classes contained with the images format as well. Then, an overall comparative study is held between various machine learning based detection and classification methods that are recently published. It is deduced that the common database used and considered large is the CBIS-DDSM which is an enhanced version from the DDSM. It is deduced that YOLO and RetinaNet is the new models that are recently used and considered more simpler than the conventional CNN networks and achieve better results and accurate performance for the mass detection and malignancy classification. Also, the models that use balanced number for the classes the model shall detect, are more fair and achieve better overall accuracy than the unbalanced sized data models.

References

1. Boyle, P., Levin, B., et al.: World cancer report 2008. IARC Press, International Agency for Research on Cancer (2008)
2. Al-antari, M.A., Al-masni, M.A., Park, S.U., Park, J.H., 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. J. Med. Biol. Eng. **38**(3), 443–456 (2017)
3. Al-masni, M., Al-antari, M., Park, J., Gi, G., Kim, T., Rivera, P., Valarezo, E., Han, S.-M., Kim, T.-S.: Detection and classification of the breast abnormalities in digital mammograms via regional convolutional neural network. In: 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2017), Jeju Island, South Korea, pp. 1230–1236(2017)
4. Al-masni, M.A., Al-antari, M., Park, J.-M.P., Gi, G., Kim, T.-Y.K., Rivera, P., Valarezo, E., Choi, M.-T., Han, S.-M., Kim, T.-S.: Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. Comput. Methods Programs Biomed. **157**, 85–94 (2018)
5. Al-antari, M.A., Al-masni, M.A., Park, S.U., Park, J.H., Kadah, Y.M., Han, S.M., Kim, T.S.: Automatic computer-aided diagnosis of breast cancer in digital mammograms via deep belief network. In: Global Conference on Engineering and Applied Science (GCEAS), Japan, pp. 1306–1314 (2016)
6. Al-antari, M.A., Al-masni, M.A., Kadah, Y.M.: Hybrid model of computer-aided breast cancer diagnosis from digital mammograms. J. Sci. Eng. **04**(2), 114–126 (2017)
7. Dromain, C., Boyer, B., Ferré, R., Canale, S., Delaloge, S., Balleyguier, C.: Computed aided diagnosis (CAD) in the detection of breast cancer. Eur. J. Radiol. **82**(3), 417–423 (2013)
8. Dhungel, N., Carneiro, G., Bradley, A.P.: Automated mass detection in mammograms using cascaded deep learning and random forests. In: 2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Australia (2015)
9. Wang, Y., Tao, D., Gao, X., Li, X., Wang, B.: Mammographic mass segmentation: embedding multiple features in vector-valued level set in ambiguous regions. Pattern Recognit. **44**(9), 1903–1915 (2011)
10. Rahmati, P., Adler, A., Hamarneh, G.: Mammography segmentation with maximum likelihood active contours. Med. Image Anal. **16**(9), 1167–1186 (2012)
11. Domínguez, A.R., Nandi, A.: Toward breast cancer diagnosis based on automated segmentation of masses in mammograms. Pattern Recognit. **42**(6), 1138–1148 (2009)
12. Qiu, Y., Yan, S., Gundreddy, R.R., Wang, Y., Cheng, S., Liu, H., Zheng, B.: A new approach to develop computer-aided diagnosis scheme of breast mass classification using deep learning technology. J. X-Ray Sci. Technol. **25**(5), 751–763 (2017)
13. Dhungel, N., Carneiro, G., Bradley, A.P.: A deep learning approach for the analysis of masses in mammograms with minimal user intervention. Med. Image Anal. **37**(1), 114–128 (2017)
14. LeCun, Y., Boser, B.E., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W.E., et al.: Handwritten digit recognition with a back-propagation network. In: Advances in Neural Information Processing Systems, pp. 396–404 (1990)
15. Becker, A.S., Marcon, M., Ghafoor, S., Wurnig, M.C., Frauenfelder, T., Boss, A.: Deep learning in mammography: diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer. Invest. Radiol. **52**(7), 434–440 (2017)

16. Kooi, T., Litjens, G., van Ginneken, B., Gubern-Mérida, A., Sánchez, C.I., Mann, R., et al.: Large scale deep learning for computer aided detection of mammographic lesions. Med. Image Anal. **35**, 303–312 (2017)
17. Dhungel, N., Carneiro, G., Bradley, A.P.: Automated mass detection in mammograms using cascaded deep learning and random forests. In: 2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA), pp. 1–8. IEEE (2015)
18. Dhungel, N., Carneiro, G., Bradley, A.P.: A deep learning approach for the analysis of masses in mammograms with minimal user intervention. Med. Image Anal. **37**, 114–128 (2017)
19. Akselrod-Ballin, A., Karlinsky, L., Alpert, S., Hasoul, S., Ben-Ari, R., Barkan, E.: A region based convolutional network for tumor detection and classification in breast mammography. In: Deep Learning and Data Labeling for Medical Applications, pp. 197–205. Springer (2016)
20. Akselrod-Ballin, A., Karlinsky, L., Hazan, A., Bakalo, R., Horesh, A.B., Shoshan, Y., et al.: Deep learning for automatic detection of abnormal findings in breast mammography. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 321–329. Springer (2017)
21. Ribli, D., Horváth, A., Unger, Z., Pollner, P., Csabai, I.: Detecting and classifying lesions in mammograms with deep learning. Sci. Rep. **8**(1), 4165 (2018)
22. Choukroun, Y., Bakalo, R., Ben-Ari, R., Akselrod-Ballin, A., Barkan, E., Kisilev, P.: Mammogram classification and abnormality detection from nonlocal labels using deep multiple instance neural network (2017)
23. Jalalian, A., Mashohor, S., Mahmud, H., Saripan, M., Rahman, A., Ramli, B., et al.: Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. Clin. Imaging **37**(3), 420–426 (2013)
24. Digital Database for Screening Mammography. http://www.eng.usf.edu/cvprg/Mammography/Database.html
25. Heath, M., Bowyer, K., Kopans, D., Moore, R., Kegelmeyer, W.P.: The digital database for screening mammography. In: Yaffe, M.J. (ed.) Proceedings of the Fifth International Workshop on Digital Mammography, pp. 212–218. Medical Physics Publishing (2001). ISBN 1–930524–00–5
26. Lee, R.S., Gimenez, F., Hoogi, A., Rubin, D.: Curated breast imaging subset of DDSM. Cancer Imaging Arch. **8** (2016)
27. Curated Breast Imaging Subset of DDSM. https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM
28. The Mini–MIAS Database of Mammograms. http://peipa.essex.ac.uk/info/mias.html
29. Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S.: INbreast: toward a full-field digital mammographic database. Acad. Radiol. **19**(2), 236–248 (2012)
30. Orel, S.G., Kay, N., Reynolds, C., Sullivan, D.C.: BI-RADS categorization as a predictor of malignancy. Radiology **211**(3), 845–850 (1999)
31. The INbreast Dataset. http://medicalresearch.inescporto.pt/breastresearch/index.php/Get_INbreast_Database
32. The Breast Cancer Digital Repository. https://bcdr.eu/
33. Al-antari, M.A., Al-masni, M.A., Choi, M.-T., Han, S.-M., Kim, T.-S.: A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. Int. J. Med. Inform. **117**, 44–54 (2018)

34. Jiang, F., Liu, H., Yu, S., Xie, Y.: Breast mass lesion classification in mammograms by transfer learning. In: Proceedings of the 5th International Conference on Bioinformatics and Computational Biology, pp. 59–62. ACM (2017)
35. Samala, R.K., Chan, H.-P., Hadjiiski, L.M., Helvie, M.A., Cha, K.H., Richter, C.D.: Multi-task transfer learning deep convolutional neural network: application to computer-aided diagnosis of breast cancer on mammograms. Phys. Med. Biol. **62**(23), 8894 (2017)
36. Levy, D., Jain, A.: Breast mass classification from mammograms using deep convolutional neural networks. arXiv:1612.00542 (2016)
37. Yuan, Z.-W., Jun, Z.: Feature extraction and image retrieval based on AlexNet. In: Eighth International Conference on Digital Image Processing (ICDIP 2016), vol. 10033, p. 100330E. International Society for Optics and Photonics (2016)
38. Wu, C., Wen, W., Afzal, T., Zhang, Y., Chen, Y.: A compact DNN: approaching GoogleNet-level accuracy of classification and domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5668–5677 (2017)
39. Jung, H., Kim, B., Lee, I., Yoo, M., Lee, J., Ham, S., Woo, O., Kang, J.: Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network. PloS One **13**(9), e0203355 (2018)