

Abnormality Detection in Mammography using Deep Convolutional Neural Networks

Pengcheng Xi
Carleton University and
National Research Council Canada
Ottawa, Ontario, Canada
pengcheng.xi@nrc-cnrc.gc.ca

Chang Shu
Computer Vision and Graphics
National Research Council Canada
Ottawa, Ontario, Canada
chang.shu@nrc-cnrc.gc.ca

Rafik Goubran
Systems and Computer Engineering
Carleton University
Ottawa, Ontario, Canada
goubran@sce.carleton.ca

Abstract—Breast cancer is the most common cancer in women worldwide. The most common screening technology is mammography. To reduce the cost and workload of radiologists, we propose a computer aided detection approach for classifying and localizing calcifications and masses in mammogram images. To improve on conventional approaches, we apply deep convolutional neural networks (CNN) for automatic feature learning and classifier building. In computer-aided mammography, deep CNN classifiers cannot be trained directly on full mammogram images because of the loss of image details from resizing at input layers. Instead, our classifiers are trained on labelled image patches and then adapted to work on full mammogram images for localizing the abnormalities. State-of-the-art deep convolutional neural networks are compared on their performance of classifying the abnormalities. Experimental results indicate that VGGNet receives the best overall accuracy at 92.53% in classifications. For localizing abnormalities, ResNet is selected for computing class activation maps because it is ready to be deployed without structural change or further training. Our approach demonstrates that deep convolutional neural network classifiers have remarkable localization capabilities despite no supervision on the location of abnormalities is provided.

Index Terms—computer aided analysis, computer aided diagnosis, mammography, medical diagnostic imaging, feature extraction, neural networks, image classification

I. INTRODUCTION

According to World Health Organization (WHO), breast cancer is the most common cancer in women both in the developed and the developing world [1]. Moreover, there is an increasing incidence of breast cancer in the developing world because of the increase in life expectancy, urbanization and adoption of western lifestyles. Although some risk reduction can be achieved with prevention, early detection for improving breast cancer outcome and survival remains the cornerstone of breast cancer control [1].

Mammography is the most common breast screening technology. There are several imaging techniques for examining the breast, including ultrasound, magnetic resonance imaging (MRI), X-ray imaging and emerging technologies such as molecular breast imaging and digital breast tomosynthesis (DBT). Mammography is a type of imaging that uses a low-dose X-ray system to examine the breast and is the most reliable method for screening breast abnormalities [28] before they become clinically palpable.

There are two types of examinations in mammography: screening and diagnostic. Screening mammography is for detecting breast cancer in an asymptomatic population while diagnostic mammography is a follow-up exam on patients who have already demonstrated abnormal clinical findings [28]. Screening mammography generally consists of four views, with two views of each breast: the craniocaudal (CC) view and the mediolateral oblique (MLO) view. Besides the two views, additional diagnostic mammography may offer in-depth look at suspicious areas.

One of the challenges in mammography is low contrast in mammogram images. This poses difficulties for radiologists to interpret results. Double reading of mammograms has been advocated to lower the rate of false positives and negatives [30]; however, the cost and workload associated with double reading are high. Therefore, computer aided detection (CADe) and computer aided diagnosis (CADx) of abnormalities in mammography have been introduced. While CADx has not been approved for clinical use, CADe is playing an increasingly important role in breast cancer screening [28] [15].

Computer aided detection is a pattern recognition process that aids radiologists in detecting potential abnormalities such as calcifications, masses, and architectural distortions [6]. It identifies suspicious features in the radiology images and brings them to the attention of radiologists [6]. In its current use, the radiologists first review the exam, activates the CAD software and then re-evaluates the CAD-marked areas of concern before writing the report [6].

Because of the medical significance of screening breast cancer, there has been considerable effort on developing CAD approaches for detecting abnormalities, including calcifications, masses, architectural distortion and bilateral asymmetry [28] [15] [11] [32]. Traditional CAD approaches rely on manually designed image features [28] [15] in detecting subtle yet crucial abnormalities in mammograms. In general, the detection of calcifications followed the procedure of image enhancement, stochastic modelling, frequency decomposition and machine learning; the detection of masses have relied on pixel-based and region-based approaches [28].

Recent advances in deep neural networks have enabled automatic feature learning from large amount of training data, providing an end-to-end solution from feature extraction to

classifier building [13] [26] [27] [9]. Moreover, this learning scheme is robust to dataset noise, making it suitable for detecting abnormalities in mammography.

In this work, we present an abnormality detection approach using deep Convolutional Neural Networks (CNN). Using transfer learning [31], we fine tune pre-trained deep CNNs on cropped image patches of calcifications and masses. After feeding a full mammogram image to input of the CNN tuned on patch images, we compute Class Activation Maps (CAM) for localizing abnormalities [9].

Our contributions are three-fold:

- Significantly leveraged deep CNNs' hierarchical feature extraction capabilities through transfer learning. This enables automatic extraction of features for classifying and localizing calcification and mass in mammograms.
- Compared the performance of state-of-the-art deep CNN architectures by training with a limited dataset without over-fitting.
- Successfully adapted patch-based CNN classifiers to full mammogram images for the localization of abnormalities without segmentation.

II. LITERATURE REVIEW

We review computer-aided approaches to detecting and classifying the two main abnormalities found in screening mammography: micro-calcification (MC) and mass. Most approaches to detecting calcifications follow a similar procedure: image enhancement, segmentation or extracting Region of Interests (ROIs), feature computation and classification. Mass detection algorithms first detect suspicious regions in a mammogram and then classify it as mass or normal tissues.

MCs are tiny deposits of calcium that appear as bright spots in mammograms. Filter banks were used to decompose mammogram images followed by ROI selection and Bayesian classifications [16] [12]. Pal et al. [18] introduced a multi-stage system for detecting MCs in mammograms. They used a back-propagation neural network to find candidate calcification regions first, cleaned network output to remove thin elongated structures and used a measure of local density for final classification. Similarly, Harirchi et al. [8] applied a two-level algorithm for the detection of MCs using diverse-Adaboost-SVM. Six features (four wavelet plus two gray level features) were computed for neural network to detect candidate MC pixels. As a result, 25 features from candidate MCs were extracted and further reduced with geometric linear discriminant analysis (GLDA). The classifier was built with diverse Adaboost SVM. Oliver et al. [17] extracted local features for morphology of MCs and then used a learning approach to select the most salient feature for a boosted classifier. Zhang et al. [33] enhanced the MCs using well-designed filters and then conducted subspace learning for feature selection. A twin SVM (TWSVM) was used for classification.

A mass in mammogram is defined as a space-occupying lesion seen in more than one projection [2]. The general procedure for detecting mass is first to detect suspicious regions, then extract shape and texture features, and finally

detect mass regions through classification or removing false positive regions [15]. Petrosian et al. [20] used texture features to distinguish mass and non-mass regions. Petrick et al. [19] used an adaptive density-weighted contrast enhancement filter to obtain potential masses and used Laplacian Gaussian for edge detection. Morphological features were extracted for classifying normal and mass ROIs. Cascio et al. [5] first segmented the boundary of ROI using an edge-based approach and then computed geometric and shape features. Neural networks were trained to distinguish true mass from normal regions.

While previous classifiers mostly used shallow neural networks, recent years witnessed great advancement on applying deep learning to computer aided detection. Wang et al. [29] introduced ChestX-ray8, a hospital-scale chest X-ray database, and provided benchmarks on weakly-supervised classification and localization of common thorax diseases. They applied deep CNNs and added transition layers to produce heatmap for localization. Following this work, Rajpurkar et al. [22] introduced CheXNet, a 121-layer Dense Convolutional Network (DenseNet) trained on the ChestX-ray 14 dataset, producing radiologist-level pneumonia detection. Moreover, Rajpurkar et al. [21] introduced MURA dataset for detecting radiologist-level abnormality in musculo-skeletal radiographs.

Machine learning has also been widely applied to medical measurements and imaging applications. Rosati et al. [23] used multiparametric MRI along with a clustering procedure based on self-organizing map (SOM) to improve the detection of prostate cancer. Andria et al. [4] investigated the relation between the radiation dose on patient and the resulting image quality, through comparing the tomosynthesis performance with 2D digital mammography. Roza et al. [24] presented an artificial neural network (ANN) and feature extraction methods to identify two types of arrhythmias in ECG signals. Alkabawi et al. [3] proposed an approach for computer-aided classification of multi-types of dementia using convolutional neural networks. The proposed approach outperforms the state-of-the-art CAD methods.

III. METHODOLOGY

Computer-aided mammography is a challenging problem and cannot be treated as an image classification task. The reason is that abnormalities within a whole image are located in small regions. For example, a typical full mammogram with a resolution of 3000x4600 (width and height in pixels) contains an abnormality region of size only about 200x200 (pixels). Training recent deep CNNs requires resizing full images to 224x224 (pixels) at input layer, making it difficult to train and detect abnormalities. To deal with this challenge, we propose training deep CNNs on cropped image patches (labelled ROIs) and adapting them to full mammogram images.

Figure 1 illustrates the data-flow of our approach. With training image patches from calcification and mass cases, a binary classifier is trained with state-of-the-art deep CNN architectures using transfer learning [31]. The pre-trained CNNs are modified at output layers to have two output classes.

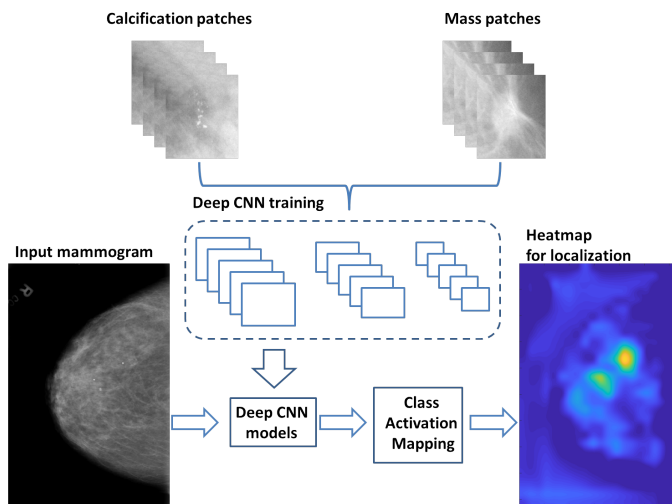


Fig. 1. Diagram of our approach.

The output layers are then fine-tuned while the first part of the network is frozen.

The fine-tuned patch neural network is then used to localize mammographic abnormalities in full-size mammograms. Traditional approaches used the classifier to scan the whole image with a sliding window and therefore have a low efficiency [31]. In contrast, our approach enables localizing abnormalities in one single forward pass. Feeding the full-size mammogram image into the patch classifier and computing class activation mapping [34] near the end of the output layers produces a heatmap for the localization of abnormalities. The computation of CAM is explained in more details at section III-D.

A. Data Selection

In mammography, there is a lack of standard evaluation data and most CAD algorithms are evaluated on private dataset. Most mammographic databases are not publicly available. This poses a challenge to compare performance of methods or to replicate prior results. The most commonly used databases are the Mammographic Image Analysis Society (MIAS) database [7] and the Digital Database for Screening Mammography (DDSM) [10]. MIAS contains left and right breast images for 161 patients. There are 208 normal, 63 benign and 51 malignant images. It also includes radiologist’s ‘truth’-markings on the locations of any abnormalities that may be present. DDSM is the largest mammography dataset that is publicly available. The database contains approximately 2,500 studies, each includes two images of each breast, along with associated patient information and image information. Images containing suspicious areas have associated pixel-level ‘ground truth’ about the locations and types of suspicious regions. Sample mammograms of a patient are shown in Figure 2.

Recently, Lee et al. [14] released an updated and standardized version of the DDSM for the evaluation of CAD systems in mammography. Their dataset, the CBIS-DDSM (Curated Breast Imaging Subset of DDSM), includes decompressed

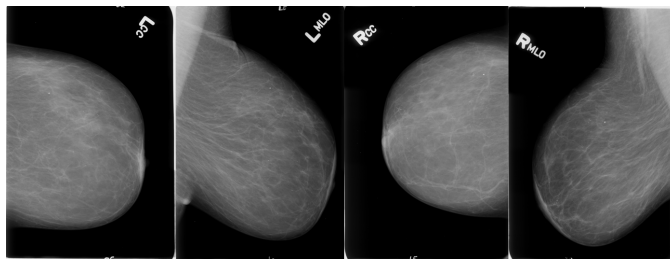


Fig. 2. Mammograms of a patient in DDSM dataset (different views from left to right: left CC, left MLO, right CC and right MLO).

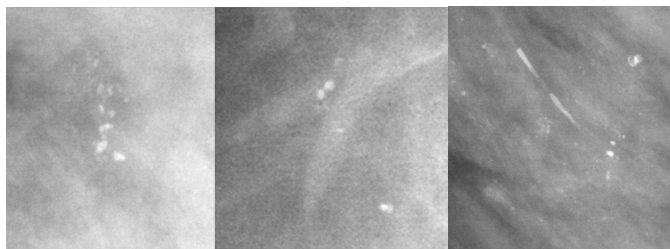


Fig. 3. Sample images of calcification patches in CBIS-DDSM.

images, data selection and curation by trained mammographers, updated mass segmentation and bounding boxes, and pathologic diagnosis for training data. The dataset contains 753 calcification cases and 891 mass cases. Sample image patches are shown in Figure 3 and 4.

We use image patches from CBIS-DDSM for classification and test on full mammograms for localization. We merge the training and testing dataset in CBIS-DDSM and conduct new 85/15 split for training and testing sets. The number of image patches are listed in Table I.

TABLE I
SIZE OF TRAINING AND TESTING IMAGE PATCHES FOR CBIS-DDSM.

Abnormality	Training	Testing	Overall
Calcification	1284	227	1511
Mass	1353	239	1592

B. Data Augmentation

To avoid over-fitting during training, we applied the following data augmentation on the training data: random rotation between zero and 360 degrees, random X and Y reflections. It

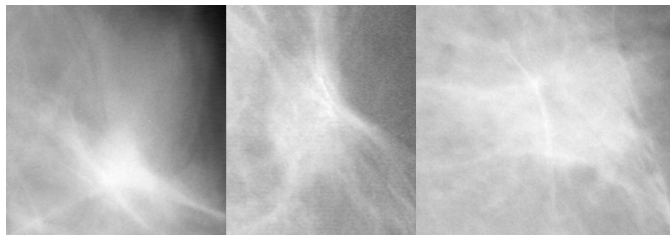


Fig. 4. Sample images of mass patches in CBIS-DDSM.

is based on our observation of the variations within the training and testing dataset.

C. Architectures of Deep CNN

In visual computing, tremendous progress has been made in object classification and recognition thanks to the availability of large scale annotated datasets such as ImageNet Large Scale Visual Recognition Competition (ILSVRC) [25]. The ImageNet dataset contains over 15 million annotated images from a total of over 22,000 categories.

Recent years witnessed great performance advancement on ILSVRC using deep CNNs. Comparing to traditional hand-crafted image features, deep CNNs automatically extract features from a large dataset for tasks they are trained for. In this work, we adapt four of the best-performing models in recent ImageNet challenges and compare their performance on classifying calcification and mass in mammograms.

- AlexNet. In 2012, Krizhevsky et al. [13] entered ImageNet ILSVRC with a deep CNN and achieved top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry. The network was made up of 5 conv layers, max-pooling layers, dropout layers, and 3 fully connected layers. This work led to a series of deep CNN variants in the following years which consistently improved the state-of-the-art in the benchmark tasks.
- VGGNet. In 2014, Simonyan and Zisserman [26] introduced a deeper 19-layer CNN and achieved top result in the localization task of ImageNet ILSVRC. The network used very small 3x3 convolutional filters and showed significant improvement. This influential work indicated that CNNs need to have a deep network of layers in order for the hierarchical feature representations to work.
- GoogLeNet. In 2014, Szegedy et al. [27] introduced a deeper CNN to ILSVRC and achieved top 5 error rate of 6.7%. Instead of sequentially stacking layers, this network was one of the first CNNs that used parallel structures in its architecture (9 Inception modules with over 100 layers in total).
- ResNet. In 2015, He et al. [9] introduced a new 152-layer network architecture and set new records in ILSVRC. ResNet achieved 3.57% error rate in the classification task. The residual learning framework is 8 times deeper than VGGNet but still has lower complexity.

All the deep CNN architectures were designed for a 1000-class classification task. To adapt them to our task, the last three layers were removed from each network. Three new layers (fully connected layer, soft-max layer and classification layer) were appended to the remaining structure of each network. Higher learning rates were set for the newly added fully connected layers so that the first part of each network remains relatively unchanged during training and the newly added layers get fine-tuned on our dataset. Five-fold cross validation is used to train and test the robustness of each architecture.

D. Class Activation Maps

Class Activation Mapping (CAM) is a technique for identifying regions in an image using a CNN for a specific class [34]. In other words, CAM identifies image regions relevant to a class. It allows re-using classifiers for localization purpose, even when no training data on locations are available. It also demonstrates that CNNs have a built-in attention capability.

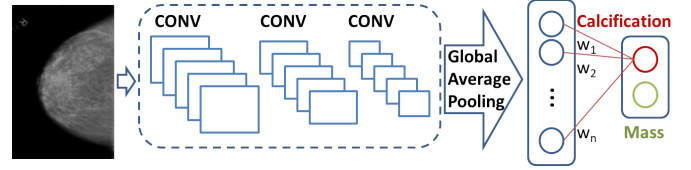


Fig. 5. Class activation mapping for heatmap production.

Computing CAM for mammograms is explained in Figure 5. A deep CNN needs to be cut after the last convolution layer and a global average pooling layer and a fully connected layer are appended. The new model needs to be retrained for learning the weights w_i ($i = 1, 2, \dots, n$) at the output layer. Within the four selected deep CNN architectures, ResNet already has the required architecture and is therefore selected for computing CAM.

A full mammogram is fed into the fine-tuned patch classifier using ResNet. The feature maps from the output of the last convolutional layer are denoted as f_i ($i = 1, 2, \dots, n$). We can identify the importance of the image regions by projecting back the weights of the output layer onto the convolutional feature maps [34] through:

$$CAM = \sum_{i=1}^n w_i f_i \quad (1)$$

The output CAM is then displayed for visualization and verification.

IV. RESULTS

A. Comparison of Different Deep CNN Architectures

We set the following parameters for training each modified deep CNN: Stochastic Gradient Descent with Momentum (SGDM) as the optimization algorithm, batch size of 16, initial learning rate as $1e - 4$, and the learning rate factor for the last fully connected layer as 20.0. Each network stops from further training if the mean accuracy on the fifty most recent batches reaches 99.5% or if the number of epochs reaches maximum setting of 200. All the models are trained on a workstation with an NVIDIA GeForce GTX TITAN X GPU (one hour for AlexNet, eight hours for VGGNet, two hours for GoogLeNet, and four hours for ResNet). The final size of fine-tuned VGGNet is about 20 times that of GoogLeNet, with in-between sizes for AlexNet and ResNet.

Running cross-validation on training and testing datasets and computing mean accuracies across the five folds give the final accuracy results in Table II. VGGNet achieves the highest accuracy for classifying calcifications and GoogleNet receives

TABLE II
MEAN CLASSIFICATION ACCURACIES OF DEEP CNNs

Model	Calcification	Mass	Overall Accuracy
AlexNet	88.81%	93.64%	91.23%
VGGNet	92.42%	92.64%	92.53%
GoogleNet	87.14%	95.06%	91.10%
ResNet	90.22%	93.39%	91.80%

the best performance for classifying masses. The highest overall accuracy is also achieved by VGGNet at 92.53%.

B. Localization Results using Class Activation Mapping

We use the fine-tuned ResNet to compute class activation mapping for localizing abnormalities. The selection is based on the fact that ResNet is ready to be used for computing CAM without further training. Without losing generality, we use one full mammogram image from the calcification class, feed it to ResNet and compute the CAM. The result is shown in Figure 6. The heatmap on the right highlights the location of calcifications found from the input mammogram. The highlighted regions correspond to the calcifications within the full mammogram (best viewed in color).

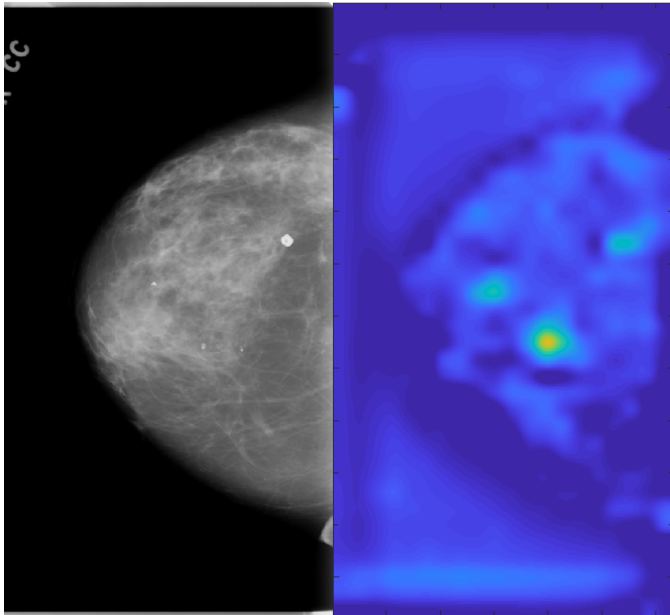


Fig. 6. Calcification localization result (left: full mammogram, right: class activation map output).

Similarly a full mammogram from the mass class is fed into the ResNet for computing the CAM. Results are demonstrated in Figure 7. To add to the comparison, we also include the ground-truth binary mask image provided by the training dataset. The highlighted heatmap region corresponds to the identified abnormality region labelled in the binary mask image.

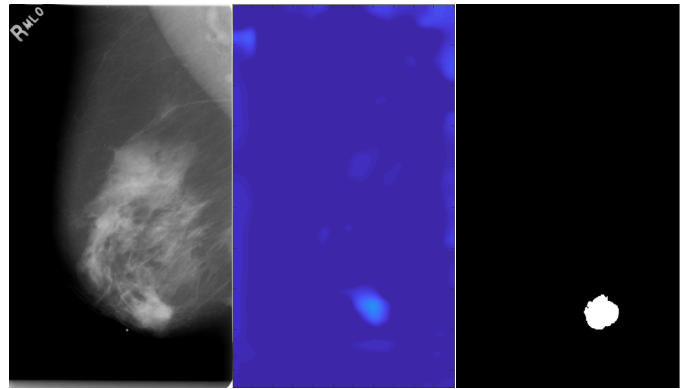


Fig. 7. Mass localization result (from left to right: full mammogram, class activation map output, ground-truth binary mask image).

V. DISCUSSIONS

Because of the low-contrast and noise in mammogram images, it is challenging to train classifiers on calcification and mass cases. Deep neural networks has a limitation on the size of input images (224x224 or 227x227 in pixels). Resizing mammogram images to these sizes will inevitably reduce the quality of images and may also lose the subtle details that are needed for classification. Therefore we propose training classifiers from cropped batch images in order to catch the difference between calcification and mass cases, and apply the trained deep CNN models onto full-size mammogram images. Using a technique called class activation mapping, we successfully reuse the patch classifier for the localization of abnormalities in full mammogram images.

VI. CONCLUSION

We successfully apply deep convolutional neural networks to localizing calcifications and masses in mammogram images without training directly on the full images. This is achieved by conducting the training on cropped image patches through transfer learning and data augmentation. State-of-the-art deep CNN architectures are trained and compared on their performance of classifying the abnormalities. Moreover, we successfully adapt the patch classifier to localizing abnormalities in full mammogram images through class activation mapping.

At the time of preparing this paper, we have found no publications on using CBIS-DDSM; therefore, our results provide a baseline for future studies on improving the performance of detecting calcification and mass in computer-aided mammography. Our future work includes extending the approach to computer aided diagnosis (benign or malignant) using mammograms.

REFERENCES

- [1] "Breast cancer: prevention and control," <http://www.who.int/cancer/detection/breastcancer/en/>, accessed: 2018-02-13.
- [2] "Acr bi-rads-mammography, ultrasound and magnetic resonance imaging," 4th ed., American College of Radiology, 2003.
- [3] E. M. Alkabawi, A. R. Hilal, and O. A. Basir, "Computer-aided classification of multi-types of dementia via convolutional neural networks," in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, May 2017, pp. 45–50.

- [4] G. Andria, F. Attivissimo, A. D. Nisio, A. M. L. Lanzolla, and M. Spadavecchia, "Image quality evaluation of breast tomosynthesis," in *2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, May 2016, pp. 1–6.
- [5] D. Cascio, F. Fauci, R. Magro, G. Raso, R. Bellotti, F. D. Carlo, S. Tangaro, G. D. Nunzio, M. Quarta, G. Forni, A. Lauria, M. E. Fantacci, A. Retico, G. L. Masala, P. Oliva, S. Bagnasco, S. C. Cheran, and E. L. Torres, "Mammogram segmentation by contour searching and mass lesions classification with neural network," *IEEE Transactions on Nuclear Science*, vol. 53, no. 5, pp. 2827–2833, Oct 2006.
- [6] R. Castellino, "Computer aided detection (cad): an overview," *Cancer Imaging*, vol. 5, no. 1, pp. 17–19, 2005.
- [7] J. S. et al, "The mammographic image analysis society digital mammogram database," *Excerpta Medica. International Congress Series*, vol. 1069, pp. 375–378, 1994.
- [8] F. Harirchi, P. Radparvar, H. A. Moghaddam, F. Dehghan, and M. Giti, "Two-level algorithm for mcs detection in mammograms using diverse-adaboost-svm," in *2010 20th International Conference on Pattern Recognition*, Aug 2010, pp. 269–272.
- [9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: <http://arxiv.org/abs/1512.03385>
- [10] M. Heath, K. Bowyer, D. Kopans, R. Moore, and W. P. Kegelmeyer, "The digital database for screening mammography," in *Proceedings of the Fifth International Workshop on Digital Mammography*, 2001, pp. 212–218.
- [11] A. Jalalian, S. B. Mashohor, H. R. Mahmud, M. I. B. Saripan, A. R. B. Ramli, and B. Karasfi, "Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review," *Clinical Imaging*, vol. 37, no. 3, pp. 420 – 426, 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0899707112002938>
- [12] K. Kavitha and N. Kumaravel, "A comparative study of various microcalcification cluster detection methods in digitized mammograms," in *2007 14th International Workshop on Systems, Signals and Image Processing and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services*, June 2007, pp. 405–409.
- [13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>
- [14] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific Data*, vol. 4:170177, 2017/12/19/online.
- [15] Y. Li, H. Chen, L. Cao, and J. Ma, "A survey of computer-aided detection of breast cancer with mammography," *J Health Med Informat*, vol. 7, no. 238, 2016.
- [16] R. Nakayama, Y. Uchiyama, K. Yamamoto, R. Watanabe, and K. Namba, "Computer-aided diagnosis scheme using a filter bank for detection of microcalcification clusters in mammograms," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 2, pp. 273–283, Feb 2006.
- [17] A. Oliver, A. Torrent, X. Lladó, M. Tortajada, L. Tortajada, M. Sentís, J. Freixenet, and R. Zwigellaar, "Automatic microcalcification and cluster detection for digital and digitised mammograms," *Know-Based Syst.*, vol. 28, pp. 68–75, Apr. 2012. [Online]. Available: <http://dx.doi.org/10.1016/j.knosys.2011.11.021>
- [18] N. R. Pal, B. Bhowmick, S. K. Patel, S. Pal, and J. Das, "A multi-stage neural network aided system for detection of microcalcifications in digitized mammograms," *Neurocomputing*, vol. 71, no. 13, pp. 2625 – 2634, 2008, artificial Neural Networks (ICANN 2006) / Engineering of Intelligent Systems (ICEIS 2006). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925231208002269>
- [19] N. Petrick, H.-P. Chan, B. Sahiner, and D. Wei, "An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection," *IEEE Transactions on Medical Imaging*, vol. 15, no. 1, pp. 59–67, Feb 1996.
- [20] A. Petrosian, H.-P. Chan, M. A. Helvie, M. M. Goodsitt, and D. D. Adler, "Computer-aided diagnosis in mammography: classification of mass and normal tissue by texture analysis," *Physics in Medicine and Biology*, vol. 39, no. 12, pp. 2273–2288, 1994.
- [21] P. Rajpurkar, J. Irvin, A. Bagul, D. Ding, T. Duan, H. Mehta, B. Yang, K. Zhu, D. Laird, R. L. Ball, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Ng, "Mura dataset: Towards radiologist-level abnormality detection in musculoskeletal radiographs," *arXiv*, 2017. [Online]. Available: <https://arxiv.org/abs/1712.06957v3>
- [22] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *CoRR*, vol. abs/1711.05225, 2017. [Online]. Available: <http://arxiv.org/abs/1711.05225>
- [23] S. Rosati, V. Giannini, C. Castagneri, D. Regge, and G. Balestra, "Dataset homogeneity assessment for a prostate cancer cad system," in *2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, May 2016, pp. 1–7.
- [24] V. C. C. Roza, A. M. de Almeida, and O. A. Postolache, "Design of an artificial neural network and feature extraction to identify arrhythmias from ecg," in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, May 2017, pp. 391–396.
- [25] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.
- [26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [27] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Computer Vision and Pattern Recognition (CVPR)*, 2015. [Online]. Available: <http://arxiv.org/abs/1409.4842>
- [28] J. Tang, R. M. Rangayyan, J. Xu, I. E. Naqa, and Y. Yang, "Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 2, pp. 236–251, March 2009.
- [29] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," *CoRR*, vol. abs/1705.02315, 2017. [Online]. Available: <http://arxiv.org/abs/1705.02315>
- [30] R. Warren and W. Duffy, "Comparison of single reading with double reading of mammograms, and change in effectiveness with experience," *The British Journal of Radiology*, vol. 68, no. 813, pp. 958–962, 1995, pMID: 7496693.
- [31] P. Xi, R. Goubran, and C. Shu, "Cardiac murmur classification in phonocardiograms using deep convolutional neural networks," in *Proceedings of the International Conference on Pattern Recognition and Artificial Intelligence*, 2018.
- [32] N. Zemmal, N. Azizi, and M. Sellami, "Cad system for classification of mammographic abnormalities using transductive semi supervised learning algorithm and heterogeneous features," in *2015 12th International Symposium on Programming and Systems (ISPS)*, April 2015, pp. 1–9.
- [33] X. Zhang and X. Gao, "Twin support vector machines and subspace learning methods for microcalcification clusters detection," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 5, pp. 1062 – 1072, 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0952197612000917>
- [34] B. Zhou, A. Khosla, L. A., A. Oliva, and A. Torralba, "Learning Deep Features for Discriminative Localization." *CVPR*, 2016.