

Zoom in to where it matters: a hierarchical graph based model for mammogram analysis

Abstract

In clinical practice, human radiologists actually review medical images with high resolution monitors and zoom into region of interests (ROIs) for a close-up examination. Inspired by this observation, we propose a hierarchical graph neural network to detect abnormal lesions from medical images by automatically zooming into ROIs. We focus on mammogram analysis for breast cancer diagnosis for this study. Our proposed network consist of two graph attention networks performing two tasks: (1) node classification to predict whether to zoom into next level; (2) graph classification to classify whether a mammogram is normal/benign or malignant. The model is trained and evaluated on INbreast dataset and we obtain comparable AUC with state-of-the-art methods.

Introduction

Recently, graph convolutional network (GCN) and its variants have generated considerable recent research interest in learning graph representations (Mamalet and Garcia 2012). Applications of GCN and its variants have demonstrated new state-of-the-art results in various domains, such as applied chemistry, social network, citation network, computer vision and natural language processing (Kipf and Welling 2016; Liao et al. 2019; Mamalet and Garcia 2012; Chen, Ma, and Xiao 2018; Yao, Mao, and Luo 2019). Additionally, GCNs and their variants generated considerable recent research interest in medical imaging field. In 2018, Parisot et al employed GCNs to predict disease using a graph where nodes present individuals and features consist of both image and non-image data. GCNs are used as a semi-supervised method to train on the labelled node and infer the labels of the unknown nodes, based on the graph structure and both image and non-image features (Parisot et al. 2018). Shin et al. demonstrated combining both CNNs and GCNs to perform deep vessel segmentation (Shin et al. 2019). CNNs are used to generate features and vessel probabilities while GCNs are employed to predict the presence of a vessel. By combining the outputs of both CNNs and GCNs, the model generates the final segmentation.

Copyright © 2020, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

We further observed that, in clinical practice, human radiologists actually review medical images with high resolution monitors and zoom into region of interests (ROIs) for a close-up examination. Inspired by this observation, we propose a hierarchical graph neural network to detect abnormal lesions from medical images by automatically zooming into ROIs. In this work, we focus on breast cancer detection in mammogram analysis.

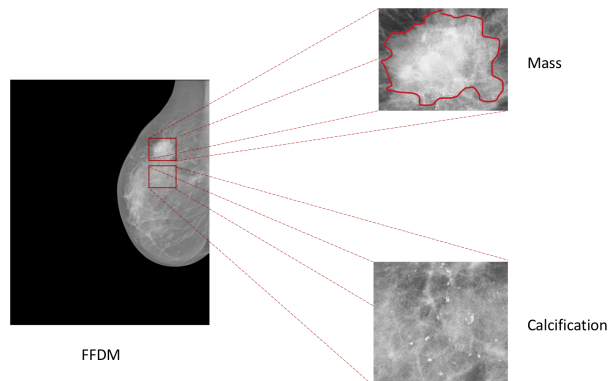


Figure 1: An example of FFDM image in the INbreast dataset. Two groups of lesions are identified by zooming into the specific patches on the right side. Mass is represented by red contours and microcalcifications are small, gray, rounded bright regions in the breast tissue.

Mammography is currently the most effective tool for breast cancer screening and early detection of the disease (Misra et al. 2010). Breast screening for one patient includes the mammography images of two views for each breast: the craniocaudal (CC) view, which is a top to bottom view, and a mediolateral oblique (MLO) view, which is a side view. Currently, the examination of mammograms mainly relies on human radiologists and facing the limitations of high false-positive rate, inconsistent interpretation accuracy and huge workload. (Noble et al. 2009; Elmore et al. 1998; Brodersen and Siersma 2013; McCann, Stockton, and God-

ward 2002). Screening mammogram classification is challenging due to the difficulties in detection of lesions especially masses and microcalcifications. As demonstrated in Figure 1, the average dimension of a full field digital mammogram (FFDM) is around 4000×3000 . However, the dimensions of mass lesions are typically less than 100×100 (Shen et al. 2019) and the sizes of microcalcifications are at most 14 pixels (Zhang et al. 2019). In clinical settings, the American College of Radiology recommends that all FFDMs should be viewed at their full acquisition resolution (of Radiology and others 2010). Radiologists review the FFDM images using high resolution monitors and zoom in for close-ups of regions of interests (ROIs).

Recently, deep learning methods have demonstrated big improvements on automated analysis of mammograms (Aboutalib et al. 2018; Kim et al. 2018; Hamidinekoo et al. 2018; Burt et al. 2018; Kooi et al. 2017; Agarwal et al. 2019; Ribli et al. 2018; Shen et al. 2019). The challenges of lesion dimensions are addressed using established object detection and anomaly detection methods (Ribli et al. 2018; Zhang et al. 2019) or fully image classification methods on resized images with pre-trained models (Shen et al. 2019). In this study, we propose a graph neural network (GNN) based method for end-to-end breast cancer detection in FFDM images. We use hierarchical graph based method to model the zoom-in mechanism of radiologists' operations. The proposed method is able to automatically zoom into the lesion ROIs and detect breast cancer based on overall graph hierarchy and specific ROIs in the mammograms.

Methods

Problem Definition

The mammography dataset is represented by $\mathcal{D} = \{(x^{(j)}, l^{(j)}, y^{(j)})_i\}$ where $j \in \{left, right\}$, indicating the left and right breast of a patient in dataset. $x^{(j)} = \{x_{cc}^{(j)}, x_{mlo}^{(j)}\}$ represents the CC and MLO views of each side of a patient. $l^{(j)} = \{l_{cc}^{(j)}, l_{mlo}^{(j)}\}$ represents the segmentation labels of lesions (mass and calcification): $\Omega \rightarrow \{0, 1, 2, 3, 4\}$, where 0 indicates healthy or normal, 1 indicates benign calcification, 2 indicates benign mass, 3 indicates malignant calcification and 4 indicates malignant mass. $y \in \{0, 1\}$ represents the class label of a mammogram, where 0 indicates normal (BI-RADS $\in \{1, 2, 3\}$) and 1 indicates malignant (BI-RADS $\in \{4, 5, 6\}$).

We use a graph instance to model the zooming operations in a mammogram $(x^{(j)})_i$, which is denoted as $(g^{(j)})_i = (V^{(j)}, A^{(j)}, X^{(j)})_i$. $(V^{(j)})_i$ is the node set and $|V_i^{(j)}| = n_i^{(j)}$, which indicates the number of zoomed regions in the mammogram. $A_i^{(j)}$ is an $n_i^{(j)} \times n_i^{(j)}$ matrix representing the connectivity in $(g^{(j)})_i$ and $(X^{(j)})_i \in n_i^{(j)} \times m \times m$ is a matrix recording the zoom-in regions of all nodes in g . The original mammogram dataset is transformed into a graph instance dataset $G = \{(g^{(j)}, l^{(j)}, y^{(j)})_i\}$. The zoom-in operation in this study refers that in a down-sized image, for a small region $M_{w \times h}$, we find the corresponding region in the original image and resize it to a larger image $M'_{w' \times h'}$, where

$w' \times h' > w \times h$. M' includes more information from original image compared to M .

Network Design

In this paper, we study the mammogram classification as a graph classification problem. The graph instances and their connections are model as a hierarchical graph, as demonstrated in Figure 1. The number of zooming levels in the hierarchical graph is denoted as R . Top node A represents the full mammogram image. The levels after are created by zoom-in operations. Let us denote the graphs truncated at different zoom-in levels by G^1, G^2, \dots, G^R . At level r , graph is defined by its adjacency matrix $A_r \in \mathbb{R}^{N_r \times N_r}$. Its feature matrix is defined as $X_r \in \mathbb{R}^{N_r \times D \times D}$ (features on the nodes are cropped zoom-in regions from mammogram images, resized to $D \times D$). A convolutional neural network (CNN), pre-trained on lesion patches is used to extract feature vectors $H_r \in \mathbb{R}^{N_r \times H}$ from original features X_r . A graph attention network (GAT_{node}) is used to classify node into two classes: to zoom into next level and not to zoom into next level. The output of r th level in the hierarchical graph is:

$$P_r = \begin{cases} 1 & r = 1 \\ \text{softmax}(GAT_{node}(A_r, CNN_{node}(X_r))) & 1 < r \leq R \end{cases} \quad (1)$$

The elements in $P_r \in \mathbb{R}^{N_r \times 2}$ gives the probability zooming into the next level in the hierarchical graph. For (i, j) th elements in P_r , if the prediction is to zoom into the next level, $S \times S$ nodes will be generated in the next level and mapped to (i, j) th node. The features on this node is divided into a $S \times S$ grid, resized to the same dimension $D \times D$ and assigned to the generated nodes in the next level. After the operation is performed on all nodes in r th level, there are K nodes predicted to be zoomed into the next level. G^{r+1} and X^{r+1} will generated accordingly with $N_{r+1} \leftarrow N_r + K \times S^2$, $A_{r+1} \in \mathbb{R}^{N_{r+1} \times N_{r+1}}$ and $X_{r+1} \in \mathbb{R}^{N_{r+1} \times D \times D}$.

At the final zoom level R , an attention based GNN (GAT_{graph}) is used to perform graph classification to classify the mammogram into normal/benign or malignant as follows:

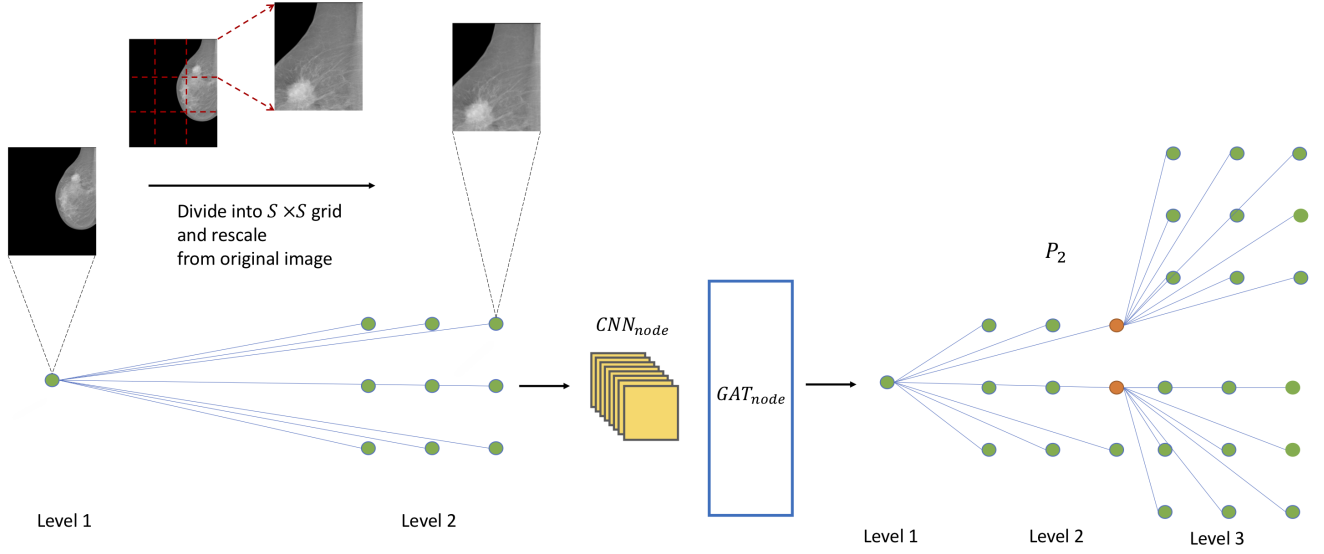
$$\begin{aligned} H_{graph}^R &= GAT_{graph}(A_R, CNN_{graph}(X_R)) \\ \hat{Y} &= \text{softmax}(H_{graph}^R W) \end{aligned} \quad (2)$$

The objective function for graph classification is defined as cross entropy loss between prediction and mammogram label:

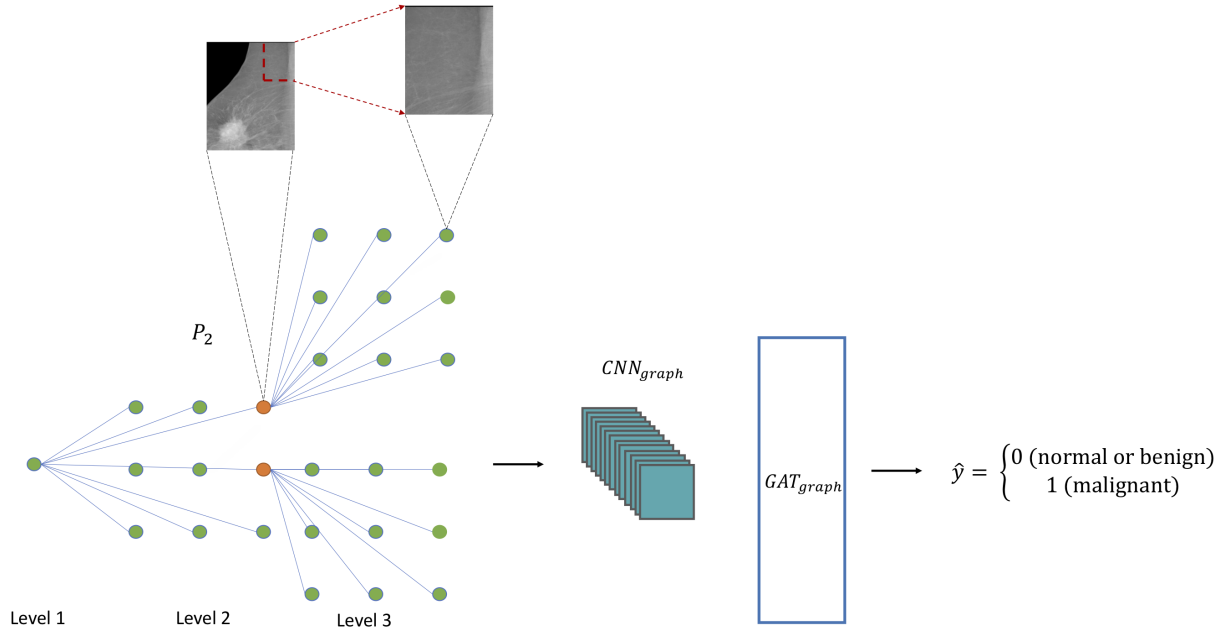
$$\mathcal{L}_{graph} = \frac{1}{|\mathcal{D}|} \sum_{i,j} y_i^{(j)} \log \hat{y}_i^{(j)} + (1 - y_i^{(j)}) \log(1 - \hat{y}_i^{(j)}) \quad (3)$$

In addition to \mathcal{L}_{graph} , we construct node loss from zoom labels to supervise the zoom-in operation in node classification network:

$$\mathcal{L}_{node} = \frac{1}{|\mathcal{D}|} \frac{1}{N_R} \sum_{i,j,v} z_{i,v}^{(j)} \log p_{i,v}^{(j)} + (1 - z_{i,v}^{(j)}) \log(1 - p_{i,v}^{(j)}) \quad (4)$$



(a) A demonstration of zoom-in mechanism in level $r = 2$: each node in level r represents a grid region in the connected node from previous level. CNN_{node} and GAT_{node} are applied to perform node classification. The nodes labelled in orange color are predicted to be zoomed in $r + 1$ level. G_r grows to G_{r+1} accordingly.



(b) When $r = R$, full graph is fed into CNN_{graph} and GAT_{graph} for preform graph classification into binary classes.

Figure 2: An illustration of proposed network

Zoom labels are obtained from lesion segmentation label $(l^{(j)})_i$. For a node v in final hierarchical graph G_R , the zoom label z_v is obtained by from the lesion segmentation label of zoom-in region, $l^{(j)}_{i,v}$, corresponding to node v . We take the

maximum label of $l^{(j)}_{i,v}$. If the maximum label is malignant,

the zoom label $z_{i,v}^{(j)}$ is defined to 1, else 0:

$$z_{i,v}^{(j)} = \begin{cases} 1 & \text{if } \max(l_{i,v}^{(j)}) \in \{0, 1, 2\} \\ 0 & \text{if } \max(l_{i,v}^{(j)}) \in \{3, 4\} \end{cases} \quad (5)$$

Experiments and Performance Evaluation

Datasets and Baseline Algorithms: The proposed method is evaluated using the INbreast (Moreira et al. 2012) dataset. INbreast is a publicly available database with 116 cases comprising 410 FFDM images. INbreast dataset contains the BI-RADS assessment categories (Orel et al. 1999) on each mammogram. In addition, INbreast contains pixel-level lesion segmentation labels on each lesion detected by radiologists. The model performance is evaluated using area under receiver operating characteristic curve (AUC) (Huang and Ling 2005). We compare the performance of the proposed model with state-of-the-art models on INbreast dataset, developed by Ribli et al. and Shen et al. (Ribli et al. 2018; Shen et al. 2019).

Experiment Setup: The experiments are carried out with 80% of the cases for training and 20% for testing. The maximum number of zoom-in levels R is selected to be 2. Zoom-in resize dimension D is selected to be 224. The feature extraction CNN is defined as a VGG like network pre-trained on lesion patches, as demonstrated in (Nikulin 2017; Shen et al. 2019). Zoom-in $S \times S$ grid is selected to be 3×3 . The code and model of the experiment will be publicly available online for reproduction of this work.

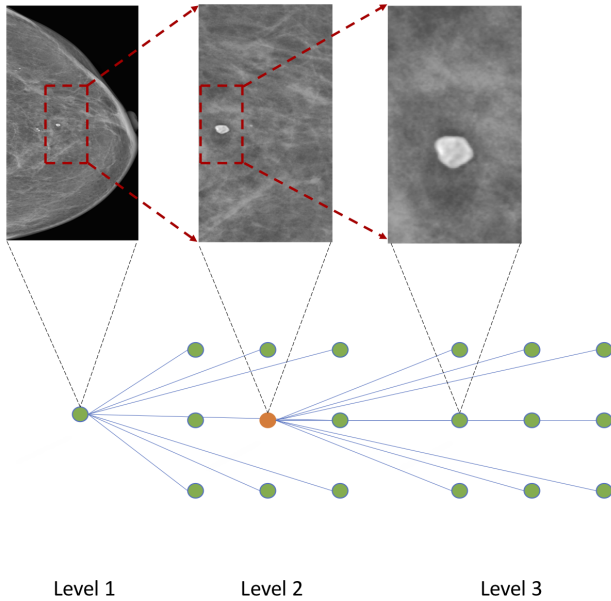


Figure 3: An example data and its result from testing dataset. The images are features assigned to nodes accordingly. In level 2, the node in orange color is predicted to be zoomed in the next level and the corresponding region contains a cluster of malignant calcifications.

Performance Analysis: Our proposed method obtained AUC 0.943 in mammogram classification into normal/benign breast or malignant breast cancer. One example from testing set was selected for demonstration purpose in Figure 3. In Figure 3, there are several calcifications (tiny, bright, grey dots marked) in the mammogram and the proposed model is able to capture the calcification and successfully zoom into the specific region in the next two levels. Compare with the state-of-the-art models by (Shen et al. 2019) with AUC 0.95 and (Ribli et al. 2018) with AUC 0.95, our proposed method produces similar results. The advantages of our model are: firstly, our model improves model interpretability by adding in zoom-in mechanism. The model can highlight the ROIs of lesions for clinicians both as a reminder in case they may have missed them or as a confirmation of their diagnosis. Secondly, our method collapses pixel-level lesion segmentation to zoom label. Compared to state-of-the-arts methods using pixel-level labels, our method is more robust to annotation errors.

Limitations and future work: There are a number of areas that we plan to continue working on to improve our model: firstly, the INbreast dataset is a relatively small dataset to demonstrate the capabilities of the method. Inspired by the achieved promising results, we plan to move on further validate the proposed algorithm locally at our own hospital with the ultimate goal to deploy it as an decision support tool for breast cancer screening. Secondly, the node classification network is difficult to be optimized using cross entropy loss on the zoom label. We plan to further investigate the problem and design a loss to better supervise the zoom-in mechanism in further studies. Lastly, we are now formally evaluating the robustness of our method against annotation noise in the actual clinical settings, and we will report our findings in our future work.

Acknowledgements

This research was supported by the National Research Foundation Singapore under its AI Singapore Programme [Award No. AISG-GC-2019-002] and Health Service Research Grant HSRG-OC17nov004.

References

- Aboutalib, S. S.; Mohamed, A. A.; Berg, W. A.; Zuley, M. L.; Sumkin, J. H.; and Wu, S. 2018. Deep learning to distinguish recalled but benign mammography images in breast cancer screening. *Clinical Cancer Research* 24(23):5902–5909.
- Agarwal, R.; Diaz, O.; Lladó, X.; Yap, M. H.; and Martí, R. 2019. Automatic mass detection in mammograms using deep convolutional neural networks. *Journal of Medical Imaging* 6(3):031409.
- Brodersen, J., and Siersma, V. D. 2013. Long-term psychosocial consequences of false-positive screening mammography. *The Annals of Family Medicine* 11(2):106–115.
- Burt, J. R.; Torosdagli, N.; Khosravan, N.; RaviPrakash, H.; Mortazi, A.; Tissavirasingham, F.; Hussein, S.; and Bagci, U. 2018. Deep learning beyond cats and dogs: recent advances in diagnosing breast cancer with deep neural networks. *The British journal of radiology* 91(1089):20170545.
- Chen, J.; Ma, T.; and Xiao, C. 2018. Fastgcn: fast learning with graph convolutional networks via importance sampling. *arXiv preprint arXiv:1801.10247*.
- Elmore, J. G.; Barton, M. B.; Mocerri, V. M.; Polk, S.; Arena, P. J.; and Fletcher, S. W. 1998. Ten-year risk of false positive screening mammograms and clinical breast examinations. *New England Journal of Medicine* 338(16):1089–1096.
- Hamidinekoo, A.; Denton, E.; Rampun, A.; Honnor, K.; and Zwiggelaar, R. 2018. Deep learning in mammography and breast histology, an overview and future trends. *Medical image analysis* 47:45–67.
- Huang, J., and Ling, C. X. 2005. Using auc and accuracy in evaluating learning algorithms. *IEEE Transactions on knowledge and Data Engineering* 17(3):299–310.
- Kim, E.-K.; Kim, H.-E.; Han, K.; Kang, B. J.; Sohn, Y.-M.; Woo, O. H.; and Lee, C. W. 2018. Applying data-driven imaging biomarker in mammography for breast cancer screening: preliminary study. *Scientific reports* 8(1):2762.
- Kipf, T. N., and Welling, M. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.
- Kooi, T.; Litjens, G.; Van Ginneken, B.; Gubern-Mérida, A.; Sánchez, C. I.; Mann, R.; den Heeten, A.; and Karssemeijer, N. 2017. Large scale deep learning for computer aided detection of mammographic lesions. *Medical image analysis* 35:303–312.
- Liao, R.; Zhao, Z.; Urtasun, R.; and Zemel, R. S. 2019. Lanczosnet: Multi-scale deep graph convolutional networks. *arXiv preprint arXiv:1901.01484*.
- Mamalet, F., and Garcia, C. 2012. Simplifying convnets for fast learning. In *International Conference on Artificial Neural Networks*, 58–65. Springer.
- McCann, J.; Stockton, D.; and Godward, S. 2002. Impact of false-positive mammography on subsequent screening attendance and risk of cancer. *Breast Cancer Research* 4(5):R11.
- Misra, S.; Solomon, N. L.; Moffat, F. L.; and Koniaris, L. G. 2010. Screening criteria for breast cancer. *Advances in surgery* 44(1):87–100.
- Moreira, I. C.; Amaral, I.; Domingues, I.; Cardoso, A.; Cardoso, M. J.; and Cardoso, J. S. 2012. Inbreast: toward a full-field digital mammographic database. *Academic radiology* 19(2):236–248.
- Nikulin, Y. 2017. Digital mammography dream challenge: Participant experience 1 (conference presentation). In *Medical Imaging 2017: Computer-Aided Diagnosis*, volume 10134, 101344J. International Society for Optics and Photonics.
- Noble, M.; Bruening, W.; Uhl, S.; and Schoelles, K. 2009. Computer-aided detection mammography for breast cancer screening: systematic review and meta-analysis. *Archives of gynecology and obstetrics* 279(6):881–890.
- of Radiology, A. C., et al. 2010. Practice guideline for determinants of image quality in digital mammography. *Available at: (Accessed January 5, 2010)*.
- Orel, S. G.; Kay, N.; Reynolds, C.; and Sullivan, D. C. 1999. Bi-rads categorization as a predictor of malignancy. *Radiology* 211(3):845–850.
- Parisot, S.; Ktena, S. I.; Ferrante, E.; Lee, M.; Guerrero, R.; Glocker, B.; and Rueckert, D. 2018. Disease prediction using graph convolutional networks: Application to autism spectrum disorder and alzheimer’s disease. *Medical image analysis* 48:117–130.
- Ribli, D.; Horváth, A.; Unger, Z.; Pollner, P.; and Csabai, I. 2018. Detecting and classifying lesions in mammograms with deep learning. *Scientific reports* 8(1):4165.
- Shen, L.; Margolies, L. R.; Rothstein, J. H.; Fluder, E.; McBride, R.; and Sieh, W. 2019. Deep learning to improve breast cancer detection on screening mammography. *Scientific reports* 9(1):1–12.
- Shin, S. Y.; Lee, S.; Yun, I. D.; and Lee, K. M. 2019. Deep vessel segmentation by learning graphical connectivity. *Medical image analysis* 58:101556.
- Yao, L.; Mao, C.; and Luo, Y. 2019. Graph convolutional networks for text classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 7370–7377.
- Zhang, F.; Luo, L.; Sun, X.; Zhou, Z.; Li, X.; Yu, Y.; and Wang, Y. 2019. Cascaded generative and discriminative learning for microcalcification detection in breast mammograms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 12578–12586.