Early Screening of Breast Cancer on Calcification Abnormality from Mammography using Deep Convolutional Neural Networks

Lam Kin Ho
Department of Computer Science
The Chinese University of Hong Kong
khlam1@cse.cuhk.edu.hk

Leung Ka Ka
Department of Computer Science
The Chiense University of Hong Kong
kkleung1@cse.cuhk.edu.hk

Li Wing Lok
Department of Computer Science
The Chiense University of Hong Kong
wlli1@cse.cuhk.edu.hk

Abstract

Breast cancer is one of the most frequent kinds of cancer in women across the world. Mammography is a commonly used for early screening of breast cancer. Calcification abnormality is early signal of breast cancer. However due to its nature, many noise is presented in the mammography, a significant number of case are misdiagnosed. In this work, we present a deep Convolutional Neural Network that can be used to aid screening process without any prior annotation or segmentation from clinic experts. Transfer learning was used to overcome inadequate training data. We obtain satisfactory results on the CBIS-DDSM dataset, allowing for easy and rapid application in real clinical data.

1 Introduction

Breast cancer is one of the most common types of cancer among women worldwide. Early detection of breast cancer is crucial for successful treatment and improved patient outcomes. Medical imaging, particularly mammography, is a widely used screening tool for breast cancer detection . Mammography screening reduce 41% risk of dying withn 10 years due to breast cancer, and 25% of advanced breast cancers [1]. However, the interpretation of mammograms can be challenging. In US community practice, 1 out of 8 patitent missed the interpretation [2]. There is a need for more accurate and efficient methods to aid in breast cancer detection.

The University of South Florida collaboratively maintained public dataset Digital Database for Screening Mammography (DDSM) [3]. Which has been widely used in many breast cancer related study. However the file format is difficult to use in modern environment. Therefore in our experiments, we use the Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) [4]. Certain preprocessing and decompression work has been done with CAD to convert the image into standard JPEG file format.[5]

We aim to leverage the CBIS-DDSM dataset to develop data analytics approaches that can aid in breast cancer detection in early screening. Convolutional Nerual Networks (CNN) achieved a impressive results on computer vision task, especially on classification[6]. Other than general image classification, CNN also having appericating performance on medical image[7, 8]. Our goal is to develop accurate and efficient methods with machine learning for breast cancer detection that can assist radiologists in their clinical practice.

Preprint. Under review.

Figure 1: Benign(Left) and Malignant(Right) sample image from training set

Figure 2: Benign(Left) and Malignant(Right) sample image from testing set

2 Related work

Even medical images mostly in grayscale and having significant difference from natural image. CNN architecture still received a great success in diagnosis[7, 8]. And many of the work has reached the state of the art[9]. Proving the performance of deep learning-based approaches with CNN.

Other than gerenal medical image identification, various model has been applied and tested on multiple breast dataset[10–13]. Many of them already reaching more than 90% of accuracy. However, some of them are focusing on other breast image rather than mammograms. While some other required marking on interest area before apply to the model. Some may focusing on mass abnormality which already developed to certain size, resulting in higher chance of malignant. To the best of our knowledge, ours is the first work to screening the cancer by calcification abnormality using CNN architecture.

3 Data

3.1 Overview

The CBIS-DDSM dataset is standarized version of the DDSM[4]. In the database, 2620 scanned film mammography studies is labeled as "Benign", "Benign without call back", and "Malignant". As well as some other detailed annotations, including type, location, size of abnormality etc.. Each of the image is in JPEG format with a resolution of around 3000(W) $\times$ 4000(H) pixels. The mammography containing both mediolateral oblique (MLO) and cranio caudal (CC) views of each side of chest. The MLO views is taken from the center of the chest outward and pressing the breast sideways. While the CC views is taken from above of the chest and pressing the breast from both up and down. Example of raw image from the training and testing dataset are shown in Fig. 1 and Fig. 2 respectively.

3.2 Statistic

The dataset consist 1546 and 326 image of calcification abnormality in training and testing set respectively. While in each set, left or right breast view and CC or MLO view in approximately 50%. In the training set, there is 35% of Malignant, 34% of benign, and 31% of benign without call back. While in the testing set, there is 40% of the malignant, 40% of benign, and 21% of benign without call back. Other detail in the dataset, for example, breast density and calcification type are omitted because we aim to develop a easy-to-use model without any information other than mammography itself.

3.3 Preprocessing

Most of the preprocessing job has been already done when creating CBIS-DDSM from DDSM[5]. Including data cleaning, file format conversion, and reannotation mis-labeled data. The image data is in grayscale with only one channel, a direct copy is applied to create a 3 channel image. Before feeding the image into model, a normalization with mean 0.1959 and standard deviation 0.2591 is applied to all image. The normalization value is calculating from all data in the training set. We also count benign and benign without call back as same category as we want to screen on malignant only.

4 Method

4.1 Baseline model

We use the same baseline model as Daniel and Arzav in previous breast cancer detection model[12]. The model is made up of 3 convolutional blocks with 3×3 Convolutions - Batch Norm - ReLU - Max Pooling. The convolution layer having stride of 1 and each of $32, 32, 64$ channels. While the all of the Max Pooling layers are configured as kernel size and stride 2. After convolutional blocks, the output of last Max Pooling layer is connected to 3 fully-connected layers with sizes of $128, 64$, and 2. Finally the class is connected to a CrossEntropy loss function in Pytorch[14], which have a softmax layer before calculating the cross entropy. We use Adam optimizer[15] with learning rate of 10^{-3}, batch size 64, and weight decay of 10^{-5} as regulation term.

4.2 Transfer Learning

Although medical image differ from natural image. A pretrained model based on natrual image dataset, ImageNet1K[16], is proven having faster inference time compared to model trained from scratch[17]. Pretraining on natural image is also useful to increase accuracy on small medical dataset[18]. Consider the small dataset in our experiment (around 1500 images), fine-tuning on a pre-train model is incorporated.

4.3 Augmentation

Data augmentation is usually included in experiment of small dataset. We apply random rotation, horizontal flipping, and cropping to increase the size effectively of our relatively small dataset. Each image is first resize to 256×256 image, following by rotation from -5 to $+5$ degree. A mirroring transformations is applied with probability of 0.5. We also included a random resize crop to 224×224.

4.4 Model Selection

We conduct the experiment using pretrain model avaliable on Pytorch[14], including AlexNet[6],VGG13 with batch normalization(VGG13BN)[19], and ResNet18[20]. The pretrained model weight on ImgeNet1K is also avaliable on PyTorch[14, 16]. We train all the model, including the baseline model, under same hyper-parameter and compare the result on present of data augmentation work.

4.5 Implementation Detail

We train all of the model with Adam update algorithm[15], base learning rate of 10^{-4}, batch size 64, and weight decay of 10^{-5}. A cosine annealing learning rate scheduler[21] and warm up[22] strategy is used. The implementation of warm up scheduler is avaliable at github[23]. We set total number of epoch to 200 with patience of 20 epoch. If the valid accuracy have no improvement after 20 epoch, we early stop of our training process. All of our experiment is conducted on a single Nvidia RTX 3090 Ti

5 Result and Analysis

We compare our result with respect to the baseline architecture. Then conduct a quantitative analysis of our best model. All of the experiment is implemented using Pytorch[14]. We also create a validation set using 20% data from training set. The number of mammography of training, validation, testing set is 1237, 309, 326 respectively. The result of our experiment is shown in Table 1

Table 1: Result of Experiment on different model

Model	Final Train Acc	Final Valid Acc	Best Valid Acc	Best Epoch
Without Augmentation				
Baseline	97.993%	65.210%	66.958%	15
AlexNet	78.496%	74.110%	77.670%	17
VGG13BN	98.6719%	75.407%	78.101%	11
ResNet18	99.218%	78.890%	81.604%	19
With Augmentation				
Baseline	99.084%	62.238%	62.937%	26
AlexNet	62.975%	72.168%	74.757%	49
VGG13BN	84.293%	68.706%	70.280%	25
ResNet18	90.622%	74.340%	74.416%	23

Table 2: Performance of ResNet18 without Data Augmentation on testing set

Accuracy	Precision	Recall	AUC
71.17%	75.73%	59.20%	70.11%

5.1 Analysis

Transfer Learning Transfer Learning provide a more efficient way to train the model and overcome the difficulty of relatively small dataset. In our experiment, we train the base model from scratch and fine-tuning on pretrain model. The pretrain model tends to obtain higher accuracy in smaller epoch.

Data augmentation We would like to overcome the common limitation of most medical dataset, relatively small number of sample by data augmentation. However in our experiment, model without augmentation tends to perform better compared to the same model with augmentation. The reason behinds may be the augmentation method is not suitable for whole mammography. Such augmentation method may work for cropped or segmented image data only[12].

5.2 Performance

We choose ResNet18 without data augmentation to be our final model as it outperform all others model in both final and best validation accuracy. The performance of our model is concluded in Table 2. Confusion matrix and receiver operating characteristic(ROC) curve is provided in Fig.3to visualize the performance as well.

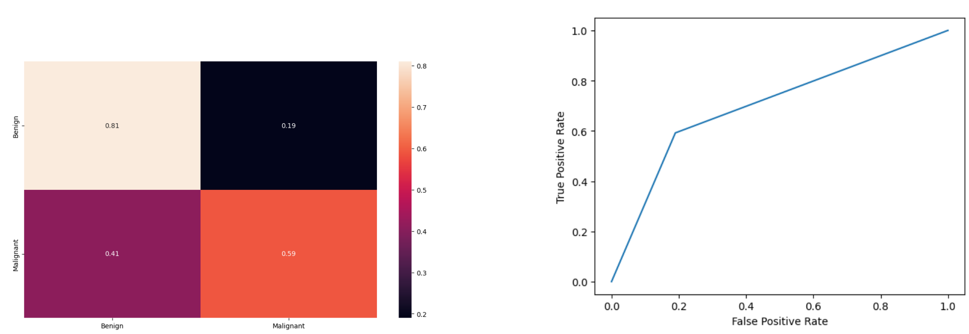

Figure 3: Confusion Matrix(Left) and ROC(Right) of ResNet18 on testing set

6 Conclusion

In our work, we proposed a deep learning model to early detected the breast cancer in the calcification abnormality. Transfer learning of pretrained model on natural image and is included in our work to show effectively to tackle the issue of small dataset. Which is typical for medical data.

Our approach requried minimal human annotation, only a single mammography is needed to feed into the model. There is no requirement to mark a interest area. The result of our work enable a easy and quick adoption in the real clinical data.

Future work includes experiment on other CNN architectures, and integration of latest attention mechanism. A pipeline including segmentation to mark interest area before feeding the mammography into the classifier may also be feasible work to provide a more concrete result.

References

[1] S. W. Duffy, L. Tabár, A. M.-F. Yen, P. B. Dean, R. A. Smith, H. Jonsson, S. Törnberg, S. L.-S. Chen, S. Y.-H. Chiu, J. C.-Y. Fann, M. M.-S. Ku, W. Y.-Y. Wu, C.-Y. Hsu, Y.-C. Chen, G. Svane, E. Azavedo, H. Grundström, P. Sundén, K. Leifland, E. Frodis, J. Ramos, B. Epstein, A. Åkerlund, A. Sundbom, P. Bordás, H. Wallin, L. Starck, A. Björkgren, S. Carlson, I. Fredriksson, J. Ahlgren, D. Öhman, L. Holmberg, and T. H.-H. Chen, "Mammography screening reduces rates of advanced and fatal breast cancers: Results in 549, 091 women," *Cancer*, vol. 126, no. 13, pp. 2971–2979, May 2020. [Online]. Available: https://doi.org/10.1002/cncr.32859

[2] C. D. Lehman, R. F. Arao, B. L. Sprague, J. M. Lee, D. S. M. Buist, K. Kerlikowske, L. M. Henderson, T. Onega, A. N. A. Tosteson, G. H. Rauscher, and D. L. Miglioretti, "National performance benchmarks for modern screening digital mammography: Update from the breast cancer surveillance consortium," *Radiology*, vol. 283, no. 1, pp. 49–58, Apr. 2017. [Online]. Available: https://doi.org/10.1148/radiol.2016161174

[3] M. Heath, K. Bowyer, D. Kopans, P. Kegelmeyer, R. Moore, K. Chang, and S. Munishkumaran, "Current status of the digital database for screening mammography," in *Computational Imaging and Vision*. Springer Netherlands, 1998, pp. 457–460. [Online]. Available: https://doi.org/10.1007/978-94-011-5318-8_75

[4] R. Sawyer-Lee, F. Gimenez, A. Hoogi, and D. Rubin, "Curated breast imaging subset of digital database for screening mammography (cbis-ddsm)," 2016. [Online]. Available: https://wiki.cancerimagingarchive.net/x/lZNXAQ

[5] 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, no. 1, Dec. 2017. [Online]. Available: https://doi.org/10.1038/sdata.2017.177

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

[7] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1207–1216, May 2016. [Online]. Available: https://doi.org/10.1109/tmi.2016.2535865

[8] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *2014 13th International Conference on Control Automation Robotics Vision (ICARCV)*, 2014, pp. 844–848.

[9] H. Yu, L. T. Yang, Q. Zhang, D. Armstrong, and M. J. Deen, "Convolutional neural networks for medical image analysis: State-of-the-art, comparisons, improvement and perspectives," *Neurocomputing*, vol. 444, pp. 92–110, Jul. 2021. [Online]. Available: https://doi.org/10.1016/j.neucom.2020.04.157

[10] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, and N. Navab, “AggNet: Deep learning from crowds for mitosis detection in breast cancer histology images,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1313–1321, May 2016. [Online]. Available: https://doi.org/10.1109/tmi.2016.2528120

[11] Y.-D. Zhang, C. Pan, X. Chen, and F. Wang, “Abnormal breast identification by nine-layer convolutional neural network with parametric rectified linear unit and rank-based stochastic pooling,” *Journal of Computational Science*, vol. 27, pp. 57–68, Jul. 2018. [Online]. Available: https://doi.org/10.1016/j.jocs.2018.05.005

[12] D. Lévy and A. Jain, “Breast mass classification from mammograms using deep convolutional neural networks,” *CoRR*, vol. abs/1612.00542, 2016. [Online]. Available: http://arxiv.org/abs/1612.00542

[13] D. Yi, R. L. Sawyer, D. C. I. au2, J. Dunnmon, C. Lam, X. Xiao, and D. Rubin, “Optimizing and visualizing deep learning for benign/malignant classification in breast tumors,” 2017.

[14] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, “Automatic differentiation in pytorch,” 2017.

[15] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2017.

[16] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[17] Y. Xie and D. Richmond, “Pre-training on grayscale imagenet improves medical image classification,” in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, September 2018.

[18] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, “Transfusion: Understanding transfer learning for medical imaging,” 2019.

[19] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2015.

[20] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” 2015.

[21] I. Loshchilov and F. Hutter, “Sgdr: Stochastic gradient descent with warm restarts,” 2017.

[22] P. Goyal, P. Dollár, R. Girshick, P. Noordhuis, L. Wesolowski, A. Kyrola, A. Tulloch, Y. Jia, and K. He, “Accurate, large minibatch sgd: Training imagenet in 1 hour,” 2018.

[23] Tony-Y, “pytorch_warmup,” https://github.com/Tony-Y/pytorch_warmup, 2022.