Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 135 (2018) 400–407

Procedia Computer Science

www.elsevier.com/locate/procedia

3rd International Conference on Computer Science and Computational Intelligence 2018

Transfer Learning from Chest X-Ray Pre-trained Convolutional Neural Network for Learning Mammogram Data

Bens Pardamean[a,c], Tjeng Wawan Cenggoro[b,c,*], Reza Rahutomo[c], Arif Budiarto[b,c], Ettikan Kandasamy Karuppiah[d]

[a] *Computer Science Department, BINUS Graduate Program - Master of Computer Science Program, Bina Nusantara University, Jakarta, Indonesia 11480*

[b] *Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480*

[c] *Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta, Indonesia 11480*

[d] *NVIDIA Corporation*

Abstract

Breast cancer is one of the deadliest cancer for female nowadays. Despite of the rapid advancement in medical image analysis with the rise of deep learning, development of breast cancer detection system is limited due to relatively small size of the publicly available mammogram dataset. In this paper, we discover an effective configuration for transfer learning from Chest X-Ray pre-trained Convolutional Neural Network to overcome the small-size mammogram dataset problem. We found that the best configuration achieve 90.38% validation accuracy for modified.

Keywords: Transfer learning; Convolutional neural network; Medical image analysis; Chest X-ray; Mammogram;

1. Introduction

Breast Cancer is one of the most dangerous cancer in the world. Back in 2012, at least 1.7 million cases has been diagnosed and over 500,000 deaths were caused by cancer, with 25% of the cases are breast cancer[1]. Not to mention that, while breast cancer is commonly occurs in women, the probability for it to occurs in men is not zero. Fortunately, it has been proved that survival rate for breast cancer cases can be increased by the adoption of cancer screening as a detection method[2].

One of the tools for early detection that can be used is Chest X-ray It has been used in many surgical intensive care units including breast cancer early detection[3]. In particular for screening breast cancer, a low dose X-ray can be used to

* Corresponding author. Tel.: +62-21-534-5830 ext. 2456; fax: +62-21-530-0244.
E-mail address: wcenggoro@binus.edu

10.1016/j.procs.2018.08.190

deliver scanned image of the breast called as mammogram[4]. This screening method, called as mammography, played a major role in reducing breast cancer morbidity over the years. Both American Cancer Society and WHO recommend women to do mammogram screening start from the age 40 at least once in two years[5]. Despite of the importance of early detection methods, there are still difficulties in interpreting the data given by early detection methods[6].

Meanwhile in image analysis, Convolutional Neural Network (CNN) has become state-of-the-art technique tasks[7,8,9,10,11]. In particular for X-ray image analysis, a CNN named CheXNet has achieved human-level performance for chest X-ray image analysis[12]. This performance is possible by the use of ChestX-ray[13] dataset, which contains 112,120 Chest X-ray images, as CheXNet training data.

In this paper, we report the result of our study on how to effectively transfer learning from CheXNet for CNN to learn from DDSM dataset[14,15]. We see that both ChestX-ray and DDSM are X-Ray images, but taken from different body parts. Thus, we want to find out if transfer learning from CheXNet could help CNN to learn from DDSM. This motivation comes from the fact that CNN typically requires massive dataset to show its peak performance. On the other hand, DDSM contains only 2,605 set of unique cases, with 4 X-ray images per case, which can be considered small compared to ChestX-Ray. In order to develop an optimal CNN for this purpose, we conduct a preliminary research to find the best transfer learning configuration from CheXNet to learn from DDSM. The results from this preliminary research are reported in this paper.

2. Related Works

Recently, deep learning has become a preferred technique for mammogram data analysis[16]. It has been applied in a broad aspect of mammogram data processing such as detection[17,18], segmentation[19], and classification[20]. Furthermore, Wang et al.[20] has proved that the accuracy obtained by deep learning is slightly higher than SVM. Not to mention that Becker et al.[17] even claimed that deep learning has achieved a human-level accuracy for breast cancer detection using mammogram data. However, the size of current public dataset for mammogram is relatively small compared to common public dataset for deep learning[21,22,13]. Therefore, CNN model for mammogram can potentially perform better given more training data.

In the attempt to improve deep learning model over small-size dataset, transfer learning is currently the most popular strategy to be employed. Transfer learning enables deep learning model to effectively learn from small dataset by transferring learned features from other deep learning model that previously learned from similar dataset with large size. The first effort in transfer learning is performed by Girshick et al[23] that uses supervised pretraining CNN on large-scale image classification dataset[21] for learning from relatively small dataset for object detection[24]. Yosinski et al[25] also supported the research from Girshick et al. by extensively studying the effect of transfer learning to improve deep learning model performance.

Before 2014, the idea similar to transfer learning for deep learning is actually has been studied, which is called as unsupervised pre-training. This method use a model which is pre-trained unsupervisedly before the weights is transferred to the main model. This idea was popularized by Hinton et al.[26] by introducing greedy layer-wise pre-training to train Deep Belief Network. This unsupervised pre-training strategy is generally considered inferior to its supervised version, which popularly termed as transfer learning nowadays. However, there are numerous application of this strategy in medical imaging[27,28,29] and also in other areas of research[30,31,32,33].

In a similar approach to this research, Kim et al. proves that transfer learning gives a boosted performance compared to directly train Inception v4[34] with the training dataset[35]. They use Inception v4 which is pretrained with ImageNet dataset to be transferred for learning DDSM dataset. The research shows that transfer learning enables Inception v4 to reach 72% of classification accuracy.

3. Research Methodology

The work in this research is divided into three main phases: Pre-Training, Transfer Learning, and Comparison, as seen in figure 1. Firstly, in pre-training phase, we acquire a pre-trained model to be transferred for learning mammogram data. Afterwards, we transfer the weights of the pre-trained model to a new model then train the new model with mammogram dataset. Lastly, we compare several configuration of new models to determine the best configuration for the case of transfer learning in this research.

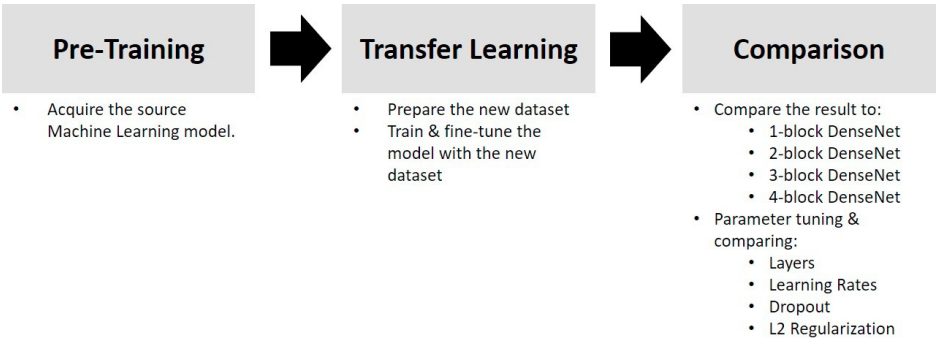

Fig. 1: Research Workflow

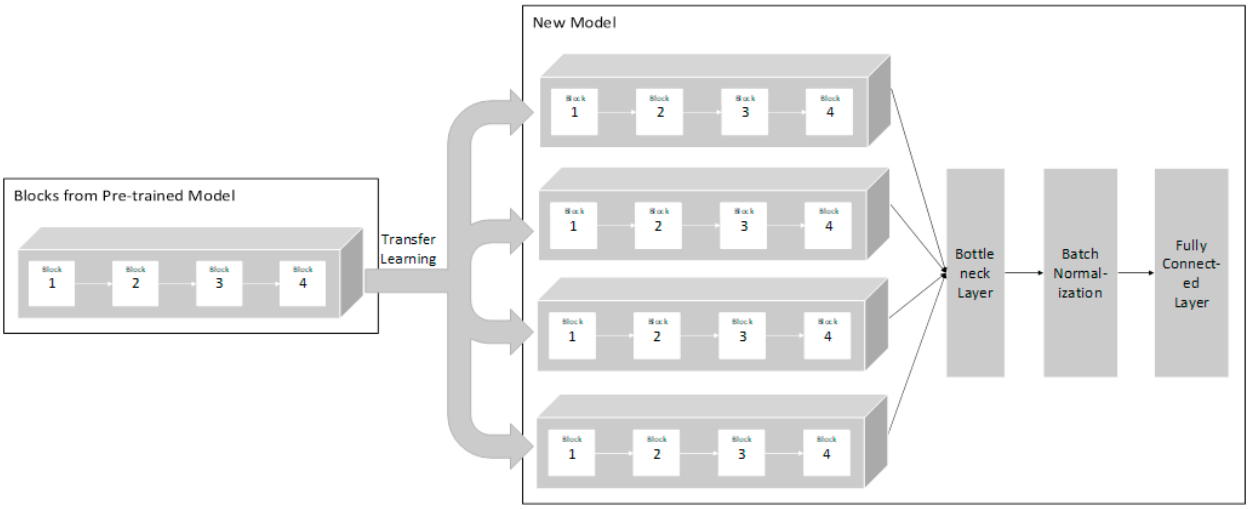

Fig. 2: Illustration of Transfer Learning Scheme

In the pre-training phase, we acquired CheXNet[12] as our pre-trained deep learning model. CheXNet is a 121-layers DenseNet model[11], which layers are grouped into 4 dense blocks. This model has been pre-trained on Chest X-Ray[13] dataset and was acquired from CheXNet reimplementation project of Machine Intelligence Lab, Institute of Computer Science & Technology, Peking University[36].

The weights of the pre-trained model is then transferred to a new model that learns from mammogram data, which we call as transfer learning phase. As mammogram data consists of four images per case, we built the new model with four parallel input paths. Each of the input path contains same dense blocks architecture from the pre-trained model. Afterwards, the same weights of the pre-trained model blocks are transferred to all four parallel input paths. This transfer learning scheme is illustrated in fig. 2. When the model is trained, each of the input path read one of the four images in each case.

In the comparison phase, we conduct experiments to compare several transfer learning configurations. The configuration we use is the variation in number of employed dense blocks, number of employed layers in the last dense block, learning rate, dropout rate, and L2 regularization rate. Each configuration run for 20 epochs and the best validation accuracy is recorded for comparison. All configurations are optimized using Adam optimization method[37], with categorical cross-entropy as loss function. All configurations are run on a server with an NVIDIA Tesla P100 GPU.

For experiment in number of blocks, we try to remove the blocks one by one starting from the last blocks. This strategy is employed because we want to find out what are the best low-level features to be transferred from CheXNet to learn DDSM. It has been known that CNN generally learn more complex features as the layers are stacked[38], thus we can the best features complexity level for our case by using this strategy. We also apply same strategy for

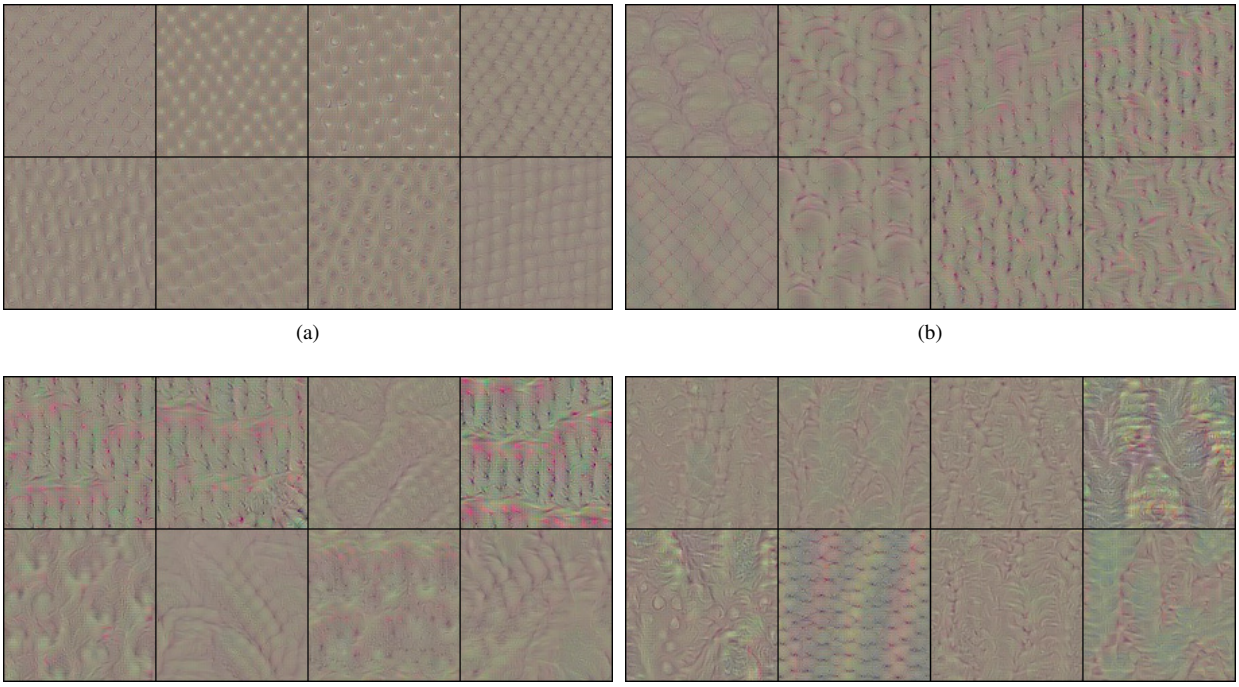

Fig. 3: Samples of Features Visualization for: (a) block 1; (b) block 2; (c) block 3; (d) block 4.

experiment in number of layers in the last block. The reason why we choose to reduce the number of blocks first before reducing the number of layers is mainly for computational efficiency. There will be 120 different configuration to try if we directly try to reduce the number of layers. On contrary, we only need 4 configurations plus the number of layers in the last block in the best blocks configuration by using our strategy.

As there exist visualization techniques to see what features that are learned by CNN, it seems interesting to see the learned features from CheXNet. By seeing the features, we can directly choose the features based on the complexity of the features shown by the visualization. Unfortunately, this strategy is not suitable in our case, as the learned features has no semantic meaning to be inferred by human. This is contrasted to other cases such as face recognition, which model tends to pick semantically meaningful features like eyes, ears, and nose. Figure 3 shows the result of visualization on CheXNet block 1 to 4 using optimization-based technique from Yosinski et al.[39]. We use optimization-based technique because the other type of techniques are not suitable in our case. We can see that CheXNet is indeed learns more complex features as the blocks stacked, but the learned features has no semantic meaning to be inferred by human.

For the dataset in all experiments, we use a modified DDSM dataset. DDSM originally consists of 4 classes: normal, benign-without-callbacks, benign, and malignant. The modification we make is aggregation of benign-without-callbacks, benign, and malignant class into a single class. We do this aggregation because we find that the simple approach we use in this research cannot perform well on fine-grained classes like benign-without-callbacks, benign, and malignant, which conceptually can be grouped as a single class named cancer. Afterwards, we randomly split the dataset into training, validation, and test set with the ratio of 60:20:20. During training, we augment the data using 5-crop technique which was popularized by Krizhevsky et al.[7]. We only use 5-crop augmentation because this is the only technique that is suitable given the characteristic of mammogram data and the parallel input paths we use.

4. Results and Discussion

To determine the best number of employed dense blocks, we compare the performance of models with 4, 3, 2, and 1 dense blocks in each input paths. These 3, 2, and 1 dense blocks configurations are constructed by removing the blocks from the original CheXNet starting from the last block. The weights of the *n* used dense blocks are transferred from the corresponding dense blocks of the original CheXNet model. To keep the necessary learned features from CheXNet, all dense blocks except the last dense block is frozen during training. We set the batch size to 24 when on each configuration. This batch size is chosen as it is the largest batch size that could fit into the GPU for configuration with 4 dense blocks. The experiment use learning rate of 1e-3, as recommended in Kingma et al.[37]. Based on the table 1, we can see that the best performance is achieved by the configuration with only 2 dense blocks, in terms of both loss value and accuracy. This result indicates that it is not necessary to use all sets of dense blocks from the original CheXNet model for transfer learning to learn mammogram data. In order to show the feasibility of our transfer learning approach, we also show the computation time needed to train each configuration in table 1.

Table 1: The result of the experiment with blocks.

Blocks	Training		Validation		Running Time (20 Epochs)
	Loss	Accuracy	Loss	Accuracy	
4	0.3493	96.31 %	0.4380	87.31 %	2h 38m 58s
3	0.3407	97.20 %	0.4408	87.12 %	2h 31m 56s
2	**0.3960**	**91.68** %	**0.4324**	**88.27** %	**2h 19m 32s**
1	0.4625	85.76 %	0.4588	85.77 %	2h 6m 49s

Afterwards, we conduct an experiment to try several configurations of the number of employed layers in the second dense block. The number of layers we try is ranged from 1 to 12 layers, as the second block of original CheXNet model consists of 12 layers. We only train the last layer and froze the rest to keep necessary learned features. The batch size used in this experiment is 32, as it is the largest batch size to fit in the GPU if we use 12 layers. From table 2, the best result performed by using 6 layers with 90.38% of validation accuracy and 0.4107 of loss value.

Table 2: The result of the experiment with layers.

Layers	Training		Validation		Running Time (20 Epochs)
	Loss	Accuracy	Loss	Accuracy	
12	0.4013	91.39 %	0.4290	88.46 %	2h 11m 38s
11	0.3262	98.80 %	0.4232	88.65 %	2h 10m 41s
10	0.3293	98.47 %	0.4240	88.65 %	2h 6m 6s
9	0.3456	97.01 %	0.4250	88.46 %	2h 9m 2s
8	0.3436	97.12 %	0.4258	88.46 %	2h 6m 56s
7	0.3220	99.22 %	0.4279	88.65 %	2h 6m 7s
6	**0.3345**	**98.02 %**	**0.4107**	**90.38 %**	**2h 3m 57s**
5	0.3314	98.34 %	0.4258	88.85 %	2h 2m 19s
4	0.3366	97.76 %	0.4149	89.61 %	2h 2m 1s
3	0.3442	97.16 %	0.4155	89.62 %	1h 59m 48s
2	0.3392	97.60 %	0.4135	90.00 %	2h 0m 1s
1	0.3465	96.82 %	0.4263	88.85 %	1h 59m 21s

We also try a few configuration on learning rate value. In this experiment, we only try three learning rate value: 1e-2, 1e-3, and 1e-4. As we see in table 3, the best result achieved by learning rate value 1e-3 with 90.38% of validation accuracy and 0.4107 of loss value, which is the default learning rate we use in the previous experiments.

As we can see from table 2 and 3, the best performing configuration suffers a noticeable overfitting, which is implied by the gap of training and validation accuracy of 7.37%. Therefore, we try several configurations of regularization. We employ two types of regularization in the experiment: dropout[40] and L2 regularization. We use three value of dropout

Table 3: The result of the experiment with learning rates.

Learning Rate	Training		Validation		Running Time (20 Epochs)
	Loss	Accuracy	Loss	Accuracy	
1e-2	0.4349	87.76 %	0.4523	86.35 %	2h 3m 33s
1e-3	**0.3345**	**98.02 %**	**0.4107**	**90.38 %**	**2h 3m 57s**
1e-4	0.3325	98.52 %	0.4284	88.46 %	2h 0m 35s

rate: 0, 0.2, and 0.5 as displayed in table 4. Dropout rate of 0.2 was employed in original DenseNet paper[11], while 0.5 was the default dropout value in the original dropout paper[40]. For L2 regularization, we also use three rate value: 0, 1e-4, and 1e-2 as seen in table 5. These rate values are commonly used for L2 regularization. Even though we are able to decrease the gap to 6.72% using dropout and 1.45% using L2, we cannot achieve better validation accuracy compared to model with no regularization. To evaluate the performance of our best model over test sub-dataset, we use Receiver Operating Characteristic (ROC) as seen in figure 4, with AUC (Area Under ROC Curve) of 0.916. The accuracy over test sub-dataset is 89.42%.

Table 4: The result of the experiment with dropouts.

Dropout Rate	Training		Validation		Running Time (20 Epochs)
	Loss	Accuracy	Loss	Accuracy	
0	**0.3345**	**98.02 %**	**0.4107**	**90.38 %**	**2h 3m 57s**
0.2	0.3392	97.51 %	0.4265	89.23 %	2h 0m 49s
0.5	0.3612	95.37 %	0.4200	88.65 %	2h 3m 37s

Table 5: The result of the experiment with L2 regularization.

L2 Regularization Rate	Training		Validation		Running Time (20 Epochs)
	Loss	Accuracy	Loss	Accuracy	
0	**0.3345**	**98.02 %**	**0.4107**	**90.38 %**	**2h 3m 57s**
1e-04	0.3849	92.98 %	0.4308	88.46 %	2h 4m 6s
1e-02	0.4335	88.96 %	0.4508	87.50 %	2h 3m 36s

5. Conclusion

In this paper, we try to find the best configuration for transfer learning from CheXNet to learn mammogram data. We find that the best configuration only employ the first two dense blocks from the original CheXNet model. The optimal number of employed layer in the last used block is also fewer than the original model, which is 6 out of 12 layers. We also find that the best learning rate is 1e-3, which is the default learning rate of Adam optimization method. The experiment on several regularization configuration cannot produce a better performance than the configuration without regularization. However, regularization successfully reduce the overfitting suffered by the best performing model. Through this research, we find that the best performing configuration achieve 90.38% of validation accuracy.

For future works, it is interesting to conduct a more extensive research in finding optimal non-architectural hyper-parameter value. A better hyperparameter searching method such as grid search and random search might be able to find more optimal configuration as opposed to trial-and-error approach we use in this research. It is also interesting to develop a more complex model for transfer learning that is able to learn fine-grained classes such as benign-without-callback, benign, and malignant.

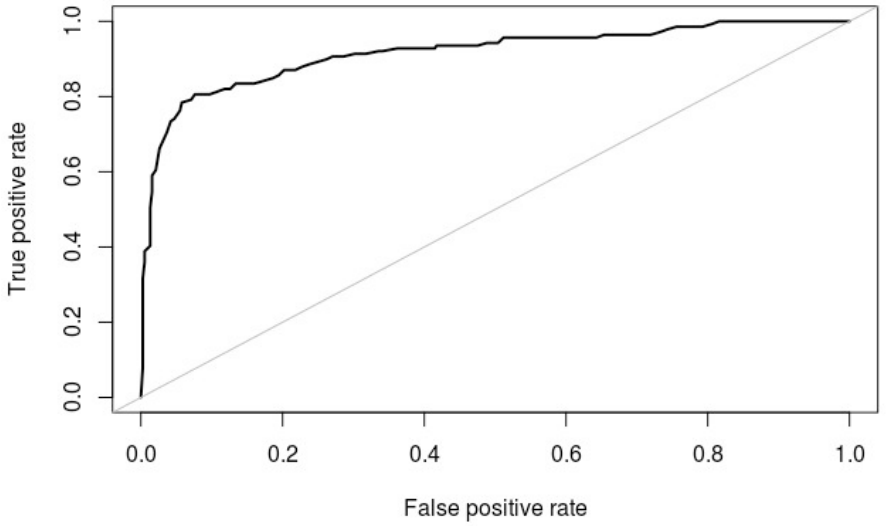

Fig. 4: ROC Curve

Acknowledgements

The experiments in this research are conducted using NVIDIA Tesla P100 GPU granted by NVIDIA Corporation to NVIDIA - Bina Nusantara University AI R&D Center.

References

1. Nounou, M.I., ElAmrawy, F., Ahmed, N., Abdelraouf, K., Goda, S., Syed-Sha-Qhattal, H.. Breast cancer: conventional diagnosis and treatment modalities and recent patents and technologies. *Breast cancer: basic and clinical research* 2015;**9**:BCBCR–S29420.
2. Monticciolo, D.L., Newell, M.S., Hendrick, R.E., Helvie, M.A., Moy, L., Monsees, B., et al. Breast cancer screening for average-risk women: recommendations from the acr commission on breast imaging. *Journal of the American College of Radiology* 2017;**14**(9):1137–1143.
3. Resnick, S., Inaba, K., Karamanos, E., Skiada, D., Dollahite, J.A., Okoye, O., et al. Clinical relevance of the routine daily chest x-ray in the surgical intensive care unit. *The American Journal of Surgery* 2017;**214**(1):19–23.
4. Dheeba, J., Singh, N.A., Selvi, S.T.. Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach. *Journal of biomedical informatics* 2014;**49**:45–52.
5. Ghanbarzadeh Dagheyan, A., Molaei, A., Obermeier, R., Westwood, A., Martinez, A., Martinez Lorenzo, J.A.. Preliminary results of a new auxiliary mechatronic near-field radar system to 3d mammography for early detection of breast cancer. *Sensors* 2018;**18**(2):342.
6. Siegel, R.L., Miller, K.D., Jemal, A.. Cancer statistics, 2018. *CA: a cancer journal for clinicians* 2018;**68**(1):7–30.
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. 2012, p. 1097–1105.
8. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. Going deeper with convolutions. Cvpr; 2015, .
9. Ren, S., He, K., Girshick, R., Sun, J.. Faster r-cnn: Towards real-time object detection with region proposal networks. In: *Advances in neural information processing systems*. 2015, p. 91–99.
10. He, K., Gkioxari, G., Dollár, P., Girshick, R.. Mask r-cnn. In: *Computer Vision (ICCV), 2017 IEEE International Conference on*. IEEE; 2017, p. 2980–2988.
11. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.. Densely Connected Convolutional Networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. ISBN 978-1-5386-0457-1; 2017, p. 2261–2269. doi:10.1109/CVPR.2017.243. URL http://ieeexplore.ieee.org/document/8099726/.
12. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:171105225* 2017;.
13. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases 2017;doi:10.1109/CVPR.2017.369. URL http://arxiv.org/abs/1705.02315.
14. Heath, M., Bowyer, K., Kopans, D., Kegelmeyer, P., Moore, R., Chang, K., et al. Current status of the digital database for screening mammography. In: *Digital mammography*. Springer; 1998, p. 457–460.

15. Heath, M., Bowyer, K., Kopans, D., Moore, R., Kegelmeyer, P.. The digital database for screening mammography. *Digital mammography* 2000;:431–434.
16. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., et al. A survey on deep learning in medical image analysis. *Medical image analysis* 2017;**42**:60–88.
17. 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. *Investigative radiology* 2017;**52**(7):434–440.
18. Dhungel, N., Carneiro, G., Bradley, A.P.. Automated mass detection in mammograms using cascaded deep learning and random forests. In: *Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference on*. IEEE; 2015, p. 1–8.
19. Kallenberg, M., Petersen, K., Nielsen, M., Ng, A.Y., Diao, P., Igel, C., et al. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. *IEEE transactions on medical imaging* 2016;**35**(5):1322–1331.
20. Wang, J., Yang, X., Cai, H., Tan, W., Jin, C., Li, L.. Discrimination of breast cancer with microcalcifications on mammography by deep learning. *Scientific reports* 2016;**6**:27327.
21. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* 2015;**115**(3):211–252. doi:10.1007/s11263-015-0816-y. URL http://dx.doi.org/10.1007/s11263-015-0816-y.
22. Lin, T.y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. Microsoft COCO: Common Objects in Context. In: *European Conference on Computer Vision 2014*. 2014, p. 740–755. doi:10.1007/978-3-319-10602-1_48. URL http://link.springer.com/10.1007/978-3-319-10602-1{_}48.
23. Girshick, R., Donahue, J., Darrell, T., Malik, J.. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, p. 580–587.
24. Everingham, M., Van˜Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.. The Pascal Visual Object Classes Challenge. *Ijcv* 2010;**88**(2):303–338. doi:10.1007/s11263-009-0275-4.
25. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.. How transferable are features in deep neural networks? In: *Advances in neural information processing systems*. 2014, p. 3320–3328.
26. Hinton, G.E., Osindero, S., Teh, Y.W.. A fast learning algorithm for deep belief nets. *Neural computation* 2006;**18**(7):1527–54. doi:10.1162/neco.2006.18.7.1527. URL http://www.ncbi.nlm.nih.gov/pubmed/16764513.
27. Cheng, J.Z., Ni, D., Chou, Y.H., Qin, J., Tiu, C.M., Chang, Y.C., et al. Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans. *Scientific Reports* 2016;**6**(1):24454. doi:10.1038/srep24454. URL http://www.nature.com/articles/srep24454.
28. Kallenberg, M., Petersen, K., Nielsen, M., Ng, A.Y., Diao, P., Igel, C., et al. Unsupervised Deep Learning Applied to Breast Density Segmentation and Mammographic Risk Scoring. *IEEE Transactions on Medical Imaging* 2016;**35**(5):1322–1331. doi:10.1109/TMI.2016.2532122. URL http://ieeexplore.ieee.org/document/7412749/.
29. Suk, H.I., Wee, C.Y., Lee, S.W., Shen, D.. State-space model with deep learning for functional dynamics estimation in resting-state fMRI HHS Public Access. *Neuroimage* 2016;**129**:292–307. doi:10.1016/j.neuroimage.2016.01.005. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5437848/pdf/nihms864760.pdf.
30. Lin, Z., Chen, Y., Zhao, X.. Spectral-Spatial Classification of Hyperspectral Image Using Autoencoders 2013;(61301206).
31. Oliveira, T.P., Barbar, J.S., Soares, A.S.. Multilayer Perceptron and Stacked Autoencoder for Internet Traffic Prediction 2014;:61–71.
32. Chen, L., Cai, C., Chen, V., Lu, X.. Learning a hierarchical representation of the yeast transcriptomic machinery using an autoencoder model. *BMC Bioinformatics* 2016;**17**(S1):S9. doi:10.1186/s12859-015-0852-1. URL http://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-015-0852-1.
33. Cenggoro, T.W., Isa, S.M., Kusuma, G.P., Pardamean, B.. Classification of imbalanced land-use/land-cover data using variational semi-supervised learning. In: *2017 International Conference on Innovative and Creative Information Technology (ICITech)*. IEEE. ISBN 978-1-5386-4046-3; 2017, p. 1–6. doi:10.1109/INNOCIT.2017.8319149. URL http://ieeexplore.ieee.org/document/8319149/.
34. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.. Inception-v4, inception-resnet and the impact of residual connections on learning. In: *AAAI*; vol. 4. 2017, p. 12.
35. Kim, M., Zuallaert, J., De Neve, W.. Towards novel methods for effective transfer learning and unsupervised deep learning for medical image analysis. In: *Doctoral Consortium (DCBIOSTEC 2017)*. 2017, p. 32–39.
36. Weng, X., Zhuang, N., Tian, J., Liu, Y.. Chexnet for classification and localization of thoracic diseases. https://github.com/arnoweng/CheXNet; 2017.
37. Kingma, D.P., Ba, J.. Adam: A Method for Stochastic Optimization. *The International Conference on Learning Representations 2015* 2014;:1–15URL http://arxiv.org/abs/1412.6980.
38. Zeiler, M.D., Fergus, R.. Visualizing and Understanding Convolutional Networks. In: *European Conference on Computer Vision*. ISBN 978-3-319-10589-5; 2014, p. 818–833. doi:10.1007/978-3-319-10590-1_53. URL http://arxiv.org/abs/1311.2901http://link.springer.com/10.1007/978-3-319-10590-1{_}53.
39. Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., Lipson, H.. Understanding Neural Networks Through Deep Visualization. *International Conference on Machine Learning - Deep Learning Workshop 2015* 2015;:12URL http://arxiv.org/abs/1506.06579.
40. Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.. Dropout : A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research (JMLR)* 2014;**15**:1929–1958. doi:10.1214/12-AOS1000.