Association for Information Systems
AIS Electronic Library (AISeL)

MWAIS 2020 Proceedings
Midwest (MWAIS)

5-28-2020

Transfer Learning in Medical Image Classification: Challenges and Opportunities

Rajesh Godasu
rajesh.godasu@trojans.dsu.edu

David Zeng

Kruttika Sutrave

Transfer Learning in Medical Image Classification: Challenges and Opportunities

Rajesh Godasu
Dakota State University
Rajesh.Godasu@trojans.dsu.edu

David Zeng
Dakota State University
David.Zeng@dsu.edu

Kruttika Sutrave
Dakota State University
Kruttika.Sutrave@trojans.dsu.edu

ABSTRACT

Transfer Learning is a popular technique in Medical Image classification. Transfer Learning methods are extensively applied with CNNs such as Resnet, Densenet, VGG16, and Inception, etc. for various medical diagnosis tasks. CNNs are around since the 1980s, but 60-80 percent of the TL research in MIC is done in the last three years. While CNNs can be used in traditional computer vision domain, they have been ensembled, segmented and improvised recently to resolve multiple MIC problems. This review identifies three main challenges in implementing Transfer Learning for Medical Image Classification: (1) Overparameterization of deep CNN models, (2) Expensive Computations, and (3) Insufficient availability of labeled data in the Medical field. The study also identifies research opportunities in the form of Light-weight architectures and Multi-stage Transfer Learning which could potentially mitigate the above-mentioned challenges.

Keywords

Convolutional Neural Network, Transfer Learning, Medical Image Classification, Lightweight Architectures, Multi-stage Transfer Learning

INTRODUCTION

Convolutional Neural Networks(CNNs) are powerful image recognition networks built with Convolutional, pooling, and fully-connected layers. They use deep learning(DL) methods and architectures for Medical Image Classification (MIC) tasks(Altaf et al. 2019; Gao et al. 2019). The early works of CNNs date back to 1980s with Neocognitron(Fukushima 1980), a self-organizing architecture that recognizes patterns based on similarities in geometrical shapes and unaffected by position changes. The true success of CNNs in object recognition was witnessed in 2012 after AlexNet(Krizhevsky et al. 2012) won ILSVRC, an image classification competition. Since AlexNet's success, tremendous interest has been generated in image classification research, especially in MIC(Yamashita et al. 2018). Extensive research on CNN in MIC is done in recent years among various image segmentation and classification tasks(Litjens et al. 2017; Shin et al. 2016). While CNNs can be used successfully with their original architectures, they have been ensembled (Majtner et al. 2018), segmented (Wong et al. 2018) and improvised (Xu et al. 2018) more recently to resolve multiple MIC problems.

While Deep CNNs usually require large amounts of labeled data for training purpose to avoid overfitting, Transfer Learning (TL), an effective method for data-scarce situations, is extensively applied with ImageNet (a large dataset of labeled natural images) pre-trained classic architectures like Resnet (He et al. 2016), Densenet (Huang et al. 2017), VGG16 (Simonyan and Zisserman 2014), Inception (Szegedy, Wei Liu, et al. 2015), etc. for various medical diagnosis tasks(Menegola et al. 2017). The rapid growth of CNN applications in MIC motivates us to review the current strengths and weaknesses of TL in MIC. Analyzing the complex networks and presenting future research directions would help create better MIC solutions. This study aims to 1. Identify the current challenges associated with TL in MIC, 2. Present potential opportunities to mitigate the challenges, and 3. Provide future research directions. In this review, the articles focusing on TL in MIC are collected from digital libraries such as IEEE explore, Arxiv, Pubmed and Google scholar. The search methods include the usage of keywords such as "Transfer Learning", "Convolutional Neural Networks", "Medical Image Classification". Furthermore, we follow the recommended forward and backward referencing approach, proposed for high-quality literature reviews in the IS field (Webster and Watson 2002). In the following sections, we provide a review of TL, identify challenges and opportunities with a focus on Multistage TL and Lighweight Architecture. In the discussion section, we propose a Multi-stage TL framework for MIC tasks.

TRANSFER LEARNING

TL is the application of features learned in one task for better generalization of a different task (Pan and Yang 2010; Yu et al. 2017). For instance, in task T1, a model learning the visual features of buildings will be able to learn and generalize the visual features of cars in task T2 given the considerable amount of buildings data is provided in task T1. There are two types of transfer learning in CNN, feature extraction and fine-tuning. In the above-mentioned example of TL, feature extraction is performed when a CNN learns features from Task 1, then uses the same base network of convolutional and pooling layers by replacing the fully connected layer with the Task 2-specific classifier. During this operation, the weights of the convolutional base are frozen (thus preserved) and only the new classifier is trained for classification of cars. The features learned from the convolutional base are reusable and generic, whereas in the fully-connected layers, representations learned are specific to the classes on which the network is trained. Hence it is replaced for new data(François 2017).

In fine-tuning, the weights of the convolutional base are slightly altered to match the task 2 problem after training the network on buildings data. Unlike feature extraction, fine-tuning requires unfreezing some top-level layers of convolutional base and a new classifier is added (together with its parameters learned while the entire conv base are frozen). While performing this type of TL, the unfrozen layers of the CNN are trained along with the fully-connected layer when passed through new data. Multi-stage transfer learning is applying either feature extraction or fine-tuning more than once, enabling customized training pipelines and higher flexibility of reusing representations learned from previous stages in the subsequent training tasks.

CHALLENGES

Overparameterization

Many of the MIC studies based on TL use variants of classic CNN architectures including but not limited to Alexnet, Densenet, ResNet, Inception, VGG16, etc. CNNs are overparameterized in MIC to gain performance improvements. A recent study on understanding TL for MIC(Raghu et al. 2019) reported that standard deep CNNs learn slower when compared to lightweight models during the training process because of overparameterization. This challenge could lead to an increase in training times and longer epochs because of inefficient learning. For instance, an Alexnet is combined with the Support Vector Machine algorithm(Dawud et al. 2019) for Brain Hemorrhage classification tasks which showed better accuracy than the baseline Alexnet model, but it took a larger number of epochs and longer training time to achieve higher accuracy. The research(Raghu et al. 2019) also pointed out that TL provides significant benefits instead of training the model from scratch, however, the architecture sizes and parameters have negative influences in the case of fine-tuning ImageNet pre-trained models.

Expensive Computations

A recent study (Wu et al. 2019) uses extra attention modules for gaining discriminative features from the deep layers and to decline features that are not required. Using these modules proved to improve the accuracy but increased the computational load. Intuitively, shorter training times and lower computational costs would be benefitial to the MIC field if model effectiveness is not decreased. Standard CNN models are successful in MIC because of their high amount of depth and large number of layers but incur additional computational costs. The Inception network has 7 million parameters that are fewer, comparing to other classic CNNs, but still too expensive to be trained on a regular i5 computer. While there is a big advancement in Central Processing Units and Graphics Processing Units recently, it is not cost-effective in many real-world applications. This situation provides the opportunity to explore Lightweight Architectures (LWAs) that can be applied in mobile devices for better usability and efficiency as it is still in its earlier stages (Khan et al. 2019).

Insufficient Labeled Data

A dataset is considered labeled if it contains appropriate annotation of the images for all the available classes. For example, if a chest X-ray image dataset has two classes "pneumonia" and "normal", each image must be annotated with either of these two classes. Labeled datasets are important for MIC, because deep learning architectures depend on large amounts of labeled images for training purpose to provide successful image classification(Menegola et al. 2017). Obtaining large amounts of labeled medical image datasets in any domain may be a difficult task (Altaf et al. 2019), comparing with natural images. TL can further prosper and create more robust solutions with the availability of sufficient labeled data. Traditional TL methods are performing well in the data-scarce conditions, but there is still room for further improvement. In natural images, the differences in features are significant from one image to another, with the images possessing a variety of lightning and shapes, but in medical images, these differences can be minute depending on the problem(Erickson et al. 2018), making it difficult to learn representations with limited data while using TL method. One effective method used to overcome this challenge is Multistage Transfer Learning.

OPPORTUNITIES

Multi-stage TL

While Traditional TL uses fine-tuning and feature extraction, multi-stage TL applies it more than once. This allows us to learn features from a similar domain(medical) and perform better on target because of feature similarities(Kim et al. 2017; Samala et al. 2019). Modality bridge TL by (Kim et al. 2017) is one of the first studies to implement two-stage TL in MIC, they employed pictures from MRI, CT, and X-Ray as bridge datasets and used the target databases from the same domains with less annotated images, while the source is still ImageNet(natural images). First, projection function from the source database into the feature space of bridge dataset is learned, based on this function, non-linear mapping of feature space from source to bridge is done and finally, the classifier learns based on projection function leveraged by bridge database. This process supports the domain adaptation of the problem dataset and the source dataset. VGG16 architecture is used to perform the experiments and resulted in improving the accuracy in all three modalities (MRI, CT, and X-Ray).

Another study(Samala et al. 2019), compared the results of single-stage TL with multi-stage to classify malignant and benign masses in breast cancer. In the first approach, they simply fine-tuned the pre-trained CNN with digital breast tomosynthesis(DBT) data. In a two-stage approach, the pre-trained CNN learned representations from mammography data and then in the second stage, it was fine-tuned with the target DBT dataset. The two-stage approach significantly outperformed the traditional approach. This research indicates that, in limited data conditions, TL, when applied more than once can leverage the knowledge gained through source tasks from unrelated and related domains.

Lightweight CNNs

Lightweight Architectures have fewer parameters and are smaller in size compared to standard deep CNNs, however, they seem to produce similar accuracies. LWAs enable us to train these models quickly and perform well (Iandola, Ashraf, et al. 2016). Squeezenets(Iandola, Han, et al. 2016), MobileNets(Howard et al. 2017), ShuffleNet(Zhang et al. 2017), MobileNetsV2(Sandler et al. 2019), Plexusnet(Eminaga et al. 2019), NasNets(Zoph et al. 2018), EfficientNets(Tan and Le 2019), etc. are some of LWAs that are laying path towards more efficient image classification tasks. Each of these models possesses different architectural benefits producing better results in image recognition tasks. For example, MobileNets use depthwise separable convolutions while SqueezeNets achieve better accuracy by replacing 3X3 convolutional filters with 1X1 filters to significantly reduce computational costs during training.

Applications of LWAs in MIC are still in the early stages, but the field is rapidly advancing with proposals of new approaches to improve efficiency as mentioned in the earlier section. Many of these models are applied for TL in MIC, for example, a recent study on Mobile Dermoscopy for skin cancer detection(Ech-Cherif et al. 2019) uses MobileNetV2 for binary classification of benign and malignant cases. This model achieved a 91.33% accuracy when trained with a batch size of 32. A prototype of this model was implemented in a mobile app that classifies all the sample images correctly. Another study(Shamim et al. n.d.) compares six CNN models, Alexnet, GoogleNet, InceptionV3(Szegedy, Vanhoucke, et al. 2015), ResNet 50, Squeezenet, and VGG19 for tongue lesion classifications for both binary and multi-class classification tasks. The Squeezenet model achieved the best speed when compared to all other models because of its architecture of fewer parameters. Even though Squeezenet achieved competitive results with more complex models, there were some major misclassifications of precancerous images as benign, so there still exist issues in LWAs and thus further research is required.

DISCUSSION AND CONCLUSION

While recent studies conducted in various medical domains have provided state of the art results, TL in MIC has a remarkable potential to further success. Deep CNNs with advanced features facilitated these results. However, challenges like overparameterization, expensive computations and insufficient availability of labeled data may cause major hurdles to the advancements of TL applications. We identified opportunities in the form of multi-stage TL and LWA as the potential remedies to mitigate the above-mentioned challenges to some extent. Multi-stage TL in MIC is largely focused on standard deep CNNs in recent history. We encourage future researchers to accelerate research in multi-stage TL using LWAs to reap the benefits of shorter training time that could potentially solve the challenges like insufficient labeled data and expensive computations substantially. This kind of research is suitable in numerous medical areas including but not limited to

Radiology, Neurology, Dermatology, etc. In the future, we plan to implement multi-stage TL with more than two stages of TL using LWAs in various medical areas such as Brain, Chest, tongue, etc. Fig 1 below shows the proposed multi-stage TL framework with four stages using LWA. Additionally, conducting a systematic literature review with a focus on LWAs would be beneficial since there seems a lack of systemic literature reviews that synthesize the architectural evolution and delineate the differences between standard CNNs and LWAs.

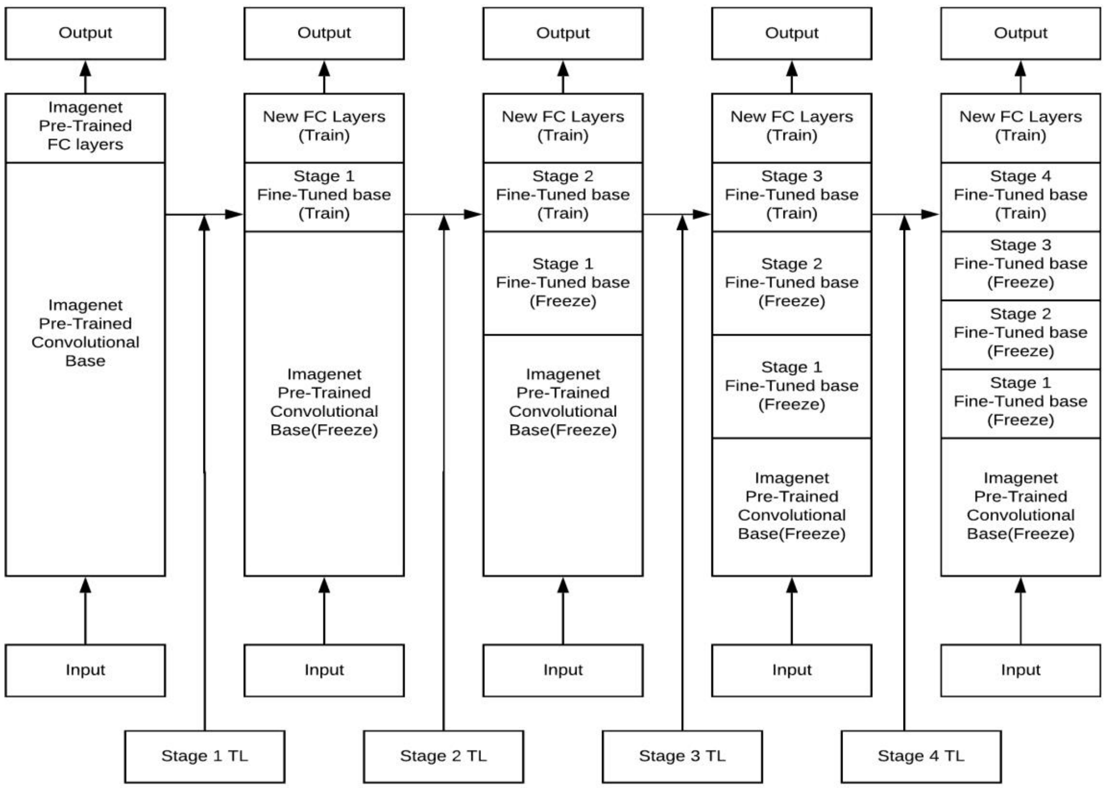

Figure 1. Multi-Stage TL System

REFERENCES

1. Altaf, F., Islam, S. M. S., Akhtar, N., and Janjua, N. K. 2019. “Going Deep in Medical Image Analysis: Concepts, Methods, Challenges and Future Directions,” *ArXiv:1902.05655 [Cs]*. (http://arxiv.org/abs/1902.05655).

2. Dawud, A. M., Yurtkan, K., and Oztoprak, H. 2019. “Application of Deep Learning in Neuroradiology: Brain Haemorrhage Classification Using Transfer Learning,” *Computational Intelligence and Neuroscience* (2019), pp. 1–12. (https://doi.org/10.1155/2019/4629859).

3. Ech-Cherif, A., Misbhauddin, M., and Ech-Cherif, M. 2019. “Deep Neural Network Based Mobile Dermoscopy Application for Triaging Skin Cancer Detection,” in *2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)*, Riyadh, Saudi Arabia: IEEE, May, pp. 1–6. (https://doi.org/10.1109/CAIS.2019.8769517).

4. Eminaga, O., Abbas, M., Kunder, C., Loening, A. M., Shen, J., Brooks, J. D., Langlotz, C. P., and Rubin, D. L. 2019. “Plexus Convolutional Neural Network (PlexusNet): A Novel Neural Network Architecture for Histologic Image Analysis,” *ArXiv:1908.09067 [Cs, Eess, q-Bio]*. (http://arxiv.org/abs/1908.09067).

5. Erickson, B. J., Korfiatis, P., Kline, T. L., Akkus, Z., Philbrick, K., and Weston, A. D. 2018. “Deep Learning in Radiology: Does One Size Fit All?,” *Journal of the American College of Radiology* (15:3), pp. 521–526. (https://doi.org/10.1016/j.jacr.2017.12.027).

6. François, C. 2017. *Deep Learning with Python*, Manning Publications Company.

7. Fukushima, K. 1980. “Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position,” *Biological Cybernetics* (36:4), pp. 193–202. (https://doi.org/10.1007/BF00344251).

8. Gao, J., Jiang, Q., Zhou, B., Chen, D., 1 College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China., and 2 Shanghai University of Medicine & Health Science, Shanghai 201308, China. 2019. “Convolutional Neural Networks for Computer-Aided Detection or Diagnosis in Medical Image Analysis: An Overview,” *Mathematical Biosciences and Engineering* (16:6), pp. 6536–6561. (https://doi.org/10.3934/mbe.2019326).

9. He, K., Zhang, X., Ren, S., and Sun, J. 2016. “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, June, pp. 770–778. (https://doi.org/10.1109/CVPR.2016.90).

10. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. 2017. “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” *ArXiv:1704.04861 [Cs]*. (http://arxiv.org/abs/1704.04861).

11. Huang, G., Liu, Z., Maaten, L. van der, and Weinberger, K. Q. 2017. “Densely Connected Convolutional Networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI: IEEE, July, pp. 2261–2269. (https://doi.org/10.1109/CVPR.2017.243).

12. Iandola, F. N., Ashraf, K., Moskewicz, M. W., and Keutzer, K. 2016. “FireCaffe: Near-Linear Acceleration of Deep Neural Network Training on Compute Clusters,” *ArXiv:1511.00175 [Cs]*. (http://arxiv.org/abs/1511.00175).

13. Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., and Keutzer, K. 2016. “SqueezeNet: AlexNet-Level Accuracy with 50x Fewer Parameters and <0.5MB Model Size,” *ArXiv:1602.07360 [Cs]*. (http://arxiv.org/abs/1602.07360).

14. Khan, A., Sohail, A., Zahoora, U., and Qureshi, A. S. 2019. “A Survey of the Recent Architectures of Deep Convolutional Neural Networks,” *ArXiv:1901.06032 [Cs]*. (http://arxiv.org/abs/1901.06032).

15. Kim, H. G., Choi, Y., and Ro, Y. M. 2017. “Modality-Bridge Transfer Learning for Medical Image Classification,” *ArXiv:1708.03111 [Cs]*. (http://arxiv.org/abs/1708.03111).

16. Krizhevsky, A., Sutskever, I., and Hinton, G. E. 2012. “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., pp. 1097–1105. (http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf).

17. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., and Sánchez, C. I. 2017. “A Survey on Deep Learning in Medical Image Analysis,” *Medical Image Analysis* (42), pp. 60–88. (https://doi.org/10.1016/j.media.2017.07.005).

18. Majtner, T., Bajić, B., Yildirim, S., Hardeberg, J. Y., Lindblad, J., and Sladoje, N. 2018. “Ensemble of Convolutional Neural Networks for Dermoscopic Images Classification,” *ArXiv:1808.05071 [Cs]*. (http://arxiv.org/abs/1808.05071).

19. Menegola, A., Fornaciali, M., Pires, R., Bittencourt, F. V., Avila, S., and Valle, E. 2017. “Knowledge Transfer for Melanoma Screening with Deep Learning,” *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 297–300. (https://doi.org/10.1109/ISBI.2017.7950523).

20. Pan, S. J., and Yang, Q. 2010. “A Survey on Transfer Learning,” *IEEE Transactions on Knowledge and Data Engineering* (22:10), pp. 1345–1359. (https://doi.org/10.1109/TKDE.2009.191).

21. Raghu, M., Zhang, C., Kleinberg, J., and Bengio, S. 2019. “Transfusion: Understanding Transfer Learning for Medical Imaging,” *ArXiv:1902.07208 [Cs, Stat]*. (http://arxiv.org/abs/1902.07208).

22. Samala, R. K., Chan, H.-P., Hadjiiski, L., Helvie, M. A., Richter, C. D., and Cha, K. H. 2019. “Breast Cancer Diagnosis in Digital Breast Tomosynthesis: Effects of Training Sample Size on Multi-Stage Transfer Learning Using Deep Neural Nets,” *IEEE Transactions on Medical Imaging* (38:3), pp. 686–696. (https://doi.org/10.1109/TMI.2018.2870343).

23. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. 2019. “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” *ArXiv:1801.04381 [Cs]*. (http://arxiv.org/abs/1801.04381).

24. Shamim, M., Syed, S., Shiblee, M., Usman, M., and Ali, S. (n.d.). *Automated Detection of Oral Pre-Cancerous Tongue Lesions Using Deep Learning for Early Diagnosis of Oral Cavity Cancer.*, p. 25.

25. Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., and Summers, R. M. 2016. “Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning.,” *IEEE Transactions on Medical Imaging* (35:5), United States, pp. 1285–1298. (https://doi.org/10.1109/TMI.2016.2528162).

26. Simonyan, K., and Zisserman, A. 2014. “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *ArXiv:1409.1556 [Cs]*. (http://arxiv.org/abs/1409.1556).

27. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. 2015. “Rethinking the Inception Architecture for Computer Vision,” *ArXiv:1512.00567 [Cs]*. (http://arxiv.org/abs/1512.00567).

28. Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. 2015. “Going Deeper with Convolutions,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA: IEEE, June, pp. 1–9. (https://doi.org/10.1109/CVPR.2015.7298594).

29. Tan, M., and Le, Q. V. 2019. “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” *ArXiv:1905.11946 [Cs, Stat]*. (http://arxiv.org/abs/1905.11946).

30. Webster, J., and Watson, R. T. 2002. “Analyzing the Past to Prepare for the Future: Writing a Literature Review,” *MIS Quarterly* (26:2), Management Information Systems Research Center, University of Minnesota, xiii–xxiii.

31. Wong, K. C. L., Syeda-Mahmood, T., and Moradi, M. 2018. “Building Medical Image Classifiers with Very Limited Data Using Segmentation Networks,” *Medical Image Analysis* (49), pp. 105–116. (https://doi.org/10.1016/j.media.2018.07.010).

32. Wu, P., Qu, H., Yi, J., Huang, Q., Chen, C., and Metaxas, D. 2019. “Deep Attentive Feature Learning for Histopathology Image Classification,” in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Venice, Italy: IEEE, April, pp. 1865–1868. (https://doi.org/10.1109/ISBI.2019.8759267).

33. Xu, X., Lin, J., Tao, Y., and Wang, X. 2018. “An Improved DenseNet Method Based on Transfer Learning for Fundus Medical Images,” in *2018 7th International Conference on Digital Home (ICDH)*, Guilin, China: IEEE, November, pp. 137–140. (https://doi.org/10.1109/ICDH.2018.00033).

34. Yamashita, R., Nishio, M., Do, R. K. G., and Togashi, K. 2018. “Convolutional Neural Networks: An Overview and Application in Radiology,” *Insights into Imaging* (9:4), pp. 611–629.

35. Yu, Y., Lin, H., Meng, J., Wei, X., Guo, H., and Zhao, Z. 2017. “Deep Transfer Learning for Modality Classification of Medical Images,” *Information* (8:3), Multidisciplinary Digital Publishing Institute, p. 91.

36. Zhang, X., Zhou, X., Lin, M., and Sun, J. 2017. “ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices,” *ArXiv:1707.01083 [Cs]*. (http://arxiv.org/abs/1707.01083).

37. Zoph, B., Vasudevan, V., Shlens, J., and Le, Q. V. 2018. “Learning Transferable Architectures for Scalable Image Recognition,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT: IEEE, June, pp. 8697–8710. (https://doi.org/10.1109/CVPR.2018.00907).