How Does Pruning Impact Long-Tailed Multi-label Medical Image Classifiers?

Gregory Holste[1], Ziyu Jiang[2], Ajay Jaiswal[1], Maria Hanna[3],
Shlomo Minkowitz[3], Alan C. Legasto[3], Joanna G. Escalon[3],
Sharon Steinberger[3], Mark Bittman[3], Thomas C. Shen[4], Ying Ding[1],
Ronald M. Summers[4], George Shih[3(✉)], Yifan Peng[3(✉)],
and Zhangyang Wang[1(✉)]

[1] The University of Texas at Austin, Austin, TX, USA
`atlaswang@utexas.edu`
[2] Texas A&M University, College Station, TX, USA
[3] Weill Cornell Medicine, New York, NY, USA
`yip4002@med.cornell.edu`
[4] Clinical Center, National Institutes of Health, Bethesda, MD, USA

Abstract. Pruning has emerged as a powerful technique for compressing deep neural networks, reducing memory usage and inference time without significantly affecting overall performance. However, the nuanced ways in which pruning impacts model behavior are not well understood, particularly for *long-tailed*, *multi-label* datasets commonly found in clinical settings. This knowledge gap could have dangerous implications when deploying a pruned model for diagnosis, where unexpected model behavior could impact patient well-being. To fill this gap, we perform the first analysis of pruning's effect on neural networks trained to diagnose thorax diseases from chest X-rays (CXRs). On two large CXR datasets, we examine which diseases are most affected by pruning and characterize class "forgettability" based on disease frequency and co-occurrence behavior. Further, we identify individual CXRs where uncompressed and heavily pruned models disagree, known as pruning-identified exemplars (PIEs), and conduct a human reader study to evaluate their unifying qualities. We find that radiologists perceive PIEs as having more label noise, lower image quality, and higher diagnosis difficulty. This work represents a first step toward understanding the impact of pruning on model behavior in deep long-tailed, multi-label medical image classification. All code, model weights, and data access instructions can be found at https://github.com/VITA-Group/PruneCXR.

Keywords: Pruning · Chest X-Ray · Imbalance · Long-Tailed Learning

1   Introduction

Deep learning has enabled significant progress in image-based computer-aided diagnosis [8,10,23,26,33]. However, the increasing memory requirements of deep neural networks limit their practical deployment in hardware-constrained environments. One promising approach to reducing memory usage and inference latency is model **pruning**, which aims to remove redundant or unimportant model weights [21]. Since modern deep neural networks are often overparameterized, they can be heavily pruned with minimal impact on overall performance [6,20,22,34]. This being said, the impact of pruning on model behavior *beyond* high-level performance metrics like top-1 accuracy remain unclear. This gap in understanding has major implications for real-world deployment of neural networks for high-risk tasks like disease diagnosis, where pruning may cause unexpected consequences that could potentially threaten patient well-being.

To bridge this gap, this study aims to answer the following guiding questions by conducting experiments to dissect the differential impact of pruning:

Q1. What is the impact of pruning on overall performance in long-tailed multi-label medical image classification?
Q2. Which disease classes are most affected by pruning and why?
Q3. How does disease co-occurrence influence the impact of pruning?
Q4. Which individual images are most vulnerable to pruning?

We focus our experiments on thorax disease classification on chest X-rays (CXRs), a challenging **long-tailed** and **multi-label** computer-aided diagnosis problem, where patients may present with multiple abnormal findings in one exam and most findings are rare relative to the few most common diseases [12].

This study draws inspiration from Hooker *et al.* [13], who found that pruning disparately impacts a small subset of classes in order to maintain overall performance. The authors also introduced **pruning-identified exemplars (PIEs)**, images where an uncompressed and heavily pruned model disagree. They discovered that PIEs share common characteristics such as multiple salient objects and noisy, fine-grained labels. While these findings uncover what neural networks "forget" upon pruning, the insights are limited to highly curated natural image datasets where each image belongs to one class. Previous studies have shown that pruning can enhance fairness [31], robustness [1], and efficiency for medical image classification [4,7,32] and segmentation [3,16,24,25,29] tasks. However, these efforts also either focused solely on high-level performance or did not consider settings with severe class imbalance or co-occurrence.

Unlike existing work, we explicitly connect class "forgettability" to the unique aspects of our problem setting: disease frequency (long-tailedness) and disease co-occurrence (multi-label behavior). Since many diagnostic exams, like CXR, are long-tailed and multi-label, this work fills a critical knowledge gap enabling more informed deployment of pruned disease classifiers. We hope that our findings can provide a foundation for future research on pruning in clinically realistic settings.

2 Methods

2.1 Preliminaries

Datasets. For this study, we use expanded versions of NIH ChestXRay14 [30] and MIMIC-CXR [18], two large-scale CXR datasets for multi-label disease classification.[1] As described in Holste *et al.* [12], we augmented the set of possible labels for each image by adding five new rare disease findings parsed from radiology reports. This creates a challenging long-tailed classification problem, with training class prevalence ranging from under 100 to over 70,000 (Supplement). **NIH-CXR-LT** contains 112,120 CXRs, each labeled with at least one of 20 classes, while **MIMIC-CXR-LT** contains 257,018 frontal CXRs labeled with at least one of 19 classes. Each dataset was split into training (70%), validation (10%), and test (20%) sets at the patient level.

Model Pruning & Evaluation. Following Hooker *et al.* [13], we focus on global unstructured **L1 pruning** [34]. After training a disease classifier, a fraction k of weights with the smallest magnitude are "pruned" (set to zero); for instance, $k = 0.9$ means 90% of weights have been pruned. While area under the receiver operating characteristic curve is a standard metric on related datasets [26, 28, 30], it can become heavily inflated in the presence of class imbalance [2, 5]. Since we seek a metric that is both resistant to imbalance and captures performance across thresholds (as choosing a threshold is non-trivial in the multi-label setting [27]), we use **average precision (AP)** as our primary metric.

2.2 Assessing the Impact of Pruning

Experimental Setup. We first train a baseline model to classify thorax diseases on both NIH-CXR-LT and MIMIC-CXR-LT. The architecture used was a ResNet50 [9] with ImageNet-pretrained weights and a sigmoid cross-entropy loss. For full training details, please see the Supplemental Materials and code repository. Following Hooker *et al.* [13], we then repeat this process with 30 unique random initializations, performing L1 pruning at a range of sparsity ratios $k \in \{0, 0.05, \ldots, 0.9, 0.95\}$ on each model and dataset. Using a "population" of 30 models allows for reliable estimation of model performance at each sparsity ratio. We then analyze how pruning impacts overall, disease-level, and image-level model behavior with increasing sparsity as described below.

Overall and Class-Level Analysis. To evaluate the overall impact of pruning, we compute the mean AP across classes for each sparsity ratio and dataset. We use Welch's t-test to assess performance differences between the 30 uncompressed models and 30 k-sparse models. We then characterize the class-level impact of pruning by considering the relative change in AP from an uncompressed model

---

[1] NIH ChestXRay14 can be found here, and MIMIC-CXR can be found here.

to its k-sparse counterpart for all k. Using relative change in AP allows for comparison of the *impact* of pruning regardless of class difficulty. We then define the **forgettability curve** of a class c as follows:

$$\left\{ \text{med} \left\{ \frac{\text{AP}_{i,k,c} - \text{AP}_{i,0,c}}{\text{AP}_{i,0,c}} \right\}_{i \in \{1,\dots,30\}} \right\}_{k \in \{0,0.05,\dots,0.9,0.95\}} \tag{1}$$

where $\text{AP}_{i,k,c} :=$ AP of the ith model with sparsity k on class c, and $\text{med}(\cdot) :=$ median across all 30 runs. We analyze how these curves relate to class frequency and co-occurrence using Pearson (r) and Spearman (ρ) correlation tests.
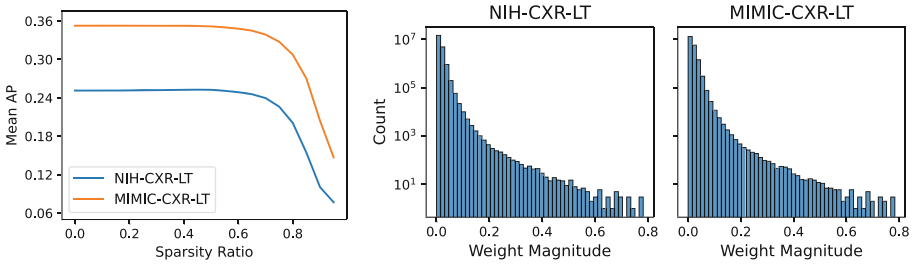
Fig. 1. Overall effect of pruning on disease classification performance. Presented is the mean AP (median across 30 runs) for sparsity ratios $k \in \{0, \dots, 0.95\}$ (left) and log-scale histogram of model weight magnitudes (right).

Incorporating Disease Co-occurrence Behavior. For each *unique pair* of NIH-CXR-LT classes, we compute the **Forgettability Curve Dissimilarity (FCD)**, the mean squared error (MSE) between the forgettability curves of each disease. FCD quantifies how similar two classes are with respect to their forgetting behavior over all sparsity ratios. Ordinary least squares (OLS) linear regression is employed to understand the interaction between difference in class frequency and class co-occurrence with respect to FCD for a given disease pair.

2.3   Pruning-Identified Exemplars (PIEs)

Definition. After evaluating the overall and class-level impact of pruning on CXR classification, we investigate which individual images are most vulnerable to pruning. Like Hooker *et al.* [13], we consider PIEs to be images where an uncompressed and pruned model disagree. Letting C be the number of classes, we compute the average prediction $\frac{1}{30} \sum_i \hat{\boldsymbol{y}}_{\mathbf{0}} \in \mathbb{R}^C$ of the uncompressed models and average prediction $\frac{1}{30} \sum_i \hat{\boldsymbol{y}}_{\mathbf{0.9}} \in \mathbb{R}^C$ of the L1-pruned models at 90% sparsity for all NIH-CXR-LT test set images. Then the Spearman rank correlation $\sigma(\frac{1}{30} \sum_i \hat{\boldsymbol{y}}_{\mathbf{0}}, \frac{1}{30} \sum_i \hat{\boldsymbol{y}}_{\mathbf{0.9}})$ represents the agreement between the uncompressed and heavily pruned models for each image; we define PIEs as images whose correlation falls in the bottom 5th percentile of test images.

Analysis and Human Study. To understand the common characteristics of PIEs, we compare how frequently (i) each class appears and (ii) images with $d = 0, \ldots, 3, 4+$ simultaneous diseases appear in PIEs relative to non-PIEs. To further analyze qualities of CXRs that require domain expertise, we conducted a human study to assess radiologist perceptions of PIEs. Six board-certified attending radiologists were each presented with a unique set of 40 CXRs (half PIE, half non-PIE). Each image was presented along with its ground-truth labels and the following three questions:

1. **Do you fully agree with the given label?** [Yes/No]
2. **How would you rate the image quality?** [1–5 Likert]
3. **How difficult is it to properly diagnose this image?** [1–5 Likert]

We use the Kruskal-Wallis test to evaluate differential perception of PIEs.
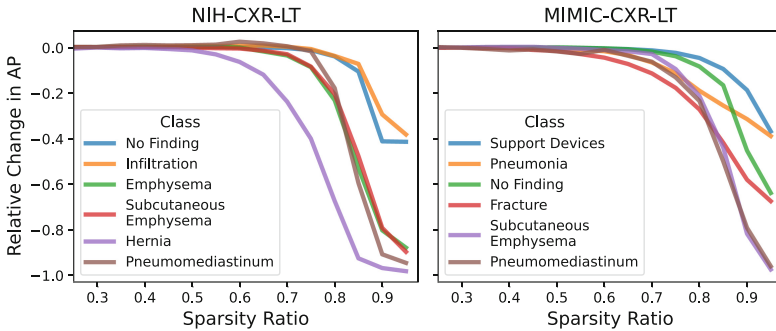
Fig. 2. "Forgettability curves" depicting relative change in AP (median across 30 runs at each sparsity ratio) upon L1 pruning for a subset of classes.

3 Results

3.1 What is the Overall Effect of Pruning?

We find that under L1 pruning, the first sparsity ratio causing a significant drop in mean AP is 65% for NIH-CXR-LT ($P < 0.001$) and 60% for MIMIC-CXR-LT ($P < 0.001$) (Fig. 1, left). This observation may be explained by the fact that ResNet50 is highly overparameterized for this task. Since only a subset of weights are required to adequately model the data, the trained classifiers have naturally sparse activations (Fig. 1, right). For example, over half of all learned weights have magnitude under 0.01. However, beyond a sparsity ratio of 60%, we observe a steep decline in performance with increasing sparsity for both datasets.
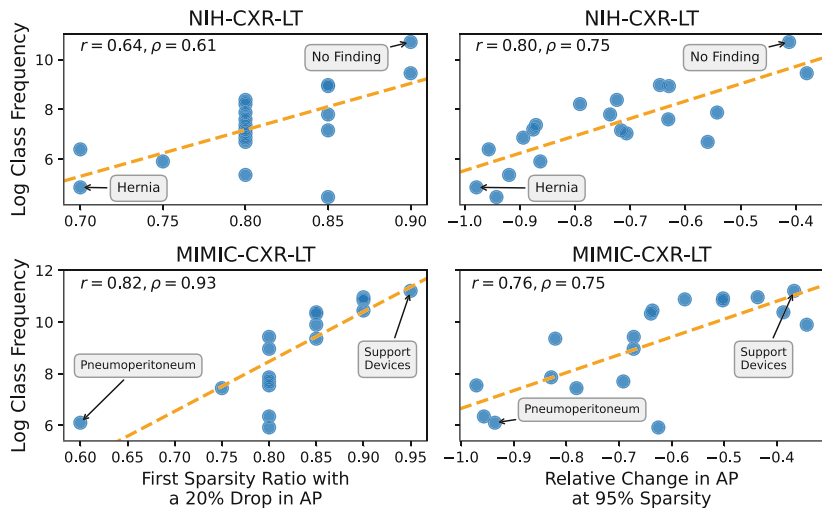
Fig. 3. Relationship between class "forgettability" and frequency. We characterize which classes are forgotten *first* (left) and which are *most* forgotten (right).
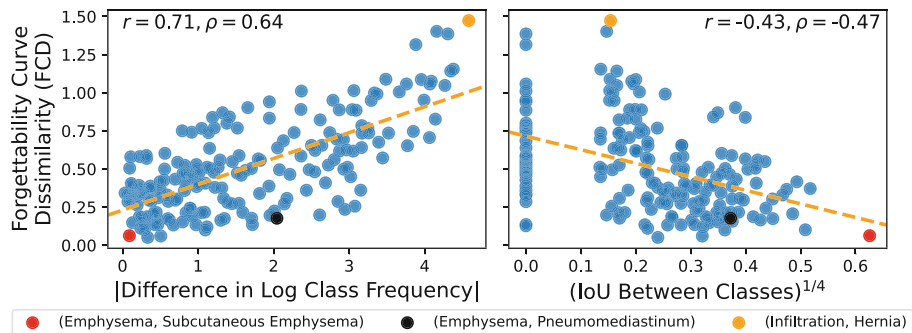
Fig. 4. Mutual relationship between pairs of diseases and their forgettability curves. For each pair of NIH-CXR-LT classes, FCD is plotted against the absolute difference in log frequency (left) and the IoU between the two classes (right).

3.2   Which Diseases are Most Vulnerable to Pruning and Why?

Class forgettability curves in Fig. 2 depict the relative change in AP by sparsity ratio for a representative subset of classes. Although these curves follow a similar general trend to Fig. 1, some curves (i) drop earlier and (ii) drop more considerably at high sparsity. Notably, we find a strong positive relationship between training class frequency and (i) the first sparsity ratio at which a class experienced a median 20% relative drop in AP ($\rho = 0.61, P = 0.005$ for NIH-CXR-LT; $\rho = 0.93, P \ll 0.001$ for MIMIC-CXR-LT) and (ii) the median relative change in AP at 95% sparsity ($\rho = 0.75, P < 0.001$ for NIH-CXR-LT; $\rho = 0.75, P < 0.001$ for MIMI-CXR-LT). These findings indicate that, in general, **rare diseases are**

forgotten earlier (Fig. 3, left) **and are more severely impacted at high sparsity** (Fig. 3, right).

3.3 How Does Disease Co-occurrence Influence Class Forgettability?

Our analysis reveals that for NIH-CXR-LT, the absolute difference in log test frequency between two diseases is a strong predictor of the pair's FCD ($\rho = 0.64, P \ll 0.001$). This finding suggests that **diseases with larger differences in prevalence exhibit more distinct forgettability behavior** upon L1 pruning (Fig. 4, left). To account for the multi-label nature of thorax disease classification, we also explore the relationship between intersection over union (IoU) – a measure of co-occurrence between two diseases – and FCD. Our analysis indicates that the IoU between two diseases is negatively associated with FCD ($\rho = -0.47, P \ll 0.001$). This suggests that **the more two diseases co-occur, the more similar their forgetting trajectories are across all sparsity ratios** (Fig. 4, right). For example, the disease pair (Infiltration, Hernia) has a dramatic difference in prevalence ($|\text{LogFreqDiff}| = 4.58$) *and* rare co-occurrence ($\text{IoU}^{1/4} = 0.15$), resulting in an extremely high FCD for the pair of diseases.

We also find, however, that there is a push and pull between differences in individual class frequency and class co-occurrence with respect to FCD. To illustrate, consider the disease pair (Emphysema, Pneumomediastinum) marked in black in Fig. 4. These classes have an absolute difference in log frequency of 2.04, which would suggest an FCD of around 0.58. However, because Emphysema and Pneumomediastinum co-occur relatively often ($\text{IoU}^{1/4} = 0.37$), their forgettability curves are *more similar than prevalence alone would dictate*, resulting in a lower FCD of 0.18. To quantify this effect, we obtain an OLS model that fitted FCD as a function of $|\text{LogFreqDiff}|$, $\text{IoU}^{1/4}$, and their interaction:

$$\text{FCD} = 0.27 + 0.21|\text{LogFreqDiff}| - 0.05(\text{IoU})^{1/4} - 0.31|\text{LogFreqDiff}| * (\text{IoU})^{1/4} \quad (2)$$

We observe a statistically significant interaction effect between the difference in individual class frequency and class co-occurrence on FCD ($\beta_3 = -0.31, P = 0.005$). Thus, for disease pairs with a very large difference in prevalence, the effect of co-occurrence on FCD is even more pronounced (Supplement).

3.4 What Do Pruning-Identified CXRs have in Common?

For NIH-CXR-LT, we find that **PIEs are more likely to contain rare diseases** and **more likely to contain 3+ simultaneous diseases** when compared to non-PIEs (Fig. 5). The five rarest classes appear 3–15x more often in PIEs than non-PIEs, and images with 4+ diseases appear 3.2x more often in PIEs.

In a human reader study involving 240 CXRs from the NIH-CXR-LT test set (120 PIEs and 120 non-PIEs), **radiologists perceived that PIEs had more label noise, lower image quality, and higher diagnosis difficulty** (Fig. 6). However, due to small sample size and large variability, these differences are not statistically significant. Respondents fully agreed with the label 55% of the time
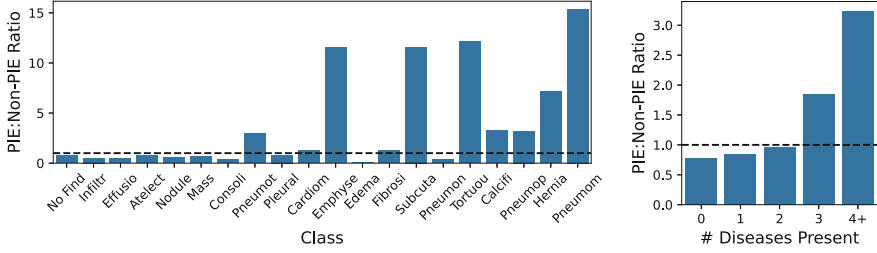
Fig. 5. Unique characteristics of PIEs. Presented is the *ratio* of class prevalence (left) and number of diseases per image (right) in PIEs relative to non-PIEs. The dotted line represents the 1:1 ratio (equally frequent in PIEs vs. non-PIEs).
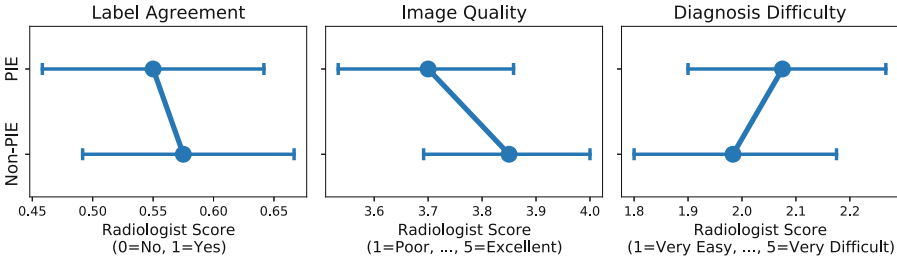
Fig. 6. Human study results describing radiologist perception of PIEs vs. non-PIEs. Mean ± standard deviation (error bar) radiologist scores are presented.

for PIEs and 57.5% of the time for non-PIEs ($P = 0.35$), gave an average image quality of 3.6 for PIEs and 3.8 for non-PIEs ($P = 0.09$), and gave an average diagnosis difficulty of 2.5 for PIEs and 2.05 for non-PIEs ($P = 0.25$).

Overall, these findings suggest that pruning identifies CXRs with many potential sources of difficulty, such as containing underrepresented diseases, (partially) incorrect labels, low image quality, and complex disease presentation.

4   Discussion and Conclusion

In conclusion, we conducted the first study of the effect of pruning on multi-label, long-tailed medical image classification, focusing on thorax disease diagnosis in CXRs. Our findings are summarized as follows:

1. As observed in standard image classification, CXR classifiers can be heavily pruned (up to 60% sparsity) before dropping in overall performance.
2. Class frequency is a strong predictor of both *when* and *how severely* a class is impacted by pruning. Rare classes suffer the most.
3. Large differences in class frequency lead to dissimilar "forgettability" behavior and stronger co-occurrence leads to more similar forgettability behavior.
   – Further, we discover a significant interaction effect between these two factors with respect to how similarly pruning impacts two classes.

4. We adapt PIEs to the multi-label setting, observing that PIEs are far more likely to contain rare diseases and multiple concurrent diseases.
   – A radiologist study further suggests that PIEs have more label noise, lower image quality, and higher diagnosis difficulty.

It should be noted that this study is limited to the analysis of global unstructured L1 (magnitude-based) pruning, a simple heuristic for post-training network pruning. Meanwhile, other state-of-the-art pruning approaches [6,20,22] and model compression techniques beyond pruning (e.g., weight quantization [14] and knowledge distillation [11]) could be employed to strengthen this work. Additionally, since our experiments only consider the ResNet50 architecture, it remains unclear whether other training approaches, architectures, or compression methods could mitigate the adverse effects of pruning on rare classes. In line with recent work [15,17,19], future research may leverage the insights gained from this study to develop an algorithm for improved long-tailed learning on medical image analysis tasks. For example, PIEs could be interpreted as salient, difficult examples that warrant greater weight during training. Conversely, PIEs may just as well be regarded as noisy examples to be ignored, using pruning as a tool for data cleaning.

References

1. Chen, L., Zhao, L., Chen, C.Y.C.: Enhancing adversarial defense for medical image analysis systems with pruning and attention mechanism. Med. Phys. **48**(10), 6198–6212 (2021)
2. Davis, J., Goadrich, M.: The relationship between precision-recall and roc curves. In: ICML, pp. 233–240 (2006)
3. Dinsdale, N.K., Jenkinson, M., Namburete, A.I.: Stamp: simultaneous training and model pruning for low data regimes in medical image segmentation. Med. Image Anal. **81**, 102583 (2022)
4. Fernandes, F.E., Yen, G.G.: Automatic searching and pruning of deep neural networks for medical imaging diagnostic. IEEE Trans. Neural Netw. Learn. Syst. **32**(12), 5664–5674 (2020)
5. Fernández, A., García, S., Galar, M., Prati, R.C., Krawczyk, B., Herrera, F.: Learning from Imbalanced Data Sets. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-98074-4
6. Frankle, J., Carbin, M.: The lottery ticket hypothesis: finding sparse, trainable neural networks. arXiv preprint arXiv:1803.03635 (2018)
7. Hajabdollahi, M., Esfandiarpoor, R., Najarian, K., Karimi, N., Samavi, S., Soroushmehr, S.R.: Hierarchical pruning for simplification of convolutional neural networks in diabetic retinopathy classification. In: IEEE Engineering in Medicine and Biology Society (EMBC), pp. 970–973. IEEE (2019)

8. Han, Y., Holste, G., Ding, Y., Tewfik, A., Peng, Y., Wang, Z.: Radiomics-Guided Global-Local Transformer For Weakly Supervised Pathology Localization in Chest X-rays. IEEE Trans. Med, Imaging PP (Oct (2022)

9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 770–778 (2016)

10. Hesamian, M.H., Jia, W., He, X., Kennedy, P.: Deep learning techniques for medical image segmentation: achievements and challenges. J. Digit. Imaging **32**, 582–596 (2019)

11. Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 (2015)

12. Holste, G., et al.: Long-tailed classification of thorax diseases on chest x-ray: A new benchmark study. In: Data Augmentation, Labelling, and Imperfections: Second MICCAI Workshop, pp. 22–32. Springer (2022). https://doi.org/10.1007/978-3-031-17027-0_3

13. Hooker, S., Courville, A., Clark, G., Dauphin, Y., Frome, A.: What do compressed deep neural networks forget? arXiv preprint arXiv:1911.05248 (2019)

14. Jacob, B., et al.: Quantization and training of neural networks for efficient integer-arithmetic-only inference. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2704–2713 (2018)

15. Jaiswal, A., Chen, T., Rousseau, J.F., Peng, Y., Ding, Y., Wang, Z.: Attend who is weak: Pruning-assisted medical image localization under sophisticated and implicit imbalances. In: WACV, pp. 4987–4996 (2023)

16. Jeong, T., Bollavaram, M., Delaye, E., Sirasao, A.: Neural network pruning for biomedical image segmentation. In: Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling. vol. 11598, pp. 415–425. SPIE (2021)

17. Jiang, Z., Chen, T., Mortazavi, B.J., Wang, Z.: Self-damaging contrastive learning. In: International Conference on Machine Learning, pp. 4927–4939. PMLR (2021)

18. Johnson, A.E., et al.: MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. Scientific Data (2019)

19. Kong, H., Lee, G.H., Kim, S., Lee, S.W.: Pruning-guided curriculum learning for semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 5914–5923 (2023)

20. Kurtic, E., Alistarh, D.: Gmp*: Well-tuned global magnitude pruning can outperform most bert-pruning methods. arXiv preprint arXiv:2210.06384 (2022)

21. LeCun, Y., Denker, J., Solla, S.: Optimal brain damage. Adv. Neural Inform. Process. Syst. **2** (1989)

22. Lee, N., Ajanthan, T., Torr, P.H.: Snip: single-shot network pruning based on connection sensitivity. arXiv preprint arXiv:1810.02340 (2018)

23. Lin, M., et al.: Automated diagnosing primary open-angle glaucoma from fundus image by simulating human's grading with deep learning. Sci. Rep. **12**(1), 14080 (2022)

24. Lin, X., Yu, L., Cheng, K.T., Yan, Z.: The lighter the better: rethinking transformers in medical image segmentation through adaptive pruning. arXiv preprint arXiv:2206.14413 (2022)

25. Mahbod, A., Entezari, R., Ellinger, I., Saukh, O.: Deep neural network pruning for nuclei instance segmentation in hematoxylin and eosin-stained histological images. In: Wu, S., Shabestari, B., Xing, L. (eds.) Applications of Medical Artificial Intelligence: First International Workshop, AMAI 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18, 2022, Proceedings, pp. 108–117. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-17721-7_12

26. Rajpurkar, P., et al.: Chexnet: radiologist-level pneumonia detection on chest -X-rays with deep learning. arXiv preprint arXiv:1711.05225 (2017)
27. Rethmeier, N., Augenstein, I.: Long-tail zero and few-shot learning via contrastive pretraining on and for small data. In: Computer Sciences & Mathematics Forum. vol. 3, p. 10. MDPI (2022)
28. Seyyed-Kalantari, L., Liu, G., McDermott, M., Chen, I.Y., Ghassemi, M.: Chexclusion: fairness gaps in deep chest x-ray classifiers. In: BIOCOMPUTING 2021: proceedings of the Pacific symposium, pp. 232–243. World Scientific (2020)
29. Valverde, J.M., Shatillo, A., Tohka, J.: Sauron u-net: Simple automated redundancy elimination in medical image segmentation via filter pruning. arXiv preprint arXiv:2209.13590 (2022)
30. 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. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 3462–3471 (2017)
31. Wu, Y., Zeng, D., Xu, X., Shi, Y., Hu, J.: FairPrune: achieving fairness through pruning for dermatological disease diagnosis. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part I, pp. 743–753. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16431-6_70
32. Yang, B., et al.: Network pruning for OCT image classification. In: Fu, H., Garvin, M.K., MacGillivray, T., Xu, Y., Zheng, Y. (eds.) Ophthalmic Medical Image Analysis: 6th International Workshop, OMIA 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, Proceedings, pp. 121–129. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32956-3_15
33. Zhou, S.K., et al.: A review of deep learning in medical imaging: imaging traits, technology trends, case studies with progress highlights, and future promises. Proc. IEEE **109**(5), 820–838 (2021)
34. Zhu, M., Gupta, S.: To prune, or not to prune: exploring the efficacy of pruning for model compression. arXiv preprint arXiv:1710.01878 (2017)