

000  
001  
002  
003  
004  
005

Lung Disease Classification - Applied AI Progress Report

Group-Q

## Abstract

## 1. Introduction-1.5

Early diagnosis of respiratory diseases like pneumonia and COVID-19 leads to decreased mortality rate [7] and is a powerful way to manage a pandemic [38]. These diseases can be diagnosed using a variety of tests like pulse oximetry, chest x-ray, CT scan [29], PCR [1] however chest X-rays are by far the most accessible [9] to low and middle income countries. Furthermore, the scan is available in minutes making it one of the fastest ways of diagnosis [30]. However, the bottleneck with this method is the need for an expert radiologists to evaluate the scan [20]. Many researchers have tried to solve this problem by creating a deep learning based lung disease classification system [34] but haven't been able to come up with models that can replace radiologists. Small [12] and highly imbalanced data [34], along with varying specifications of X-ray scanners leading to low inter-hospital accuracy [26] are the biggest problems that researchers have faced. Another issue with using deep neural networks in medical settings is its black-box nature [3], doctors and patients will not trust a model that cannot explain its results [21].

This project is an attempt to compare three CNN backbone architectures namely, ResNet-34, MobileNet V3 Large and EfficientNet B1 along with three lung disease datasets to identify the type of architecture that works best for lung disease classification. Two

of the datasets used presented a multiclass classification problem with 3 classes while the third dataset presented a multiclass, multilabel classification problem. A total of 12 models were trained in this study, four for each of the three datasets. The first three models for each dataset was trained from scratch and the fourth model was trained using transfer learning. Transfer learning was performed by deep-tuning ImageNet weights and the performance was evaluated to check improvement over the models trained from scratch. The small dataset problem and the issue of different radiographic contrast [22] is mitigated using data augmentation. Imbalanced data problem is handled by undersampling the majority class. The hyperparameters were fixed across models and the F1 scores and cross entropy loss have been used to compare models and select the best overall model. All the models were optimized using the Adam optimizer [18] with default parameters and the cosine annealing [19] learning rate scheduler was used to decrease the learning rate as training progressed. Further, an ablation study was performed to find the best learning rate for the selected model. Finally, GradCAM [11] and T-SNE were used to visualize the trained models and understand model predictions better. An F1 score of 0.8 and 0.98 was achieved for the two multiclass datasets, whereas the maximum F1 for the multilabel dataset with 7 classes was 0.46.

**Related Works:** Li *et al.* [1] were among the first to use CNNs in a medical setting. They used a single convolutional layer to classify interstitial lung diseases using CT scans, achieving better performance than existing approaches. Since then there has been a dramatic increase in application of CNNs in healthcare, deep neural networks have been used to perform various tasks like segmenting regions of interest in MRI [2], classifying X-Ray [3], MRI [4], and CT [5] scans. Further, GANs have been used to generate high quality scans [6] when there is a lack of available data due to either privacy reasons or availability of subjects. GANs have also been used to generate high quality CT scans from MRI scans [7]. Apart from radiographic scans, deep CNNs have also been used to detect malarial parasite in blood smear images [8] with an accuracy of 99.96%. Another interesting application is the use of 1-D convolutions to detect heart anomalies using ECG data [9]. Researchers have also used architectures like the Inception V3 to

108	Dataset	No. of Images	Classes	Size
109	COVID [5, 6, 25]	10000k:3.6k:1.3k	3	299 <sup>2</sup>
110	Pneumonia [17, 33]	3k:1.5k:1.5k	3	224 <sup>2</sup>
111	Chest X-Ray8 [35, 36]	7.2k:7k:7k:4.1k :3.9k:3.5k:2.9k	7	1024 <sup>2</sup>

Table 1. Shortlisted Datasets.

118 perform dermatologist level skin-cancer detection using skin lesion images [] using transfer learning.

119 In the recent years, many researchers have tried to predict lung diseases using deep CNNs, Wang *et al.* [35] used state of the art backbone architectures to train a lung disease classifier for multilabel data by training only the prediction and transition layers from scratch and leaving pre-trained ImageNet weights freezed while training. They achieved a high AUC of over 0.6 for most of the classes in the dataset with this technique. Rajpurkar *et al.* [] created a 121 layer deep CNN - CheXNet to detect pneumonia using chest X-rays with radiologist level accuracy. Labhane *et al.* [] used transfer learning with state of the art backbone architectures like VGG16, VGG19 and InceptionV3 to predict pneumonia in pediatric patients and achieved an F1 score of 0.97. Islam *et al.* combined CNN and LSTM to create a COVID-19 detector []. The CNN extracted complex features from scans and the LSTM was used as a classifier. This method resulted in an improvement over a vanilla CNN network and an F1 score of 98.9% was achieved. Abbas *et al.* [] created the De-TraC network to detect COVID in chest X-rays that improved performance of existing backbone models significantly with the highest accuracy of 98.23% using the VGG19 architecture. Guefrechi *et al.* [] on the other hand used data augmentation techniques like random rotation, flipping and noise with transfer learning on backbone architectures like ResNet50, InceptionV3 and VGG16 to achieve a high accuracy of 98.3%.

149 In the following sections methodology of the approach and the results will be discussed.

## 2. Methodology-2

153 **Datasets:** (Tab. 1) with varying disease types were chosen to ensure model robustness. Other criteria included the *number of images per class* and *image quality* as noisy scans can lead to mis-diagnosis [28].

157 The **COVID** dataset was created by a team of researchers from Qatar University, Doha, Qatar, and the University of Dhaka, Bangladesh along with collaborators from Pakistan and Malaysia in collaboration with medical doctors from the Italian Society of Medical and

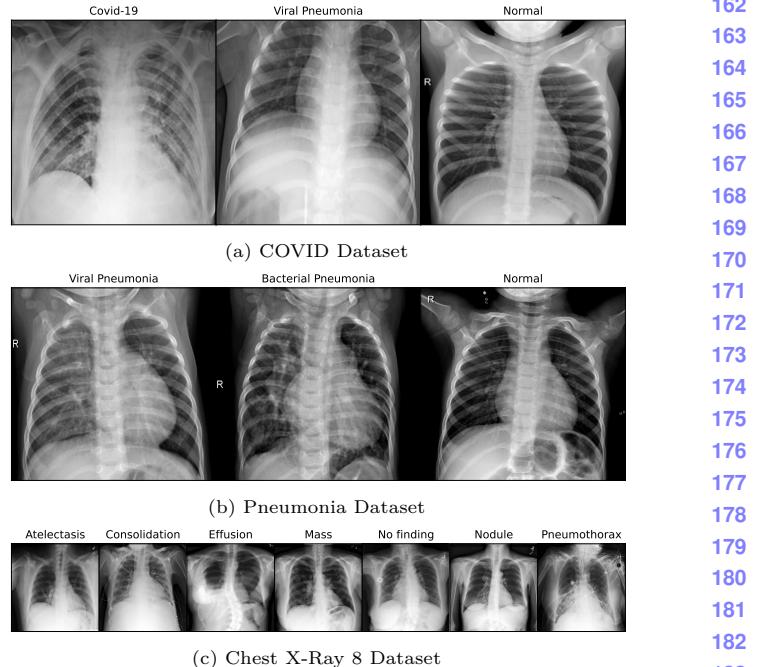


Figure 1. Sample Chest X-rays from the datasets used.

184 Interventional Radiology database using 43 [8] different publications [5, 6, 25]. It is a multiclass data with three classes, COVID, viral pneumonia and normal. X-rays with widespread, hazy, and irregular ground glass opacities are of the COVID-19 class [16]. Whereas, the ones with haziness only in the lower regions [39] are viral pneumonia cases as shown in Fig. 1. Chest X-rays of normal lungs provide a clear view of the lungs.

185 The **Pneumonia**, dataset contains scans from pediatric patients of one to five year olds collected as part of patients' routine clinical care at the Guangzhou Women and Children's Medical Center, Guangzhou, China. [17, 33] This dataset is multiclass with three classes, viral pneumonia, bacterial pneumonia and normal. Scans with one white condensed area affecting only one side of the lungs are tagged as bacterial pneumonia [2] as bacteria tends to aggressively attack one part of the lungs causing inflammation to replace the cells that were otherwise filled with air. On the other hand, X-rays which show bilateral patchy areas of consolidation are classified as viral pneumonia [13] as viruses attack both sides of the lungs producing a homogeneous inflammatory reaction causing mucus and cellular debris. Normal scans here as well produce a clear view of the lungs.

186 NIH [37] released over 100k anonymized chest X-ray images along with their radiological reports from over 30k patients. Wang *et al.* [35] used this data to create the **Chest X-ray 8** dataset by generating

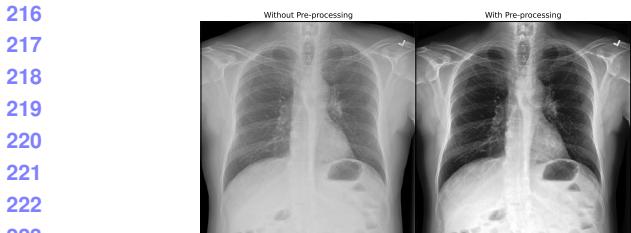


Figure 2. Effect of pre-processing on Chest X-ray images.

disease labels through NLP from the radiological reports. [36] The dataset contains 15 classes but only 7 Fig. 1 were chosen for this study. This dataset is significantly different from the other two as it is a multi-label dataset. Classes were iteratively removed, ensuring that they are not highly imbalanced to finally reach the 7 classes. With over 29,000 images of size 1024 x 1024, this dataset was the biggest and thus had to be resized down to 384 x 384 to reduce training and processing times. Furthermore, normal class images were undersampled by first choosing one scan per patient and then selecting 7000 scans out of this subset randomly. The data consists of multiple scans from the same subject which could lead to data leakage between the train, val and test sets if a random train-test-val split was performed. This was prevented with the use of GroupShuffleSplit from the scikit library.

Before training, all the images were pre-processed using histogram equalization and Gaussian blur with a 5x5 filter as Giełczyk *et al.* [10] showed that this improved the F1 score by about 4% for the chest X-ray classification task. Visually, the contrast of the scan improved and allowed irregularities to stand out as shown in Fig. 2. Next, the scans were divided into train, validation and test with the 70:15:15 split. During training, the scans were augmented using RandomAdjustSharpness and RandomAutocontrast in Pytorch [24] to increase the number of images the model gets to learn from and ensure that the model is robust to scans from different machines. RandomHorizontalFlip was also used to make the models invariant to the direction of the scan as some scans were anterior-posterior while others were posterior-anterior.

**Backbone Architectures:** (Tab. 2) of various configuration and blocks were chosen to ensure that different ideas are tested in this study. Other selection criteria were the *number of trainable parameters*, important to keep track of the total training time, *FLOPS* as we wanted models that could easily be deployed on to embedded devices and the *top 5 classification accuracy* on the ImageNet 1K benchmark dataset.

**ResNet 34** residual learning network with 34 lay-

Arch.	Params (Mil.)	Layers	FLOPS (Bil.)	Imagenet Acc.	
MobileNet	5.5	18	8.7	92.6	270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323
EfficientNet	7.8	25	25.8	94.9	
Resnet	21.8	34	153.9	91.4	

Table 2. Shortlisted Backbone Architectures.

ers that are made possible by skip connections. The 34 layer variant was chosen to decrease training time while not compromising on the accuracy much. This architecture had the highest trainable parameters and FLOPS while the lowest Imagenet accuracy. [14]

**MobileNet V3 Large** uses depthwise separable convolution from MobileNet V2 [27] along with squeeze-excitation blocks in residual layers from MnasNet [31]. This makes it really quick to train while still performing at par with other architectures. This architecture had the lowest trainable parameters and FLOPS among the three selected. Howard *et al.* [15] also used network architecture search to find the most effective model. The large configuration was chosen to not compromise on the prediction accuracy.

**EfficientNet B1** uses compound scaling to scale the model by depth, width and resolution. The B1 version was chosen to have faster training without compromising on the accuracy. [32] This architecture performs the best among the selected on the Imagenet benchmark dataset while having a third of the trainable parameters of Resnet34.

**Optimization Algorithm:** The Adam optimizer [18] is an adaptive learning rate algorithm which was chosen as the algorithm of choice as it converges faster by integrating benefits of RMSProp and momentum. It is also robust to hyperparameters but, requires tweaking of the learning rate depending on the task at hand. For this study, we used a learning rate of 0.01 and the author recommend settings for  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$  and  $\epsilon = 10^{-8}$  for the first and second order moment estimate as defined in Eq. (1) and Eq. (2) where  $\beta_1$  and  $\beta_2$  control the decay rates.

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \quad (1)$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \quad (2)$$

We further used the Cosine annealing [19] learning rate scheduler to reduce the learning rate as the training progressed down to a low of 0.001.

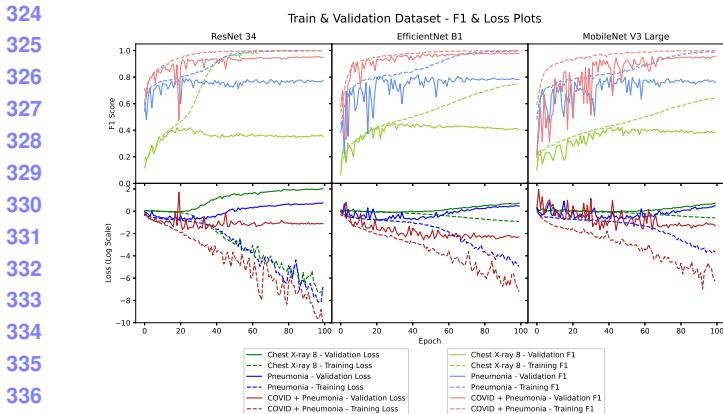


Figure 3. Train &amp; Val F1 &amp; Loss plots for the 9 models.

### 3. Results-2.5

**Experiment Setup:** Two datasets in this study presented the multiclass classification problem while the third, chest X-ray 8 dataset presented the multiclass, multilabel classification problem. Thus, the training methodology was separated for these two problems. For the multiclass problem, the softmax

Two datasets in this study had a very small number of samples which caused the models to overfit early. To mitigate this, random contrast and sharpness adjustment [23] data augmentation techniques were used. Some scans in the datasets were anterior-posterior while some others were posterior-anterior and using the horizontal flip data augmentation would make the model invariant to these differences [4]. Inception was the first model trained and each epoch took over 1 hour. To reduce the training time, the X-ray images were resized, pre-processed and split into train, test and validation sets separately. Furthermore, EfficientNet, MobileNet and ResNet 34 were chosen as they have a considerably low number of learnable parameters. Now each epoch is taking less than 4 minutes.

Nine models were trained from scratch and the training, validation F1 score and loss can be seen in Fig. 3. From the plots it is clear that going from a smaller architecture to a bigger architecture, makes the model start to overfit earlier. Another interesting observation is that cosine annealing impacted the loss of MobileNet the most every 10 epochs due to warm restarts. From the graphs it can be seen that all three datasets had similar performance across models when trained for a high number of epochs. The X-ray 8 dataset performed the worst among the three datasets which could be due to the high number of classes as compared to the other datasets. Surprisingly, the pneumonia dataset performed worse than the COVID +

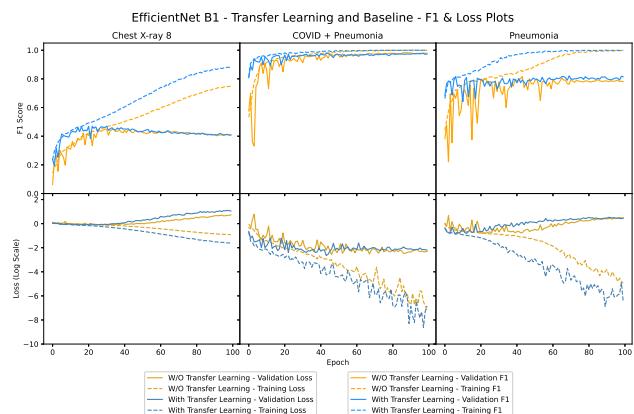


Figure 4. Train &amp; Val, F1 &amp; Loss plots for EfficientNet B1 trained from scratch and with ImageNet weights.

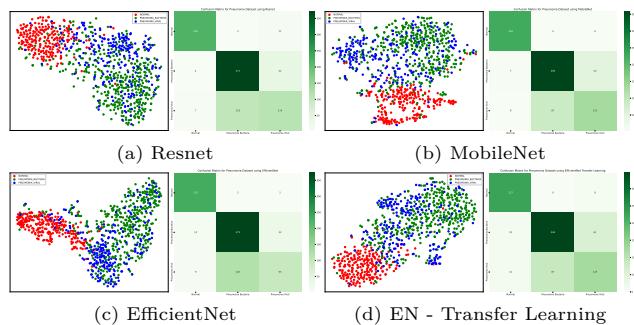


Figure 5. T-SNE and Confusion matrices of the Pneumonia dataset.

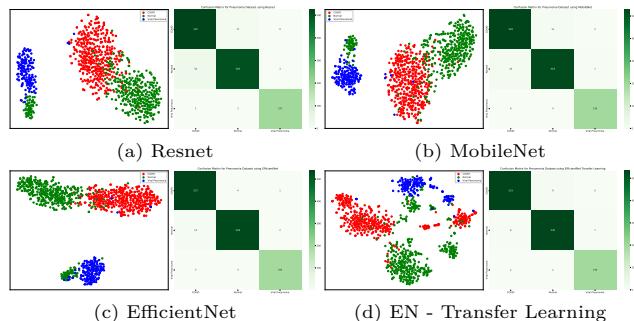


Figure 6. T-SNE and Confusion matrices of the COVID dataset.

pneumonia dataset which indicates that COVID cases are easier to distinguish from pneumonia cases.

#### Main Results:

#### Ablative Study:

Model	ResNet			MobileNet			EfficientNet			EN - Transfer Learning		
Dataset	F1	Time	Epoch	F1	Time	Epoch	F1	Time	Epoch	F1	Time	Epoch
Pneumonia	0.784	82	<b>22</b>	<b>0.804</b>	<b>75</b>	42	0.768	110	44	0.782	114	70
COVID	0.963	68	71	0.959	<b>45</b>	82	0.970	80	89	<b>0.978</b>	56	<b>46</b>
X-Ray 8	0.411	11,502	<b>19</b>	0.406	<b>7,275</b>	42	0.445	13,820	31	<b>0.457</b>	13,813	29

Table 3. F1 (higher is better), time per epoch in seconds (lower is better), and number of epochs to reach the best validation loss (lower is better) for the 12 models that were trained.

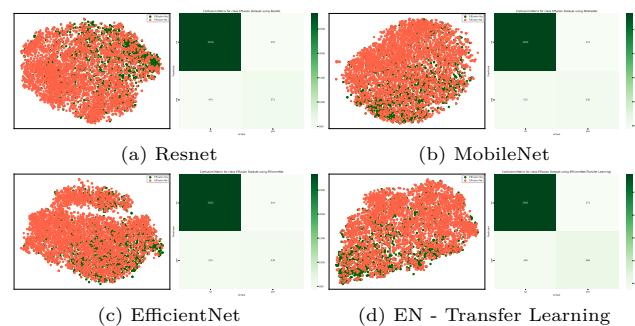


Figure 7. T-SNE and Confusion matrices of the Chest X-ray 8 dataset.

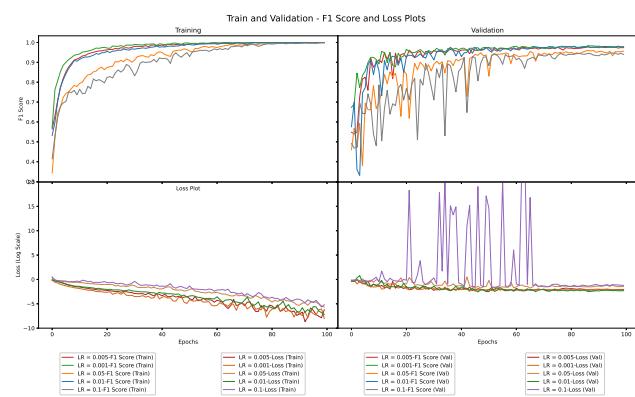


Figure 8. Train & Val, F1 & Loss plots for ablative study models.

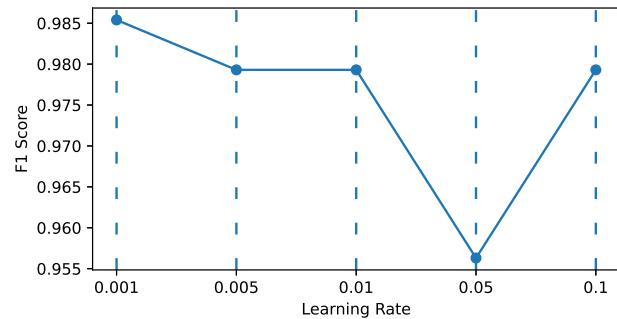


Figure 9. Ablative Study F1 scores (Higher is better).

540

## 541 References

- 542 [1] Noorullah Akhtar, Jiyuan Ni, Claire Langston, Gail J  
543 Demmler, and Jeffrey A Towbin. Pcr diagnosis of viral  
544 pneumonitis from fixed-lung tissue in children. *Bio-  
545 chemical and molecular medicine*, 58(1):66–76, 1996.  
546 1
- 547 [2] How Drugs are Made and Product List. Viral  
548 vs. bacterial pneumonia: Understanding the differ-  
549 ence. [https://www.pfizer.com/news/articles/viral\\_vs\\_bacterial\\_pneumonia\\_understanding\\_the\\_difference](https://www.pfizer.com/news/articles/viral_vs_bacterial_pneumonia_understanding_the_difference), 2020. 2
- 550 [3] Paul J. Blazek. Why we will never open deep learning’s  
551 black box. <https://towardsdatascience.com/why-we-will-never-open-deep-learnings-black-box-4c27cd335118>, 2022. 1
- 552 [4] Aleksander Botev, Matthias Bauer, and Soham  
553 De. Regularising for invariance to data augmentation  
554 improves supervised learning. *arXiv preprint arXiv:2203.03304*, 2022. 4
- 555 [5] Muhammad E. H. Chowdhury, Tawsifur Rahman,  
556 Amith Khandakar, Rashid Mazhar, Muhammad Abd-  
557 ul Kadir, Zaid Bin Mahbub, Khandakar Reajul  
558 Islam, Muhammad Salman Khan, Atif Iqbal,  
559 Nasser Al Emadi, Mamun Bin Ibne Reaz, and Moham-  
560 mad Tariqul Islam. Can ai help in screening viral and  
561 covid-19 pneumonia? *IEEE Access*, 8:132665–132676,  
562 2020. 2
- 563 [6] Muhammad E. H. Chowdhury, Tawsifur Rahman,  
564 Amith Khandakar, Rashid Mazhar, Muhammad Abd-  
565 ul Kadir, Zaid Bin Mahbub, Khandakar Reajul  
566 Islam, Muhammad Salman Khan, Atif Iqbal,  
567 Nasser Al Emadi, Mamun Bin Ibne Reaz, and Moham-  
568 mad Tariqul Islam. Covid-19 radiography  
569 database. <https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database>,  
570 2021. 2
- 571 [7] Priya Daniel, Chamira Rodrigo, Tricia M Mckeever,  
572 Mark Woodhead, Sally Welham, and Wei Shen Lim.  
573 Time to first antibiotic and mortality in adults hos-  
574 pitalised with community-acquired pneumonia: a  
575 matched-propensity analysis. *Thorax*, 71(6):568–570,  
576 2016. 1
- 577 [8] Società Italiana di Radiologia. Covid pneumonia  
578 dataset. <https://sirm.org/category/senza-categoria/covid-19/>, 2020. 2
- 579 [9] Guy Frija, Ivana Blažić, Donald P Frush, Monika  
580 Hierath, Michael Kawooya, Lluis Donoso-Bach, and  
581 Boris Brkljačić. How to improve access to medical  
582 imaging in low-and middle-income countries? *EClinicalMedicine*, 38:101034, 2021. 1
- 583 [10] Agata Gielczyk, Anna Marciniak, Martyna Tar-  
584 czewska, and Zbigniew Lutowski. Pre-processing  
585 methods in chest x-ray image classification. *Plos one*,  
586 17(4):e0265949, 2022. 3
- 587 [11] Jacob Gildenblat and contributors. Pytorch library  
588 for cam methods. <https://github.com/jacobgil/pytorch-grad-cam>, 2021. 1

- 589 [12] Sarra Guefrechi, Marwa Ben Jabra, Adel Ammar, Anis  
590 Koubaa, and Habib Hamam. Deep learning based de-  
591tection of covid-19 from chest x-ray images. *Multime-  
592 dia Tools and Applications*, 80(21):31803–31820, 2021.  
593 1
- 594 [13] W Guo, J Wang, M Sheng, M Zhou, and L Fang. Ra-  
595 diological findings in 210 paediatric patients with viral  
596 pneumonia: a retrospective case study. *The British  
597 journal of radiology*, 85(1018):1385–1389, 2012. 2
- 598 [14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian  
599 Sun. Deep residual learning for image recognition. In  
600 *Proceedings of the IEEE conference on computer vision  
601 and pattern recognition*, pages 770–778, 2016. 3
- 601 [15] Andrew Howard, Mark Sandler, Grace Chu, Liang-  
602 Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang,  
603 Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al.  
604 Searching for mobilenetv3. In *Proceedings of the  
605 IEEE/CVF international conference on computer vi-  
606 sion*, pages 1314–1324, 2019. 3
- 607 [16] Adam Jacobi, Michael Chung, Adam Bernheim, and  
608 Corey Eber. Portable chest x-ray in coronavirus  
609 disease-19 (covid-19): A pictorial review. *Clinical  
610 imaging*, 64:35–42, 2020. 2
- 611 [17] Daniel Kermany, Kang Zhang, Michael Goldbaum,  
612 et al. Labeled optical coherence tomography (oct) and  
613 chest x-ray images for classification. *Mendeley data*,  
614 2(2), 2018. 2
- 615 [18] Diederik P Kingma and Jimmy Ba. Adam: A  
616 method for stochastic optimization. *arXiv preprint  
617 arXiv:1412.6980*, 2014. 1, 3
- 618 [19] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic  
619 gradient descent with warm restarts. *arXiv preprint  
620 arXiv:1608.03983*, 2016. 1, 3
- 621 [20] P Mehrotra, V Bosemani, and J Cox. Do radiologists  
622 still need to report chest x rays? *Postgraduate medical  
623 journal*, 85(1005):339–341, 2009. 1
- 624 [21] Aleksandra Mojsilovic. Introducing ai explainability  
625 360. <https://www.ibm.com/blogs/research/2019/08/ai-explainability-360/>, 2019. 1
- 626 [22] Andrew Murphy. Radiographic contrast. <https://radiopaedia.org/articles/radiographic-contrast>, 2022. 1
- 627 [23] Loris Nanni, Michelangelo Paci, Sheryl Brahnam, and  
628 Alessandra Lumini. Comparison of different image  
629 data augmentation approaches. *Journal of Imaging*,  
630 7(12):254, 2021. 4
- 631 [24] PyTorch. Transforming and augmenting images.  
632 <https://pytorch.org/vision/stable/transforms.html>. 3
- 633 [25] Tawsifur Rahman, Amith Khandakar, Yazan Qi-  
634 blawey, Anas Tahir, Serkan Kiranyaz, Saad Bin  
635 Abul Kashem, Mohammad Tariqul Islam, Somaya  
636 Al Maadeed, Susu M. Zughaiier, Muhammad Salman  
637 Khan, and Muhammad E.H. Chowdhury. Exploring  
638 the effect of image enhancement techniques on covid-  
639 19 detection using chest x-ray images. *Computers in  
640 Biology and Medicine*, 132:104319, 2021. 2

- 648 [26] Melissa Rohman. Ai performs poorly when 702  
649 tested on data from multiple health systems. 703  
650 [https://healthimaging.com/topics/artificial- 706  
653 intelligence / ai - poorly - detects - pneumonia - 707  
654 chest-x-rays](https://healthimaging.com/topics/artificial- 704<br/>651 intelligence / ai - poorly - detects - pneumonia - 705<br/>652 chest-x-rays), 2018. 1
- 655 [27] Mark Sandler, Andrew Howard, Menglong Zhu, 708  
656 Andrei Zhmoginov, and Liang-Chieh Chen. MobileNetv2: 709  
657 Inverted residuals and linear bottlenecks. In *Proceedings 710  
658 of the IEEE conference on computer vision and 711  
659 pattern recognition*, pages 4510–4520, 2018. 3
- 660 [28] Janaki Sivakumar, K Thangavel, and P Saravanan. 712  
661 Computed radiography skull image enhancement 713  
662 using wiener filter. In *International Conference on 714  
663 Pattern Recognition, Informatics and Medical Engineering 715  
664 (PRIME-2012)*, pages 307–311. IEEE, 2012. 2
- 665 [29] Matt Smith. Common lung diagnostic tests. [https://www.webmd.com/lung/breathing-diagnostic-tests](https:// 716<br/>666 www.webmd.com/lung/breathing-diagnostic-tests), 717  
667 2022. 1
- 668 [30] Healthwise Staff. Chest x-ray. [https://www.healthlinkbc.ca/tests-treatments-medications/ 720  
671 medical-tests/chest-x-ray](https://www. 718<br/>669 healthlinkbc.ca/tests-treatments-medications/ 719<br/>670 medical-tests/chest-x-ray), 2021. 1
- 672 [31] Mingxing Tan, Bo Chen, Ruoming Pang, Vijay 721  
673 Vasudevan, Mark Sandler, Andrew Howard, and Quoc V 722  
674 Le. Mnasnet: Platform-aware neural architecture 723  
675 search for mobile. In *Proceedings of the IEEE/CVF 724  
676 Conference on Computer Vision and Pattern Recognition*, 725  
677 pages 2820–2828, 2019. 3
- 678 [32] Mingxing Tan and Quoc Le. Efficientnet: Rethinking 726  
679 model scaling for convolutional neural networks. In *International 727  
680 conference on machine learning*, pages 6105–6114. PMLR, 2019. 3
- 681 [33] Tolga. Chest x-ray images. [https://www.kaggle.com/datasets/tolgadincer/labeled-chest-xray- 730  
684 images](https://www.kaggle. 728<br/>682 com/datasets/tolgadincer/labeled-chest-xray- 729<br/>683 images), 2020. 2
- 685 [34] Guangyu Wang, Xiaohong Liu, Jun Shen, Chengdi 731  
686 Wang, Zhihuan Li, Linsen Ye, Xingwang Wu, Ting 732  
687 Chen, Kai Wang, Xuan Zhang, et al. A deep-learning 733  
688 pipeline for the diagnosis and discrimination of viral, 734  
689 non-viral and covid-19 pneumonia from chest x-ray 735  
690 images. *Nature biomedical engineering*, 5(6):509–521, 736  
691 2021. 1
- 692 [35] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, 737  
693 Mohammadhdadi Bagheri, and Ronald M Summers. 738  
694 Chestx-ray8: Hospital-scale chest x-ray database and 739  
695 benchmarks on weakly-supervised classification and 740  
696 localization of common thorax diseases. In *Proceedings 741  
697 of the IEEE conference on computer vision and pattern 742  
698 recognition*, pages 2097–2106, 2017. 2
- 699 [36] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, 743  
700 Mohammadhdadi Bagheri, and Ronald M Summers. Nih 744  
701 chest x-rays. <https://www.kaggle.com/datasets/nih-chest-xrays/data>, 2017. 2, 3
- 702 [37] Wang X, Peng Y, Lu L, Lu Z, Bagheri M, and Summers 751  
703 RM. Nih clinical center provides one of the largest 752  
704 publicly available chest x-ray datasets to scientific 753  
705 community. <https://www.nih.gov/news-events/news-releases/nih-clinical-center-provides-one-of-the-largest-publicly-available-chest-x-ray-datasets-to-scientific-community>
- 706 [one-largest-publicly-available-chest-x-ray-datasets-scientific-community](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9500000/), 2017. 2
- 707 [38] Lizhou Xu, Danyang Li, Sami Ramadan, Yanbin Li, 708  
708 and Norbert Klein. Facile biosensors for rapid 709  
709 detection of covid-19. *Biosensors and Bioelectronics*, 710  
710 170:112673, 2020. 1
- 711 [39] Na Zhan, Yingyun Guo, Shan Tian, Binglu Huang, 712  
712 Xiaoli Tian, Jinjing Zou, Qiutang Xiong, Dongling Tang, 713  
713 Liang Zhang, and Weiguo Dong. Clinical characteristics 714  
714 of covid-19 complicated with pleural effusion. 715  
715 *BMC Infectious Diseases*, 21(1):1–10, 2021. 2