

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 [39]. These diseases can be diagnosed using a variety of tests like pulse oximetry, chest x-ray, CT scan [30], 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 [31]. 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 [35] but haven't been able to come up with models that can replace radiologists. Small [12] and highly imbalanced data [35], along with varying specifications of X-ray scanners leading to low inter-hospital accuracy [27] 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, 26]	10000k:3.6k:1.3k	3	299 ²
110	Pneumonia [17, 34]	3k:1.5k:1.5k	3	224 ²
111	Chest X-Ray8 [36, 37]	7.2k:7k:7k:4.1k :3.9k:3.5k:2.9k	7	1024 ²

Table 1. Shortlisted Datasets.

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

117 In the recent years, many researchers have tried to predict lung diseases using deep CNNs, Wang *et al.* [36] 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 [29].

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

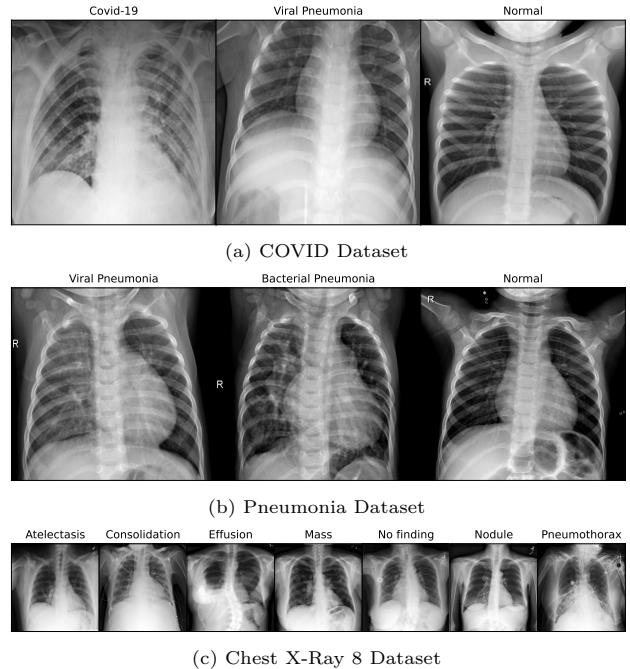


Figure 1. Sample Chest X-rays from the datasets used.

184 Interventional Radiology database using 43 [8] different publications [5, 6, 26]. 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 [40] are viral pneumonia cases as shown in Fig. 1. Chest X-rays of normal lungs provide a clear view of the lungs. The normal class was undersampled to use only ***3.6k*** scans and reduce the data imbalance.

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, 34] 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 [38] released over 100k anonymized chest X-ray images along with their radiological reports from

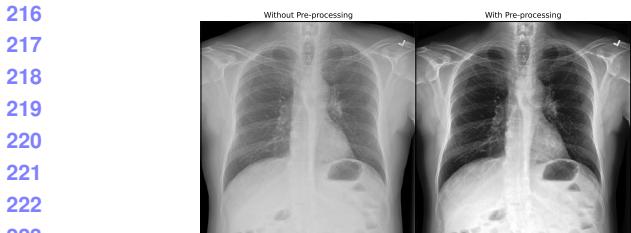


Figure 2. Effect of pre-processing on Chest X-ray images.

over 30k patients. Wang *et al.* [36] used this data to create the **Chest X-ray 8** dataset by generating disease labels through NLP from the radiological reports. [37] 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 Gielczyk *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 [23] 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 [4].

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*.

Arch.	Params (Mil.)	Layers	FLOPS (Bil.)	Imagenet Acc.	
MobileNet	5.5	18	8.7	92.6	270 271 272 273 274 275 276 277
EfficientNet	7.8	25	25.8	94.9	278 279 280 281
Resnet	21.8	34	153.9	91.4	282 283 284 285 286 287

Table 2. Shortlisted Backbone Architectures.

racy on the ImageNet 1K benchmark dataset.

ResNet 34 residual learning network with 34 layers 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 [28] along with squeeze-excitation blocks in residual layers from MnasNet [32]. 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. [33] 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 β_1 and β_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.

324

3. Results-2.5

325

Experiment Setup:

First we performed undersampling as described in Sec. 2 on the datasets. Then, the scans were pre-processed using histogram equalization and Gaussian blur before resizing them and storing them in separate directories to make it easier for PyTorch dataloaders. 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 multilabel problem, a softmax layer had to be added before the loss function to get 0 or 1 prediction for all the classes of the data. For this, the BCEWITHLOGITSLOSS function of PyTorch was used as it combines the Sigmoid layer and the BCELoss function in one single class. This makes these operations more numerically stable than their separate counterparts [25]. The backbone architectures were obtained directly from the torchvision library and the final classification layer was modified for the selected datasets. For the models which had to be trained from scratch, the weights were randomly initialized and the entire model was trained for a total of 100 epochs each. For the transfer learning models, the weights were initialized with the IMAGENET1K_V2 weights but the entire model was fine-tuned. The rationale behind performing deep-tuning was that the Imagenet data is very different from chest X-ray scans thus the model would need to learn features from X-ray scans.

While training the best model by validation loss was saved to prevent the usage of overfit models for test set analysis. The actual and predicted results from each epoch was also stored to calculate the F1 score at each step of training. While calculating the F1 score, macro averaging was used to get an average score across classes. All images were normalized before getting trained with the mean and standard deviation of the training set of each of the selected datasets.

Initial runs of the multilabel data training produced 0 F1 score due to the highly imbalanced nature of multilabel classes. To mitigate this, class wise weights were calculated using the training data and used with the loss function. This improved the F1 score considerably.

Finally, the best models from each run by validation loss were used to get the test set metrics that are displayed in Tab. 3. Training and validation F1 score and loss are also provided in Fig. 3 and Fig. 4.

Main Results:

Nine models were trained from scratch and the training, validation F1 score and loss can be seen in Fig. 3.

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

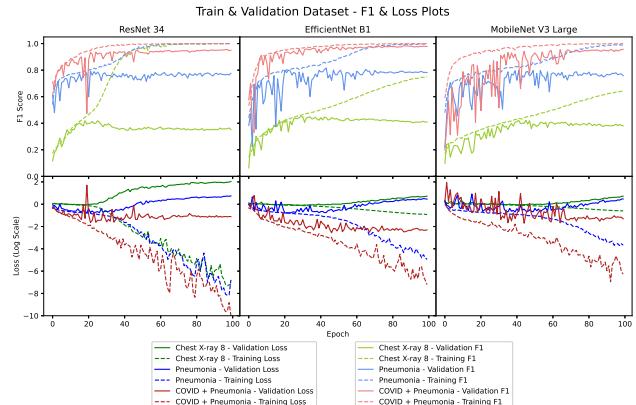


Figure 3. Train & Val F1 & Loss plots for the 9 models.

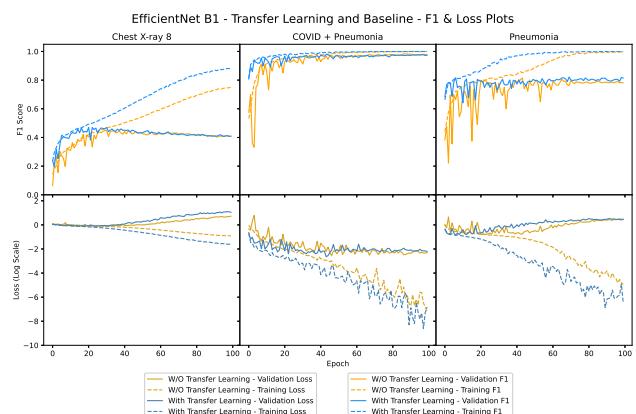


Figure 4. Train & Val, F1 & Loss plots for EfficientNet trained from scratch and with ImageNet weights.

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

Ablative Study: For the ablative study, the COVID dataset was chosen along with the EfficientNet B1 architecture trained from scratch. The learning rates chosen for the study are 0.001, 0.005, 0.01, 0.05, and 0.1. From the training and validation F1 score and loss plots given in Fig. 10 it is seen that a very high

Model	ResNet			MobileNet			EfficientNet			EN - Transfer Learning		
Dataset	F1	Time	Epoch	F1	Time	Epoch	F1	Time	Epoch	F1	Time	Epoch
Pneumonia	0.784	82	22	0.804	75	42	0.768	110	44	0.782	114	70
COVID	0.967	68	71	0.959	45	82	0.979	80	89	0.978	56	46
X-Ray 8	0.411	11,502	19	0.406	7,275	42	0.445	13,820	31	0.457	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.

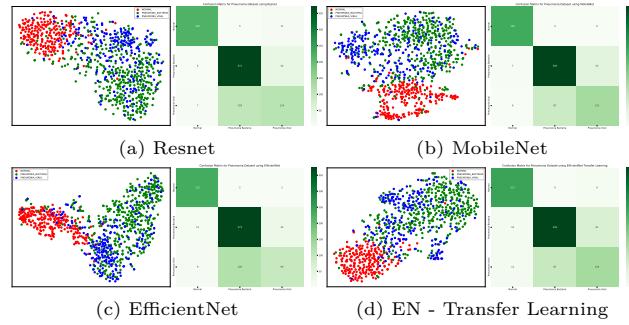


Figure 5. T-SNE and Confusion matrices of the Pneumonia dataset.

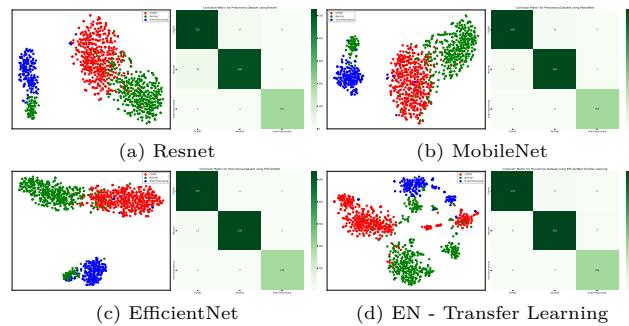


Figure 6. T-SNE and Confusion matrices of the COVID dataset.

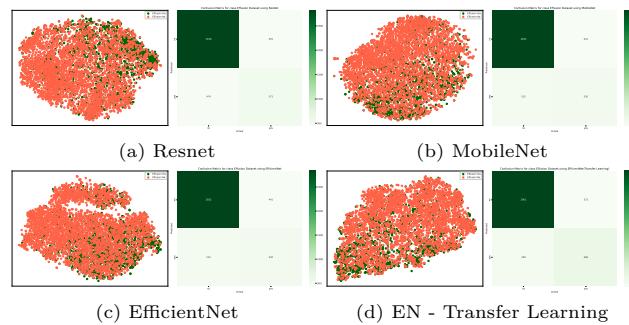


Figure 7. T-SNE and Confusion matrices of the Chest X-ray 8 dataset.

learning rate of 0.1 is highly unstable and prevents the model from reaching close to global minima. Similarly,

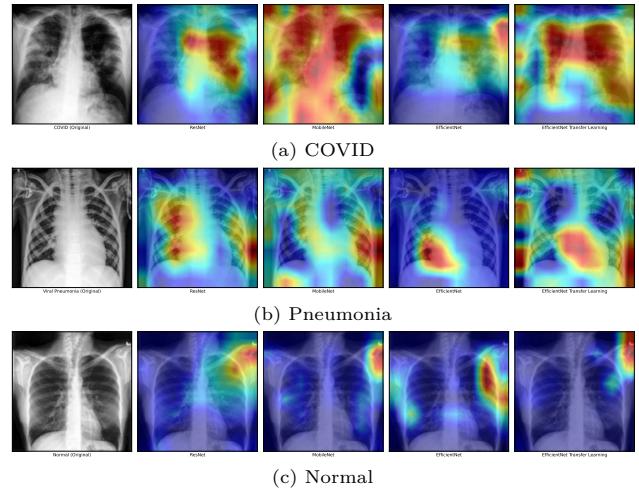


Figure 8. GradCAM visualization for the COVID dataset.

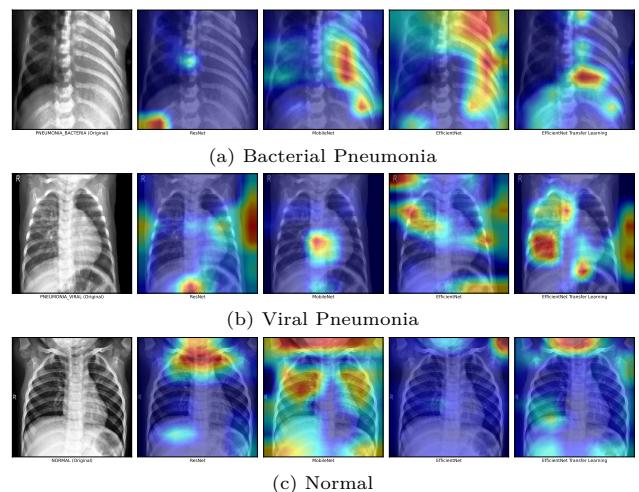


Figure 9. GradCAM visualization for the Pneumonia dataset.

learning rate of 0.05 also prevented the model from converging on the validation set even after 100 epochs. The other three learning rates all converged on the validation set but, the learning rate of 0.01 was the most stable and reached the highest F1 score earliest. On the other hand, learning rate of 0.001 performed bet-

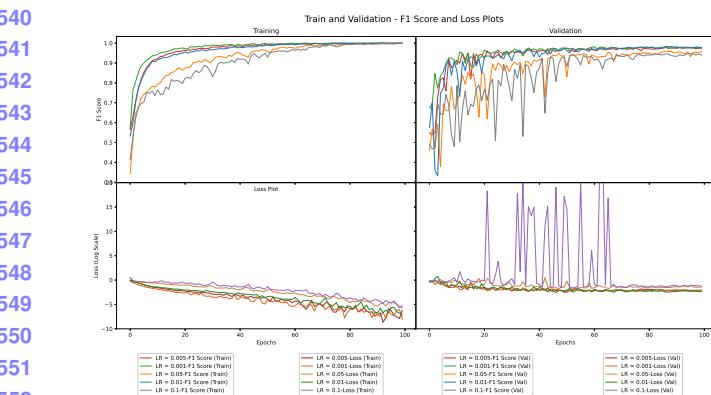


Figure 10. Train & Val, F1 & Loss plots for ablative study models.

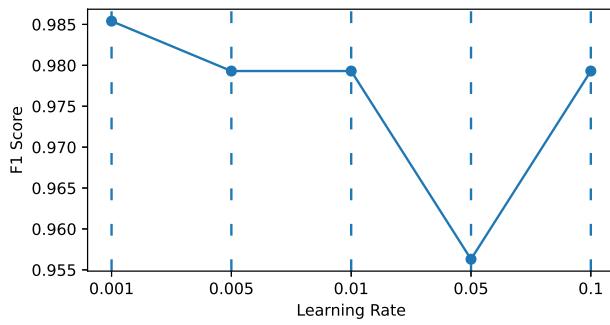


Figure 11. Ablative Study F1 scores (Higher is better).

ter on the loss plot. From Fig. 11 we can see that the best performing learning rate is 0.001 on the F1 score of the test set with 0.005, 0.01 and 0.1 close second** check**.

648

649
References

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

- [1] Noorullah Akhtar, Jiyuan Ni, Claire Langston, Gail J Demmler, and Jeffrey A Towbin. Pcr diagnosis of viral pneumonitis from fixed-lung tissue in children. *Biochemical and molecular medicine*, 58(1):66–76, 1996.
- [2] How Drugs are Made and Product List. Viral vs. bacterial pneumonia: Understanding the difference. https://www.pfizer.com/news/articles/viral_vs_bacterial_pneumonia_understanding_the_difference, 2020.
- [3] Paul J. Blazek. Why we will never open deep learning’s black box. <https://towardsdatascience.com/why-we-will-never-open-deep-learnings-black-box-4c27cd335118>, 2022.
- [4] Aleksander Botev, Matthias Bauer, and Soham De. Regularising for invariance to data augmentation improves supervised learning. *arXiv preprint arXiv:2203.03304*, 2022.
- [5] Muhammad E. H. Chowdhury, Tawsifur Rahman, Amith Khandakar, Rashid Mazhar, Muhammad Abdul Kadir, Zaid Bin Mahbub, Khandakar Reajul Islam, Muhammad Salman Khan, Atif Iqbal, Nasser Al Emadi, Mamun Bin Ibne Reaz, and Mohammad Tariqul Islam. Can ai help in screening viral and covid-19 pneumonia? *IEEE Access*, 8:132665–132676, 2020.
- [6] Muhammad E. H. Chowdhury, Tawsifur Rahman, Amith Khandakar, Rashid Mazhar, Muhammad Abdul Kadir, Zaid Bin Mahbub, Khandakar Reajul Islam, Muhammad Salman Khan, Atif Iqbal, Nasser Al Emadi, Mamun Bin Ibne Reaz, and Mohammad Tariqul Islam. Covid-19 radiography database. <https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database>, 2021.
- [7] Priya Daniel, Chamira Rodrigo, Tricia M Mckeever, Mark Woodhead, Sally Welham, and Wei Shen Lim. Time to first antibiotic and mortality in adults hospitalised with community-acquired pneumonia: a matched-propensity analysis. *Thorax*, 71(6):568–570, 2016.
- [8] Società Italiana di Radiologia. Covid pneumonia dataset. <https://sirm.org/category/senza-categoria/covid-19/>, 2020.
- [9] Guy Frija, Ivana Blažić, Donald P Frush, Monika Hierath, Michael Kawooya, Lluis Donoso-Bach, and Boris Brklačić. How to improve access to medical imaging in low-and middle-income countries? *EClinicalMedicine*, 38:101034, 2021.
- [10] Agata Gielczyk, Anna Marciniak, Martyna Tarczewska, and Zbigniew Lutowski. Pre-processing methods in chest x-ray image classification. *Plos one*, 17(4):e0265949, 2022.
- [11] Jacob Gildenblat and contributors. Pytorch library for cam methods. <https://github.com/jacobgil/pytorchGradCam>, 2021.

- [12] Sarra Guefrechi, Marwa Ben Jabra, Adel Ammar, Anis Koubaa, and Habib Hamam. Deep learning based detection of covid-19 from chest x-ray images. *Multimedia Tools and Applications*, 80(21):31803–31820, 2021.
- [13] W Guo, J Wang, M Sheng, M Zhou, and L Fang. Radiological findings in 210 paediatric patients with viral pneumonia: a retrospective case study. *The British journal of radiology*, 85(1018):1385–1389, 2012.
- [14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [15] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1314–1324, 2019.
- [16] Adam Jacobi, Michael Chung, Adam Bernheim, and Corey Eber. Portable chest x-ray in coronavirus disease-19 (covid-19): A pictorial review. *Clinical imaging*, 64:35–42, 2020.
- [17] Daniel Kermany, Kang Zhang, Michael Goldbaum, et al. Labeled optical coherence tomography (oct) and chest x-ray images for classification. *Mendeley data*, 2(2), 2018.
- [18] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [19] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.
- [20] P Mehrotra, V Bosemani, and J Cox. Do radiologists still need to report chest x rays? *Postgraduate medical journal*, 85(1005):339–341, 2009.
- [21] Aleksandra Mojsilovic. Introducing ai explainability 360. <https://www.ibm.com/blogs/research/2019/08/ai-explainability-360/>, 2019.
- [22] Andrew Murphy. Radiographic contrast. <https://radiopaedia.org/articles/radiographic-contrast>, 2022.
- [23] Loris Nanni, Michelangelo Paci, Sheryl Brahnam, and Alessandra Lumini. Comparison of different image data augmentation approaches. *Journal of Imaging*, 7(12):254, 2021.
- [24] PyTorch. Transforming and augmenting images. <https://pytorch.org/vision/stable/transforms.html>.
- [25] PyTorch. Bce with logits loss. <https://pytorch.org/docs/stable/generated/torch.nn.BCEWithLogitsLoss.html>, 2022.
- [26] Tawsifur Rahman, Amith Khandakar, Yazan Qiblawey, Anas Tahir, Serkan Kiranyaz, Saad Bin Abul Kashem, Mohammad Tariqul Islam, Somaya Al Maadeed, Susu M. Zughaier, Muhammad Salman Khan, and Muhammad E.H. Chowdhury. Exploring

- 756 the effect of image enhancement techniques on covid-
757 19 detection using chest x-ray images. *Computers in*
758 *Biology and Medicine*, 132:104319, 2021. 2
- 759 [27] Melissa Rohman. Ai performs poorly when
760 tested on data from multiple health systems.
<https://healthimaging.com/topics/artificial-intelligence/ai-poorly-detects-pneumonia-chest-x-rays>, 2018. 1
- 761 [28] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2:
762 Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and*
763 *pattern recognition*, pages 4510–4520, 2018. 3
- 764 [29] Janaki Sivakumar, K Thangavel, and P Saravanan.
765 Computed radiography skull image enhancement using wiener filter. In *International Conference on Pattern*
766 *Recognition, Informatics and Medical Engineering (PRIME-2012)*, pages 307–311. IEEE, 2012. 2
- 767 [30] Matt Smith. Common lung diagnostic tests. <https://www.webmd.com/lung/breathing-diagnostic-tests>,
768 2022. 1
- 769 [31] Healthwise Staff. Chest x-ray. <https://www.healthlinkbc.ca/tests-treatments-medications-medical-tests/chest-x-ray>, 2021. 1
- 770 [32] Mingxing Tan, Bo Chen, Ruoming Pang, Vijay Vasudevan, Mark Sandler, Andrew Howard, and Quoc V Le. Mnasnet: Platform-aware neural architecture search for mobile. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2820–2828, 2019. 3
- 771 [33] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019. 3
- 772 [34] Tolga. Chest x-ray images. <https://www.kaggle.com/datasets/tolgadincer/labeled-chest-xray-images>, 2020. 2
- 773 [35] Guangyu Wang, Xiaohong Liu, Jun Shen, Chengdi Wang, Zhihuan Li, Linsen Ye, Xingwang Wu, Ting Chen, Kai Wang, Xuan Zhang, et al. A deep-learning pipeline for the diagnosis and discrimination of viral, non-viral and covid-19 pneumonia from chest x-ray images. *Nature biomedical engineering*, 5(6):509–521, 2021. 1
- 774 [36] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhdadi Bagheri, and Ronald M Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2097–2106, 2017. 2, 3
- 775 [37] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhdadi Bagheri, and Ronald M Summers. Nih chest x-rays. <https://www.kaggle.com/datasets/nih-chest-xrays/data>, 2017. 2, 3
- 776 [38] Wang X, Peng Y, Lu L, Lu Z, Bagheri M, and Summers RM. Nih clinical center provides one of the largest
777 publicly available chest x-ray datasets to scientific
778 community. <https://www.nih.gov/news-events-news-releases/nih-clinical-center-provides-one-largest-publicly-available-chest-x-ray-datasets-scientific-community>, 2017. 2
- 779 [39] Lizhou Xu, Danyang Li, Sami Ramadan, Yanbin Li,
780 and Norbert Klein. Facile biosensors for rapid de-
781 tection of covid-19. *Biosensors and Bioelectronics*,
782 170:112673, 2020. 1
- 783 [40] Na Zhan, Yingyun Guo, Shan Tian, Binglu Huang, Xi-
784 aoli Tian, Jinjing Zou, Qiutang Xiong, Dongling Tang,
785 Liang Zhang, and Weiguo Dong. Clinical character-
786 istics of covid-19 complicated with pleural effusion.
BMC Infectious Diseases, 21(1):1–10, 2021. 2