

000
001
002
003
004
005
006
007054
055
056
057
058
059
060
061
062
063
064
065
066
067
068
069
070
071
072
073
074
075
076
077
078
079
080
081
082
083
084
085
086
087
088
089
090
091
092
093
094
095
096
097
098
099
100
101
102
103
104
105
106
107

Lung Disease Classification - Applied AI Progress Report

Group-Q

1. Introduction and Problem Statement

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 [29], PCR [1] however chest X-rays are by far the most accessible [9]. 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 using deep learning [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 to X-ray scanners are the biggest problems [26] 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 results [21].

This project is an attempt to compare three backbone architectures and lung disease datasets to identify the type of architecture that works best for lung disease classification. The small dataset problem and the issue of different radiographic contrast [22] is mitigated using data augmentation. Imbalanced data will be handled with a combination of using class weights, and a good backbone model. A detailed comparison of results across architectures and datasets along with the explanation of model results will be provided. The final system will use the best model in terms of both efficiency and F1 score allowing rapid diagnosis of lung diseases leading to immediate initiation of treatment, reducing mortality rate.

2. Proposed Methodologies

In this study, 12 models, four for each of the three datasets will be trained. The first three models will be trained from scratch and the fourth model will be trained using transfer learning. The hyperparameters will be fixed across models to produce comparable results. Next, hyperparameters will be tuned to find the best model. Finally, the models will be visualized using t-SNE and Grad-CAM [11] to explain model results. Before training, the images were pre-processed using histogram equalization and Gaussian blur with a 5x5



Figure 1. Effect of pre-processing on Chest X-ray images.

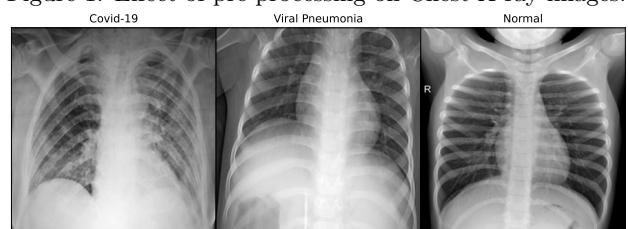


Figure 2. Sample images of Chest X-rays.

Dataset	No. of Images	Classes	Size
COVID [5, 6, 25]	10k:3.6k:1.3k	3	299 ²
Pneumonia [17, 33]	3k:1.5k:1.5k	3	224 ²
Chest X-Ray8 [36, 37]	25k:12k:6k:5k :3k:2.7k:2.6k	7	1024 ²

Table 1. Shortlisted Datasets.

filter as Giełczyk *et al.* [10] showed that this improved the F1 score by 4% for chest X-ray classification. Visually, the contrast of the scan improved and allowed irregularities to stand out as shown in Fig. 1. Next, the scans were divided into train, validation and test with the 70:15:15 split. During training, images were augmented using RandomHorizontalFlip, RandomAdjustSharpness, and RandomAutocontrast in Pytorch [24] to increase the number of images the model gets to learn from. To train the model, cosine annealing with warm restarts [19] was used along with the Adam optimizer [18] and the cross entropy loss function.

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]. The **COVID** dataset was created using 43 [8] different publications. [5, 6, 25] X-rays with widespread, hazy, and irregular ground glass opacities are of the COVID-19

108	Arch.	Params (Mil.)	Layers	FLOPS (Bil.)	Imagenet Acc.
109	MobileNet	5.5	18	8.7	92.6
110	EfficientNet	7.8	25	25.8	94.9
111	Resnet	21.8	34	153.9	91.4

Table 2. Shortlisted Backbone Architectures.

class [16]. Whereas, the ones with haziness in the lower regions only [40] are viral pneumonia cases as shown in Fig. 2. The **Pneumonia**, dataset contains scans from pediatric patients. [17, 33] Scans with one white condensed area affecting only one side of the lungs are tagged bacterial pneumonia [2]. X-rays which show bilateral patchy areas of consolidation are classified as viral pneumonia [13]. **Chest X-ray 8** dataset was released by NIH [38] with over 100k chest X-ray images and their radiological reports which Wang *et al.* [36] used to create disease labels through NLP. [37] It contains 15 classes but only 7 were chosen for this study. Furthermore, normal class images were under-sampled to choose only one scan per patient. **Backbone Architectures:** (Tab. 2) of various configuration and blocks were chosen. Other selection criteria were the *number of trainable parameters*, important as total training time and hardware resources are limited for this project and the *top 5 classification accuracy* 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. [14] **MobileNet V3 Large:** uses depthwise separable convolution from MobileNet V2 [27] along with squeeze-excitation blocks in residual layers from MnasNet [31]. 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]

3. Attempts at Solving the Problem

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

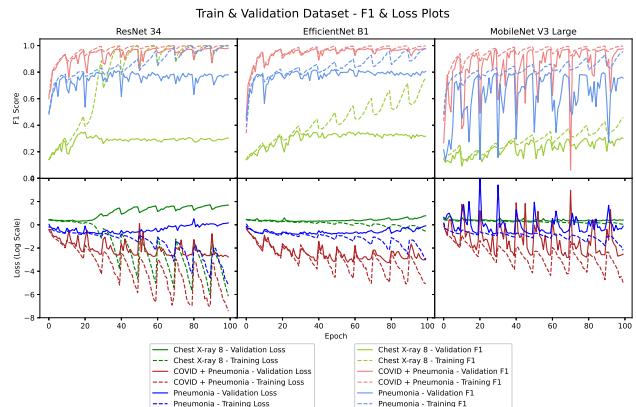


Figure 3. Train & Val F1 & Loss plots for the 9 models.

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

4. Future Improvement

The model with best validation loss is being saved to evaluate the test set but, going forward, the models will be trained for 50 epochs to save time as all the models start to overfit after that point. One model for each dataset will be trained using transfer learning and its performance will be compared with models trained from scratch. Ablation study will be performed to see how different hyperparameters affect the F1 score of the models. Class weights will be explored to improve model performance and a confusion matrix of the test set will be produced to understand model misclassifications. Further, t-SNE [34] and grad-CAM [11] will be used to visualize the trained models.

216

217
References

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

- [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. 1
- [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. 2
- [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. 1
- [4] Aleksander Botev, Matthias Bauer, and Soham De. Regularising for invariance to data augmentation improves supervised learning. *arXiv preprint arXiv:2203.03304*, 2022. 2
- [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. 1
- [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. 1
- [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. 1
- [8] Società Italiana di Radiologia. Covid pneumonia dataset. <https://sirm.org/category/senza-categoria/covid-19/>, 2020. 1
- [9] Guy Frija, Ivana Blažić, Donald P Frush, Monika Hierath, Michael Kawooya, Lluis Donoso-Bach, and Boris Brkljačić. How to improve access to medical imaging in low-and middle-income countries? *EClinicalMedicine*, 38:101034, 2021. 1
- [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. 1
- [11] Jacob Gildenblat and contributors. Pytorch library for cam methods. <https://github.com/jacobgil/pytorchGradCam>, 2021. 1, 2

- [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. 1
- [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. 2
- [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. 2
- [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. 2
- [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. 2
- [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. 1, 2
- [18] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 1
- [19] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016. 1
- [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. 1
- [21] Aleksandra Mojsilovic. Introducing ai explainability 360. <https://www.ibm.com/blogs/research/2019/08/ai-explainability-360/>, 2019. 1
- [22] Andrew Murphy. Radiographic contrast. <https://radiopaedia.org/articles/radiographic-contrast>, 2022. 1
- [23] Loris Nanni, Michelangelo Paci, Sheryl Brahnam, and Alessandra Lumini. Comparison of different image data augmentation approaches. *Journal of Imaging*, 7(12):254, 2021. 2
- [24] PyTorch. Transforming and augmenting images. <https://pytorch.org/vision/stable/transforms.html>. 1
- [25] 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 the effect of image enhancement techniques on covid-19 detection using chest x-ray images. *Computers in Biology and Medicine*, 132:104319, 2021. 1

- 324 [26] Melissa Rohman. Ai performs poorly when
325 tested on data from multiple health systems.
326 <https://healthimaging.com/topics/artificial-intelligence/ai-poorly-detects-pneumonia-chest-x-rays>, 2018. 1 378
327
328
- 329 [27] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2:
330 Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and*
331 *pattern recognition*, pages 4510–4520, 2018. 2 379
332
333
- 334 [28] Janaki Sivakumar, K Thangavel, and P Saravanan.
335 Computed radiography skull image enhancement using wiener filter. In *International Conference on Pattern*
336 *Recognition, Informatics and Medical Engineering (PRIME-2012)*, pages 307–311. IEEE, 2012. 1 380
337
338
- 339 [29] Matt Smith. Common lung diagnostic tests. <https://www.webmd.com/lung/breathing-diagnostic-tests>,
340 2022. 1 381
341
342
- 343 [30] Healthwise Staff. Chest x-ray. <https://www.healthlinkbc.ca/tests-treatments-medications-medical-tests/chest-x-ray>, 2021. 1 382
344
345
- 346 [31] 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. 2 383
347
348
- 349 [32] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019. 2 384
350
351
- 352 [33] Tolga. Chest x-ray images. <https://www.kaggle.com/datasets/tolgadincer/labeled-chest-xray-images>, 2020. 1, 2 385
353
354
- 355 [34] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning*
356 *research*, 9(11), 2008. 2 386
357
358
- 359 [35] Guangyu Wang, Xiaohong Liu, Jun Shen, Chengdi
360 Wang, Zhihuan Li, Linsen Ye, Xingwang Wu, Ting
361 Chen, Kai Wang, Xuan Zhang, et al. A deep-learning
362 pipeline for the diagnosis and discrimination of viral,
363 non-viral and covid-19 pneumonia from chest x-ray
364 images. *Nature biomedical engineering*, 5(6):509–521,
365 2021. 1 387
366
367
- 368 [36] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu,
369 Mohammadhadi Bagheri, and Ronald M Summers.
370 Chestx-ray8: Hospital-scale chest x-ray database and
371 benchmarks on weakly-supervised classification and lo-
372 calization of common thorax diseases. In *Proceedings*
373 *of the IEEE conference on computer vision and pattern*
374 *recognition*, pages 2097–2106, 2017. 1, 2 388
375
376
- 377 [37] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mo-
378 hamadhadi Bagheri, and Ronald M Summers. Nih
379 chest x-rays. <https://www.kaggle.com/datasets/nih-chest-xrays/data>, 2017. 1, 2 389
380
381
- 382 [38] Wang X, Peng Y, Lu L, Lu Z, Bagheri M, and Summers RM. Nih clinical center provides one of the largest
383 publicly available chest x-ray datasets to scientific
384 community. <https://www.nih.gov/news-events-news-releases/nih-clinical-center-provides-one-largest-publicly-available-chest-x-ray-datasets-scientific-community>, 2017. 2 390
385
386
- 387 [39] Lizhou Xu, Danyang Li, Sami Ramadan, Yanbin Li,
388 and Norbert Klein. Facile biosensors for rapid de-
389 tection of covid-19. *Biosensors and Bioelectronics*,
390 170:112673, 2020. 1 391
391
392
- 393 [40] Na Zhan, Yingyun Guo, Shan Tian, Binglu Huang, Xi-
394 aoli Tian, Jinjing Zou, Qiutang Xiong, Dongling Tang,
395 Liang Zhang, and Weiguo Dong. Clinical character-
396 istics of covid-19 complicated with pleural effusion.
397 *BMC Infectious Diseases*, 21(1):1–10, 2021. 2 398
398
399
- 400 [41] 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. 2 401
401
402
- 403 [42] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019. 2 403
403
404
- 405 [43] Tolga. Chest x-ray images. <https://www.kaggle.com/datasets/tolgadincer/labeled-chest-xray-images>, 2020. 1, 2 405
405
406
- 407 [44] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning*
408 *research*, 9(11), 2008. 2 406
408
409
- 410 [45] Guangyu Wang, Xiaohong Liu, Jun Shen, Chengdi
411 Wang, Zhihuan Li, Linsen Ye, Xingwang Wu, Ting
412 Chen, Kai Wang, Xuan Zhang, et al. A deep-learning
413 pipeline for the diagnosis and discrimination of viral,
414 non-viral and covid-19 pneumonia from chest x-ray
415 images. *Nature biomedical engineering*, 5(6):509–521,
416 2021. 1 417
417
418
- 419 [46] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu,
420 Mohammadhadi Bagheri, and Ronald M Summers.
421 Chestx-ray8: Hospital-scale chest x-ray database and
422 benchmarks on weakly-supervised classification and lo-
423 calization of common thorax diseases. In *Proceedings*
424 *of the IEEE conference on computer vision and pattern*
425 *recognition*, pages 2097–2106, 2017. 1, 2 425
425
426
- 427 [47] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mo-
428 hamadhadi Bagheri, and Ronald M Summers. Nih
429 chest x-rays. <https://www.kaggle.com/datasets/nih-chest-xrays/data>, 2017. 1, 2 429
429
430
- 431 [48] Wang X, Peng Y, Lu L, Lu Z, Bagheri M, and Summers RM. Nih clinical center provides one of the largest
432 publicly available chest x-ray datasets to scientific
433 community. <https://www.nih.gov/news-events-news-releases/nih-clinical-center-provides-one-largest-publicly-available-chest-x-ray-datasets-scientific-community>, 2017. 2 432
432
433