Using Deep Learning to Detect Pathology in Lateral View Chest X-rays

VILLANOVA UNIVERSITY
IGNITE CHANGE. GO NOVA.

Jessica Yarnall, MS[1] & Benjamin R Mitchell, PhD[2]

[1]jyarnal1@villanova.edu

[2]bmitche1@swarthmore.edu

Department of Computing Sciences, Villanova University, PA, USA

Introduction

- Chest x-rays are a valuable diagnostic tool for many different diseases.
- Deep learning has the potential to assist radiologists and help improve patient care.
- Explore the use of deep learning to classify lateral view chest x-rays from the MIMIC-CXR-JPG[4] dataset
  - Conduct multi-label classification experiments and single label classification experiments
    - Multi-label classification tasks found that VGG16[8] achieved the highest accuracy of the models tested

Figure 1: Sample image from the MIMIC-CXR dataset

- Found single label classification results for the following twelve pathologies: atelectasis, cardiomegaly, consolidation, edema, enlarged cardiomediastinum, fracture, lung lesion, lung opacity, plural effusion, plural other, pneumonia, and pneumothorax.
- Model performance varied for each pathology, with pleural effusion achieving the highest accuracy of 83.5% and among the highest recall, specificity, and precision results.
- We believe this work serves as a baseline for classifying the lateral view images of the MIMIC-CXR dataset.

Materials and Methods*

- Lateral view images from the MIMIC-CXR-JPG dataset.
  - Images are resized to 224x224 pixels
- CheXpert labeler[4]
  - Positive labels are treated as 1
  - Uncertain, negative, and unmentioned labels are treated as 0
- VGG16, ResNet50[2] and DenseNet121[3] architectures for multilabel classification tasks
  - Transfer learning using ImageNet weights
  - Accuracy and loss are measured for architectures
- VGG16 to classify with or without pathology of lateral view x-rays for twelve pathologies. Pathology splits can be found in Table 1.
  - For each pathology, an equal number of No Finding images were randomly selected and added to the dataset
  - Utilized ReLu activation function and 10 epochs

Results

- Multi-label classification experiments found VGG16 had the highest accuracy among the models tested: VGG16, ResNet50 and DesneNet121
- Results for each pathology are found in Table 1
- Figure 2 shows the confusion matrix for each pathology.

Pathology	Accuracy	Recall	Specificity	Precision	No. of Images
Atelectasis	77.30%	76.98%	77.63%	78.16%	14,268
Cardiomegaly	75.79%	71.13%	80.61%	79.13%	13,899
Consolidation	63.09%	31.42%	96.02%	89.13%	2,558
Edema	82.41%	83.02%	81.80%	82.00%	6,423
Enlarged Cardio.	75.26%	83.42%	66.67%	72.49%	1,939
Fracture	62.23%	75.69%	48.39%	60.12%	2,517
Lung Lesion	65.37%	60.29%	70.05%	65.00%	3,596
Lung Opacity	71.84%	72.96%	70.72%	71.50%	18,645
Pleural Effusion	83.52%	81.37%	85.67%	85.00%	16,659
Pleural Other	71.04%	53.10%	84.93%	73.17%	1,143
Pneumonia	66.31%	61.38%	71.47%	69.25%	7,391
Pneumothorax	76.38%	74.17%	78.60%	77.61%	2,711

Table 1: The testing results for each pathology. Also shows the total number of images before training, validation, and testing splits.

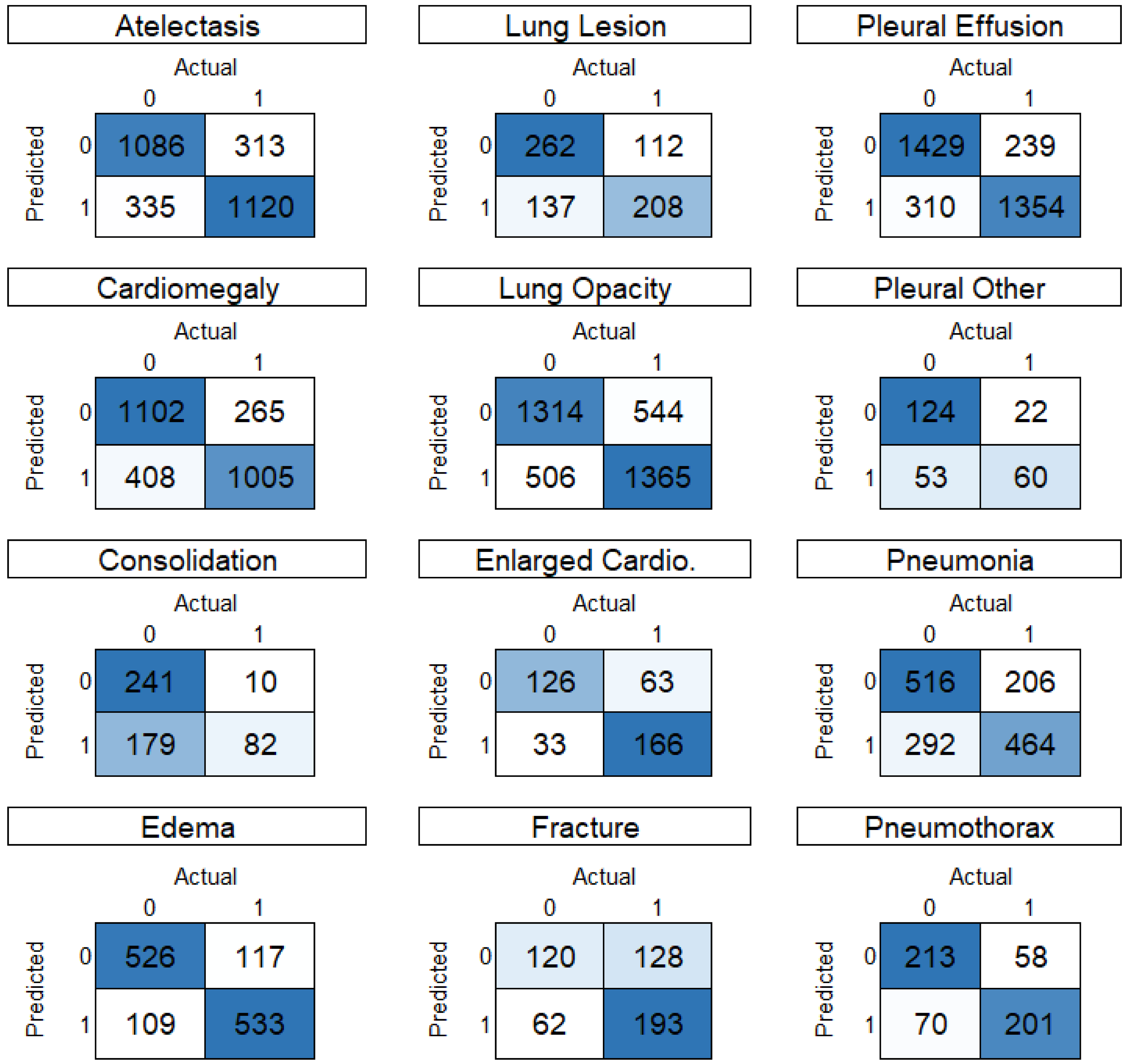

Figure 2: Confusion matrix for each pathology. 0 indicates without pathology and 1 indicates with pathology.

Discussion

One model does not fit all

- VGG16 architecture worked best on the entire lateral multi-label dataset
- Accuracy ranged from 62.23% to 83.52% for individual pathologies

Due to the nature of the classification problem, recall and specificity are more informative

- Do not want to send home sick patients
- Do not want to treat healthy patients
- Recall ranged between 31.42% and 83.42% while specificity ranged between 48.39% and 96.02%
- Consolidation has the highest specificity and lowest recall while fracture scored the lowest specificity and among the highest recall. However, both pathologies scored similarly on accuracy with 63.09% and 62.23% respectively

We are not aware of any published peer-reviewed research that studies the recall, specificity and precision of the lateral view x-rays from MIMIC-CXR

Future Work

- Test our architecture on frontal images
- Utilize a similar architecture to DuelNet[7] to test both lateral and frontal images
- Utilize a similar training technique to Monshi et al.[5] to adjust learning rates across epochs
- Add weights to uncertainty labels
- Partner with radiologists to test on novel images

References and Relevant Work

1. COHEN, J. P., HASHIR, M., BROOKS, R.,AND BERTRAND, H. On the limits of cross-domain generalization in automated x-ray prediction. In Medical Imaging with Deep Learning (2020).
2. HE, K., ZHANG, X., REN, S.,ANDSUN, J. Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016), pp. 770–778.
3. HUANG, G., LIU, Z.,VAN DERMAATEN, L., AND WEINBERGER, K. Q. Densely connected convolutional networks. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017).
4. JOHNSON, A. E. W., POLLARD, T. J., GREENBAUM, N. R., LUNGREN, M. P.,YINGDENG, C., PENG, Y., LU, Z., MARK, R. G., BERKOWITZ, S. J.,AND HORNG, S. Mimic-cxr-jpg, a large publicly available database of labeled chest radiographs, 2019.
5. MONSHI, M. M. A., POON, J.,AND CHUNG, V. Convolutional neural net-work to detect thorax diseases from multi-view chest x-rays. In Neural In-formation Processing (Cham, 2019), T. Gedeon, K. W. Wong, and M. Lee, Eds., Springer International Publishing, pp. 148–158.
6. POOCH, E., BALLESTER, P.,AND BARROS, R. C. Can we trust deep learning models diagnosis? the impact of domain shift in chest radiograph classification. ArXiv abs/1909.01940(2019).
7. RUBIN, J., SANGHAVI, D., ZHAO, C., LEE, K., QADIR, A.,ANDXU-WILSON,M. Large scale automated reading of frontal and lateral chest x-rays using dual convolutional neural networks. CoRR abs/1804.07839 (2018).
8. SIMONYAN, K.,AND ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition, 2014.

*Code used to run experiments and generate results can be found here: https://github.com/jessyarnall/thesis-code