Evaluation of LeNet5 Network for Chest X-ray Image Classification

Sadaf Asrar, David Robison, & Nathan Zencey

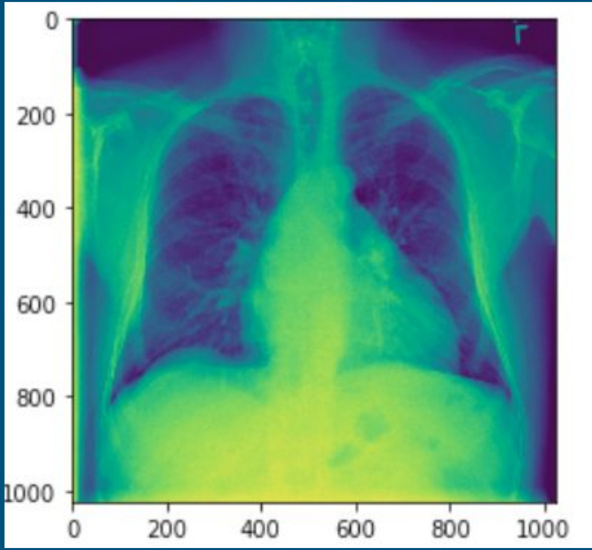
Introduction

Background and Motivation

- The interpretation of radiological images often requires the availability of a physician specialist, the radiologist.
- With the volume of radiological examinations requiring interpretation, access to radiological expertise is becoming more difficult.
- Radiological interpretation provides a paradigm for a computer vision task where a tool could be developed to assist radiologists.

Project Image Data

- National Institute of Health (NIH) Clinical Centers
- 112,120 x-ray images of 30,805 unique patients,
- Images are of size 1024 x 1024 x 3
- Images are labels with ***one or more*** of 14 disease labels or healthy



Label: Cardiomegaly | Emphysema

Previous Work in X-ray Computer Vision

01	**Wang et al**	• Used NLP to extract labels for x-ray images • Benchmarked multiple networks using pretrained Imagenet architectures • Reported that a ResNet architecture had the best overall performance
02	**Mohammad et al.**	• Used an abnormality detection framework, found that GoogLeNet performed best • Performed significant preprocessing e.g., histogram equalization • Reported that performance was not affected by image resizing
03	**Stanford Machine Learning Group**	• Developed a 121 layer Dense Convolutional Neural Network • Achieved superior results to human experts for pneumonia detection and F1 scoring across 14 distinct classes • Used an Adam optimizer with pretrained ImageNet parameters

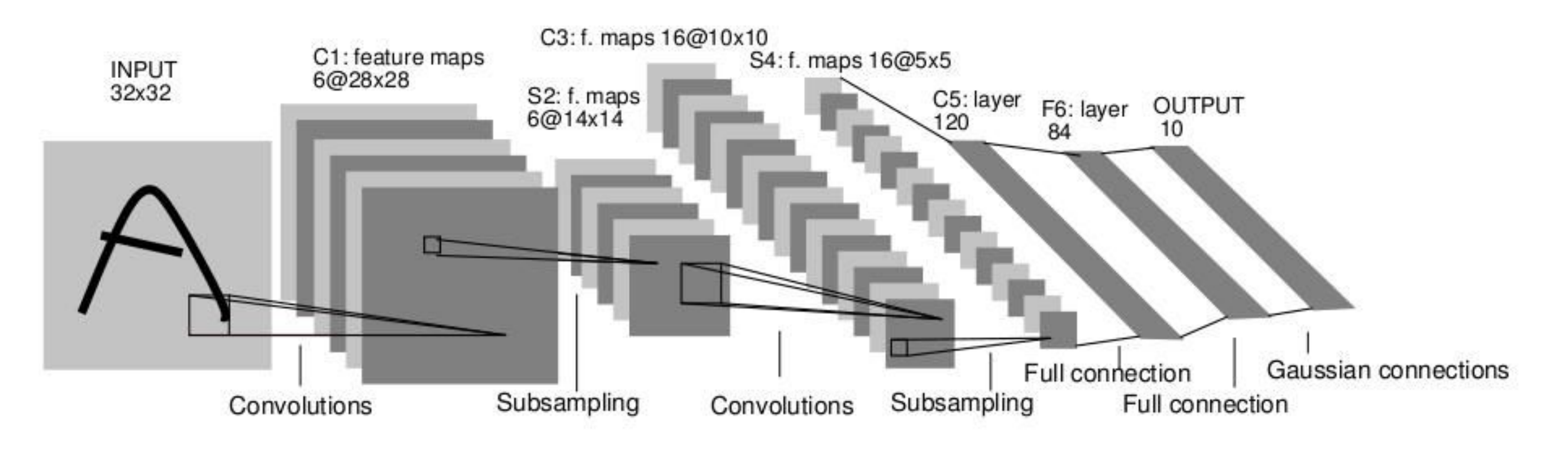
Methods and Data Processing

Project Goals and Learning Outcomes

1. Explore how the performance of the LeNet 5 for a binary classification task using the NIH Chest X-ray images is affected by changes in:
    a. optimization algorithm,
    b. transfer function, and
    c. size of convolution layer kernels
2. Gain subject matter expertise in a specific type of computer vision classification task, namely medical image classification.
3. Gain experience evaluating the quality of and trade-offs between different neural network architectures.

Overview of the LeNet5 Architecture

- Stochastic gradient descent optimization
- Two convolution and pooling layers, two fully connected layers, and an output layer.



Data Preprocessing

- Convert to *binary classification* setup
- Used tools in *skimage* and *opencv* to randomly and crop resize 24,999 images to 224 x 224 x3
- Implemented network in PyTorch
- Train, validation, and test sets of: 85% - 10% - 5% (19,999 - 4,000 - 1,000)

Label: Finding

Label: No Finding

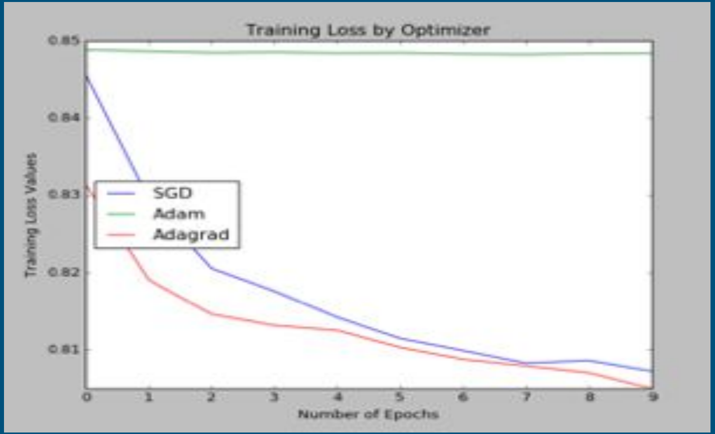
Summary of Experimental Results

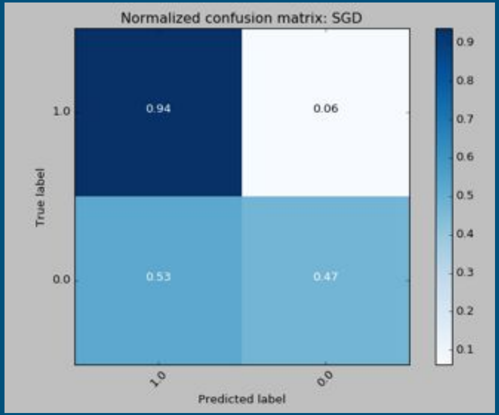
Discussion of Optimization Algorithms:

We experimented with three variants of the gradient descent algorithm:

- Stochastic Gradient Descent (SGD) with momentum,
- Adaptive Moment Estimation (Adam), and
- Adagrad

Network Performance by Optimizer





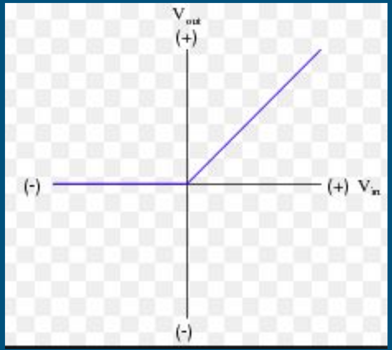
Optimization algorithm	F1 score	Accuracy (%)	Recall (%)	Precision (%)	Run time (sec.)
SGD	0.617	62	94	64	613
Adam	0.374	57	97	56	610
Adagrad	0.5	63	96	60	622

Discussion of Transfer Functions

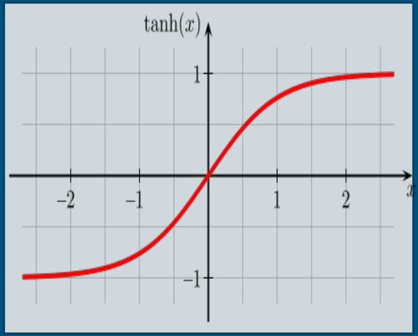
We experimented with three different transfer functions:

- ReLU
- Tanh
- Sigmoid

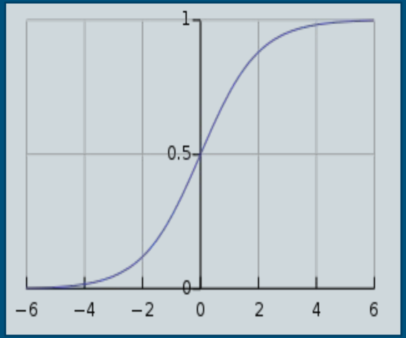
ReLU



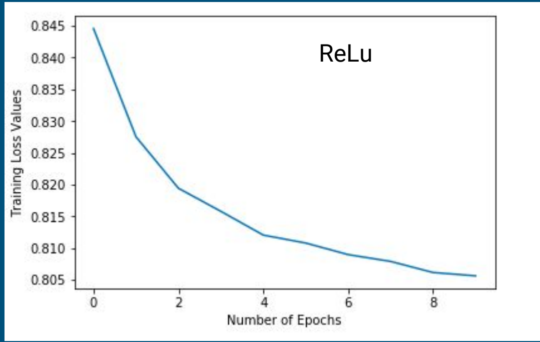
Tanh

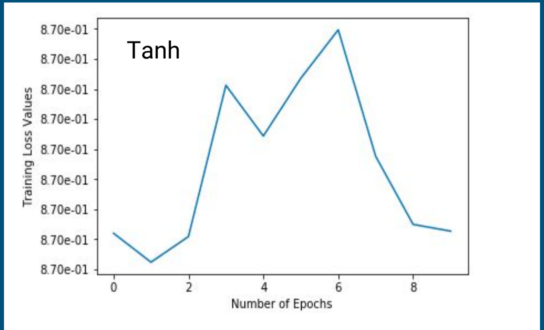


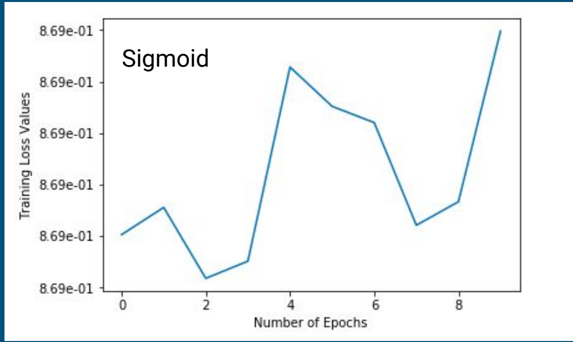
Sigmoid



Network Performance by Transfer Function





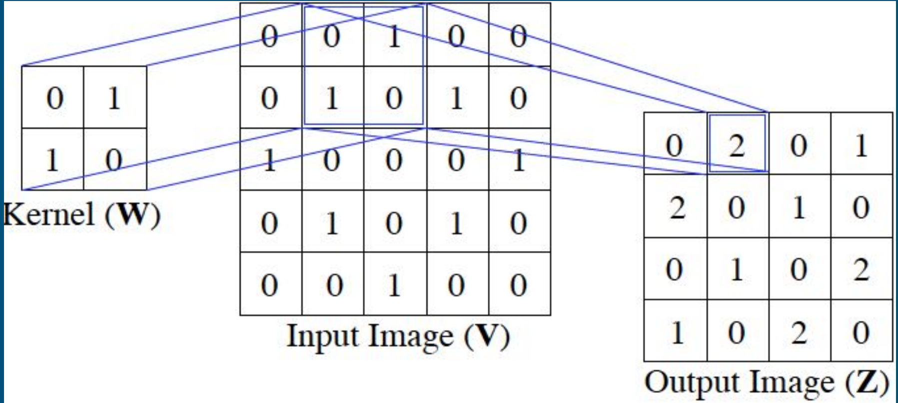


Transfer Function	F1 score	Accuracy (%)	Run time (sec.)
ReLu	0.425	61	545
Tanh	0.588	42	511
Sigmoid	0	57	532

Discussion of Kernels

The kernel is the key operator of the convolution network:

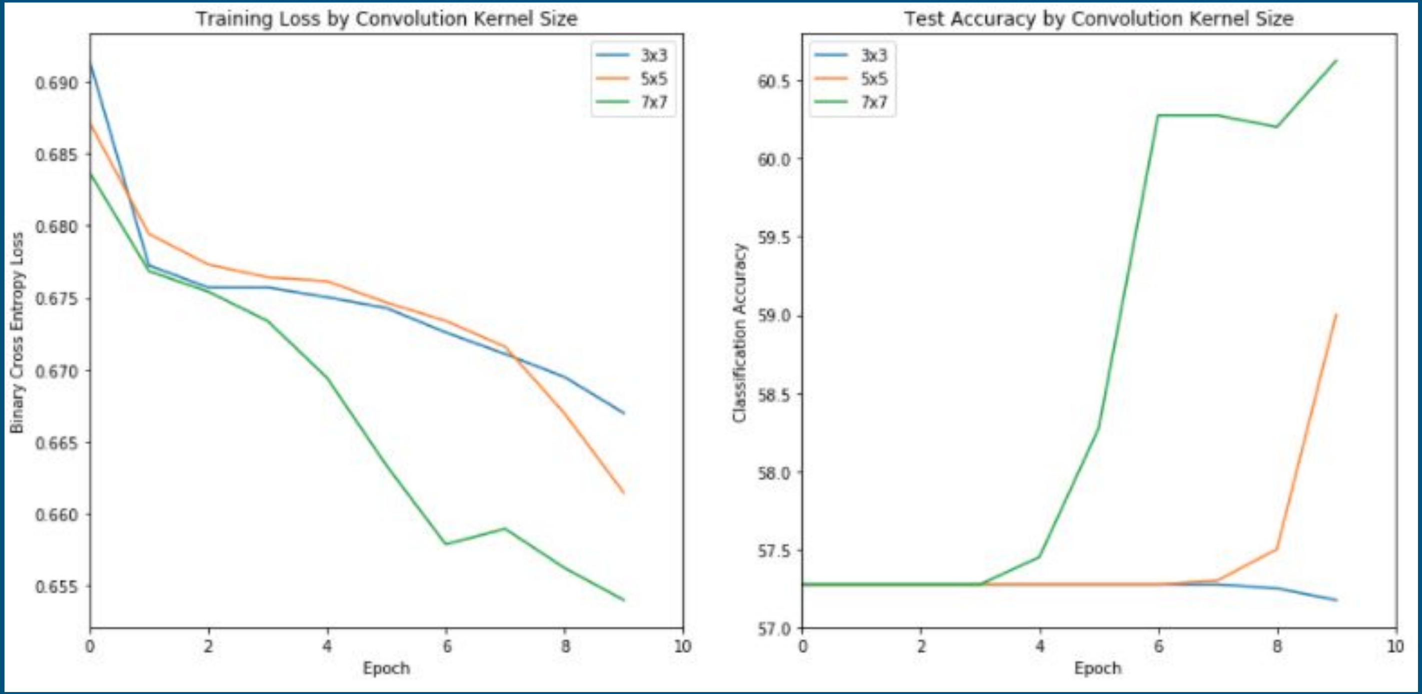
- Has height and width to define its receptive field.
- Convolves the image performing non-matrix multiplication operations to build a feature map.
- Feature map values are learned weights that are updated with each training iteration.



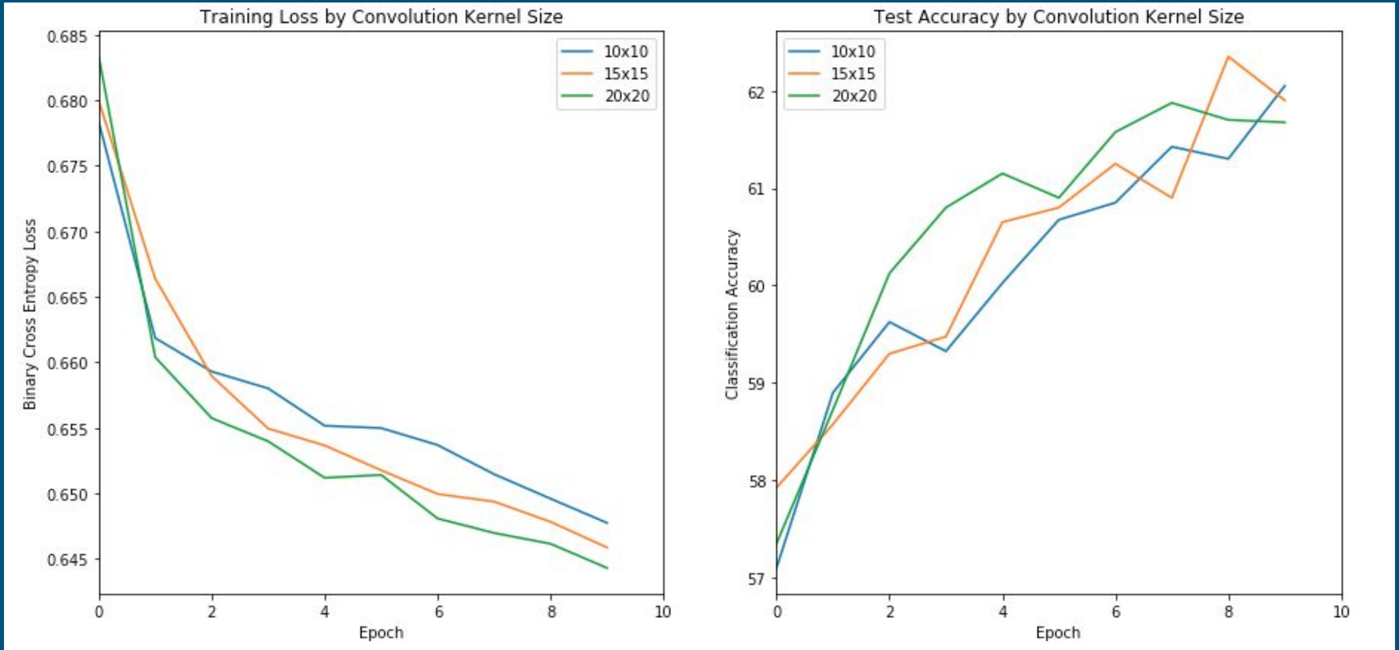
$$z_{i,j} = \sum_{k=1}^{r} \sum_{l=1}^{c} w_{k,l} v_{i+k-1, j+l-1}$$

$$\mathbf{Z} = \mathbf{W} \circledast \mathbf{V}$$

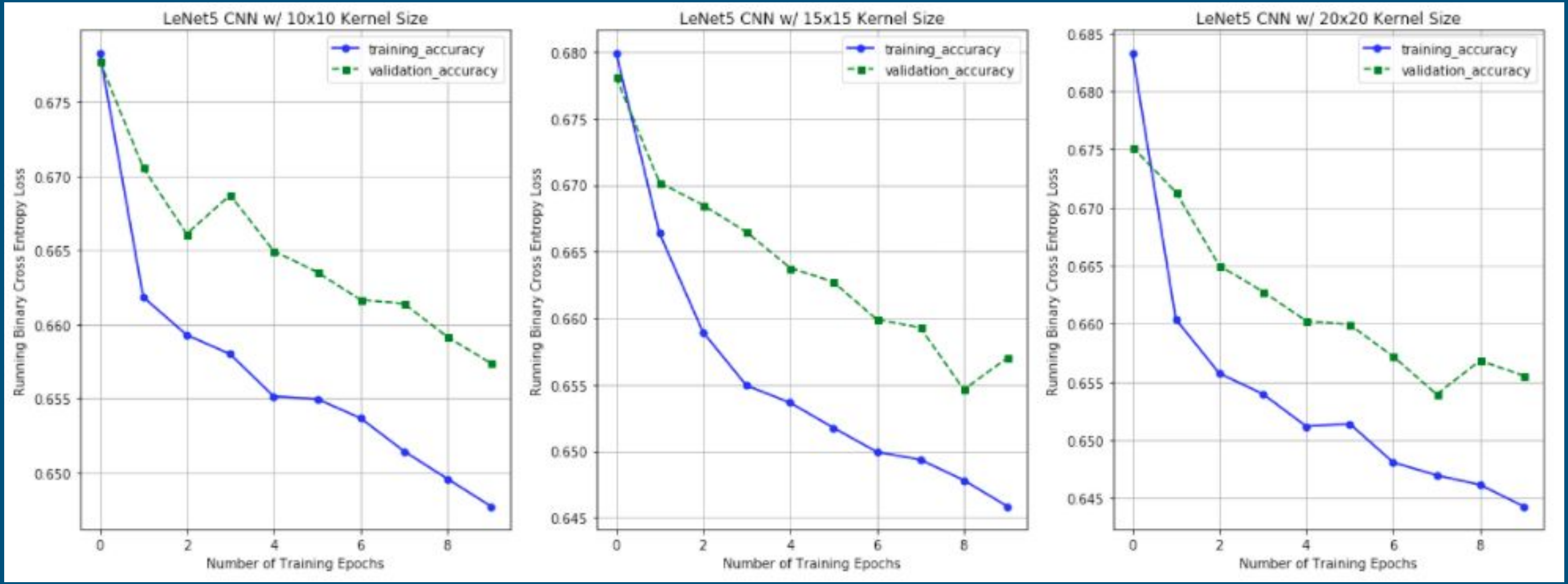
Evaluation of Kernels: Small



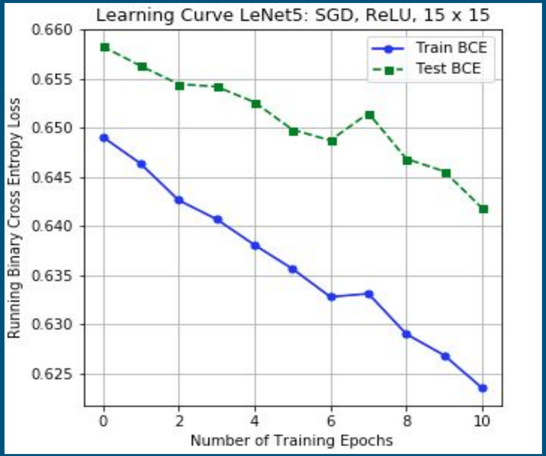
Evaluation of Kernels: Small

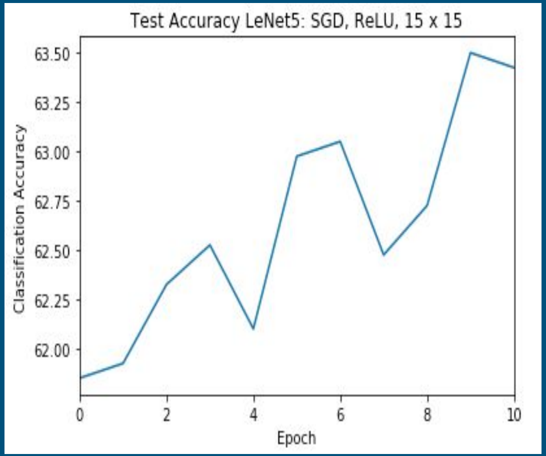


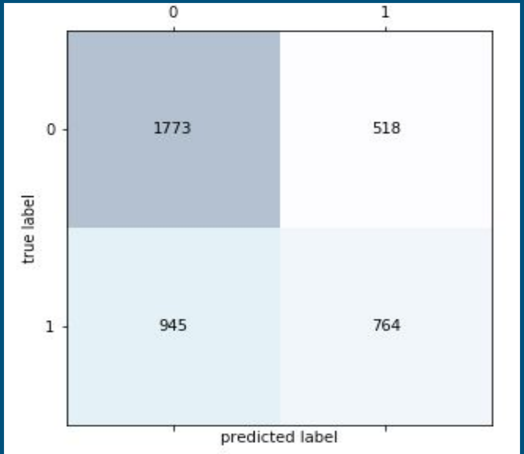
Training Loss and Validation Loss with Large Kernels



Final Model: LeNet5, SGD, ReLU, 15 x 15







Class	Precision	Recall	F1
No FInding	0.65	0.77	0.71
Finding	0.60	0.45	0.51

Conclusion

- For a LeNet 5 Architecture, larger kernel sizes, stochastic gradient descent, and ReLu transfer functions achieve highest performance.
- From a practical perspective, results are mediocre.
- Further gains could be made with dropout and data augmentation.

References

1. British Library. Roentgen's discovery of the x-ray. Available at: https://www.bl.uk/learning/cult/bodies/xray/roentgen.html.
2. Mayo Clinic. Chest X-rays. Available at: https://www.mayoclinic.org/tests-procedures/chest-x-rays/about/pac-20393494.
3. Diagnostic Imaging. In Radiology, Turnaround Time is King. Available at: http://www.diagnosticimaging.com/practice-management/radiology-turnaround-time-king.
4. National Institutes of Health. NIH Clinical Center provides one of the largest publicly available chest x-ray datasets to scientific community. Available at: https://www.nih.gov/news-events/news-releases/nih-clinical-center-provides-one-largest-publicly-available-chest-x-ray-datasets-scientific-community
5. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. IEEE CVPR 2017, http://openaccess.thecvf.com/content_cvpr_2017/papers/Wang_ChestX-ray8_Hospital-Scale_Chest_CVPR_2017_paper.pdf
6. Tariqul, Mohammad, et al. “Abnormality Detection and Localization in Chest X-Rays Using Deep Convolutional Neural Networks.” 27 Sept. 2017, arxiv.org/abs/1705.09850.
7. Rajpurkar, et al. “CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning.” 25 Dec. 2017, arxiv.org/abs/1711.05225.
8. Simonyan, et al. “Very Deep Convolutional Networks for Large-Scale Image Recognition.” 10 Apr. 2015, arxiv.org/abs/1409.1556.
9. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition, Proc. IEEE 86(11): 2278–2324, 1998.
10. Ruderm, Sebastian (2016). “An overview of gradient descent optimisation algorithms.” 15 Sept. 2016, arXiv preprint arXiv:1609.04747.
11. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: Proceedings of ICLR. (2015)