Group Proposal

Machine Learning II

Sertan Akinci

Kristin Levine

April 2021

**• What problem did you select and why did you select it?**

We decided to examine the [COVID-19 Radiography Database](https://www.kaggle.com/tawsifurrahman/covid19-radiography-database) from Kaggle. We selected this problem because it combines two of our interests: healthcare data and image classification. Also we are hoping to improve the way to identify, diagnose, and help treat critical lung diseases. Our model will be able to identify 3 different lung infections as well as a healthy lung which can be used to assist healthcare professionals where they can choose which tests require further screening based on the accuracy probabilities of the diagnosis.

**• What database/dataset will you use? Is it large enough to train a deep network?**

This is a large collection of chest X-rays from [a number of different sources](https://www.kaggle.com/tawsifurrahman/covid19-radiography-database). There are four classes of images:

3616 COVID chest X-rays

6023 Lung Opacity chest X-rays (non-COVID lung infection)

10,192 normal chest X-rays

134 Viral pneumonia X-ray

Totaling to over 20,000 x-ray images which is large enough to train a deep network.

**• What deep network will you use? Will it be a standard form of the network, or will you**

**have to customize it?**

We wanted to use a variety of pre-trained models on this dataset, including Resnet, Xception, and VGG16. We also wanted to try ensembling these pre-trained methods, both to learn how that technique works and to try to improve our overall results.

**• What framework will you use to implement the network? Why?**

We decided to use tensorflow because we had not had a chance to utilize this platform since we learned it in class. It also seems straightforward to work with the pre-trained models in that framework.

**• What reference materials will you use to obtain sufficient background on applying the**

**chosen network to the specific problem that you selected**?

We plan to use these papers, created on different subsets of this database, to give us ideas on how to implement our models.

-M.E.H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M.A. Kadir, Z.B. Mahbub, K.R. Islam, M.S. Khan, A. Iqbal, N. Al-Emadi, M.B.I. Reaz, M. T. Islam, “Can AI help in screening Viral and COVID-19 pneumonia?” IEEE Access, Vol. 8, 2020, pp. 132665 - 132676. [Paper link](https://ieeexplore.ieee.org/document/9144185)

-Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Kashem, S.B.A., Islam, M.T., Maadeed, S.A., Zughaier, S.M., Khan, M.S. and Chowdhury, M.E., 2020. Exploring the Effect of Image Enhancement Techniques on COVID-19 Detection using Chest X-ray Images. [Paper Link](https://doi.org/10.1016/j.compbiomed.2021.104319)

**• How will you judge the performance of the network? What metrics will you use?**

We will look at the categorical accuracy. We also also plan to look at the precision, recall, and f1-scores.

**• Rough Schedule for completing the project**

Two weeks to inspect and familiarize ourselves with the data.

Two weeks to implement the methods (since we are using a new platform)

A week to write a report