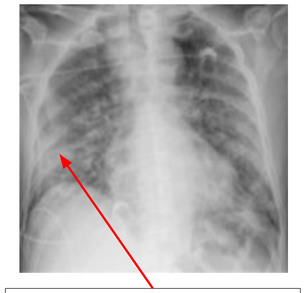
# Level-Up Lungs: Identifying COVID-19 Infected Lungs Using Machine Learning

Sanjeevani Choudhery, Samantha Ray, Sournav Bhattacharya, Rohan Singh Wilkho, and Mounika Kunduru

## Is It COVID-19?

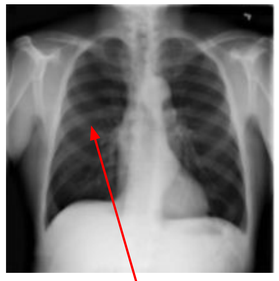
Note: These are inferences made by computer science students, not medical professionals

COVID-19 Patient X-Ray [1]



**White dense** residue in lungs makes X-ray "blurry".

Healthy Patient X-Ray [1]



**Dark** background comes through with **clear** lungs.

**Motivation**: Given lung X-rays and patient demographics (gender, location and age), we can predict with **95.6%** accuracy in 5-fold cross validation whether the patient has COVID-19.

## Making it Black and White

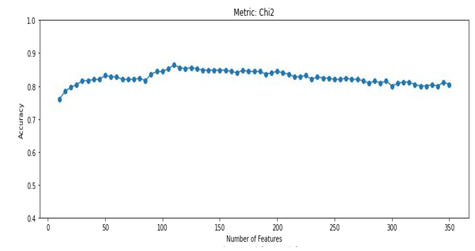
Dense matter shows up as white in x-rays, i.e., bones, certain organs such as the heart, and abnormalities in the lungs.

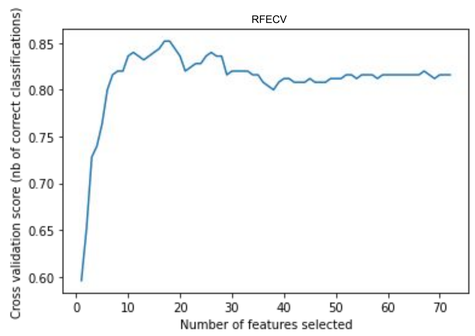
| Standard Computer Vision Features | Domain-Knowledge Features |
|---|---|
| • Edge Detection: Canny, Hessian, Meijering, Laplace, Sobel (Horizontal and Vertical)<br>• Histogram of Gradients (HoG)<br>• Gabor Filter with various values for theta | • Blob Counts<br>• Light/Dark Patch Counts<br>• Edge Detection Ratio<br>• Pixel Value Statistics (Mean, 1st and 3rd Quartile, Light/Dark Percentage) |

## Selecting the Best Features

| Filter Method | Wrapper Method |
|---|---|
| • Metrics Tested: correlation with with the target variable, chi-squared, mutual information gain, and Fisher's criterion.<br>• Accuracy : 86% with ~200 features based on Chi-Square<br>• Time: ~ 3 minutes to run each metrics with 5-Fold CV | • RFECV with SVC linear kernel<br>• Accuracy : 0.98 with 79 selected features using SVC<br>• Time : <1 minute total with GPU acceleration |

| Important Feature Categories Selected |
|---|
| • HoG<br>• Canny<br>• Gabor<br>• Hessian<br>• Laplace<br>• Sobel<br>• Blob Count - Difference of Gaussian<br>• Pixel Value Percentages<br>• Light/Dark Patch Counts<br>• Location |




## 5-Fold Cross Validation Accuracy

| Model | w/o Feature Selection | w/ Feature Selection |
|---|---|---|
| **AdaBoost (n_estimators = 200)** | 0.744 | 0.816 |
| **Random Forest (n_estimators = 100)** | 0.736 | 0.776 |
| **SVM (kernel = linear)** | 0.820 | **0.956** |
| **Multilayer Perceptron (hidden layer size = 200)** | 0.824 | 0.948 |

## Best Performing Model: SVM

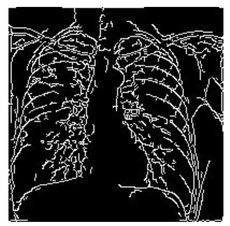
| Feature Subset | w/o Feature Selection | w/ Feature Selection |
|---|---|---|
| **Standard Computer Vision Features Only** | 0.708 | 0.800 |
| **Standard Computer Vision Features + Domain-Knowledge Features** | 0.688 | 0.828 |
| **Standard Computer Vision Features + Domain-Knowledge Features + Demographic Features** | 0.820 | **0.956** |

## Processing Pipeline

**Original Image Cropped and Resized to 200x200**

**Preprocessing: Dilate, Erode, and Sharpen**

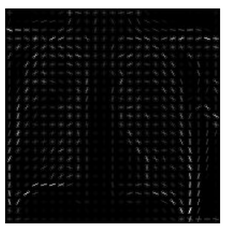
**These preprocessing steps standardize the inputs and make important features clearer**



A. Canny

B. Hessian

C. Gabor



D. HoG

E. Laplace

**And all the other features generated**

**PCA**

**Feature Selection**

## References

[1] Chandra, T. B., Verma, K., Singh, B. K., Jain, D., & Netam, S. S. (2021). Coronavirus disease (COVID-19) detection in Chest X-Ray images using majority voting based classifier ensemble. Expert systems with applications, 165, 113909. https://doi.org/10.1016/j.eswa.2020.113909

[2] Parekh, M., Donuru, A., Balasubramanya, R., & Kapur, S. (2020). Review of the Chest CT Differential Diagnosis of Ground-Glass Opacities in the COVID Era. Radiology, 297(3), E289–E302. https://doi.org/10.1148/radiol.2020202504

[3] Mayo Clinic. Chest X-ray. Last accessed Nov. 21, 2020. Available: https://www.mayoclinic.org/tests-procedures/chest-x-rays/multimedia/chest-x-ray/img-20006961