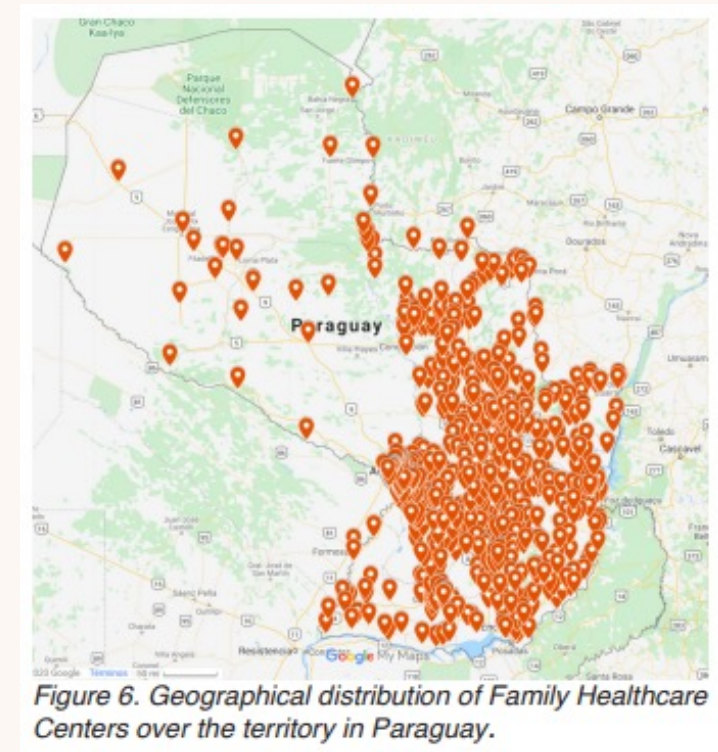


Applying ML to Lung X-Ray Images for Real-Time Diagnosis

By Alexa Wingate

Motivation: To apply ML to a Real-World Problem

- Providing early diagnosis in remote locations continues to be a healthcare challenge
 - 160,000 people live in the Chaco (Western) region of Paraguay
 - 30% of the population does not have access to health services and there are long waits to get medical appointments
- Telemedicine partially addresses these issues but is limited in two ways:
 - Diagnosis occurs after the patient leaves the treatment center
 - Flaws in examination techniques produce poor quality data from which a diagnosis can be made
- Machine learning at the point of examination can mitigate these issues



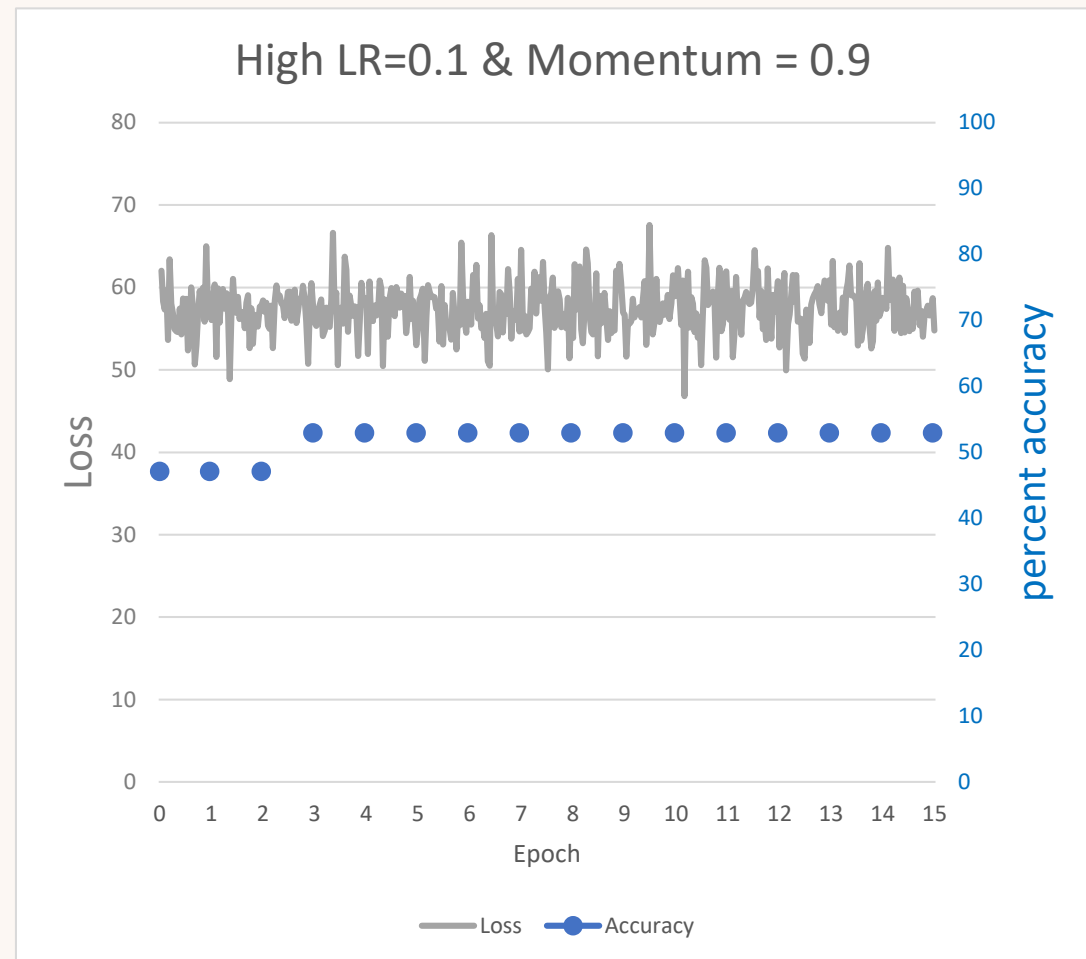
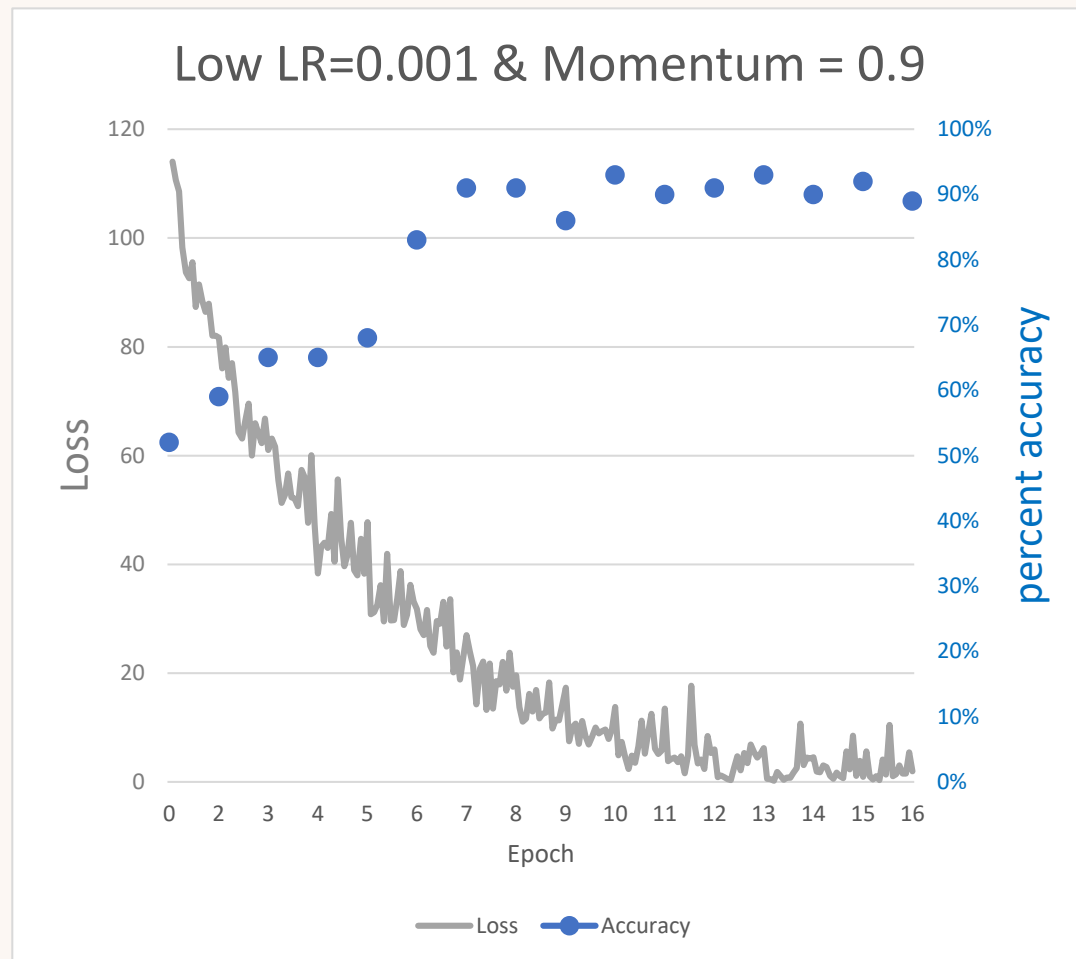
The data

- The data includes 21,165 lung X-ray images in four categories:
 1. Normal
 2. COVID
 3. Lung Opacity
 4. Viral Pneumonia
- The data may not be perfect for training because:
 - Almost half (10,192) of the images are normal
 - The age and sex proportions of the images are unknown
 - The quality of the images varies (obstruction, angle/perspective)

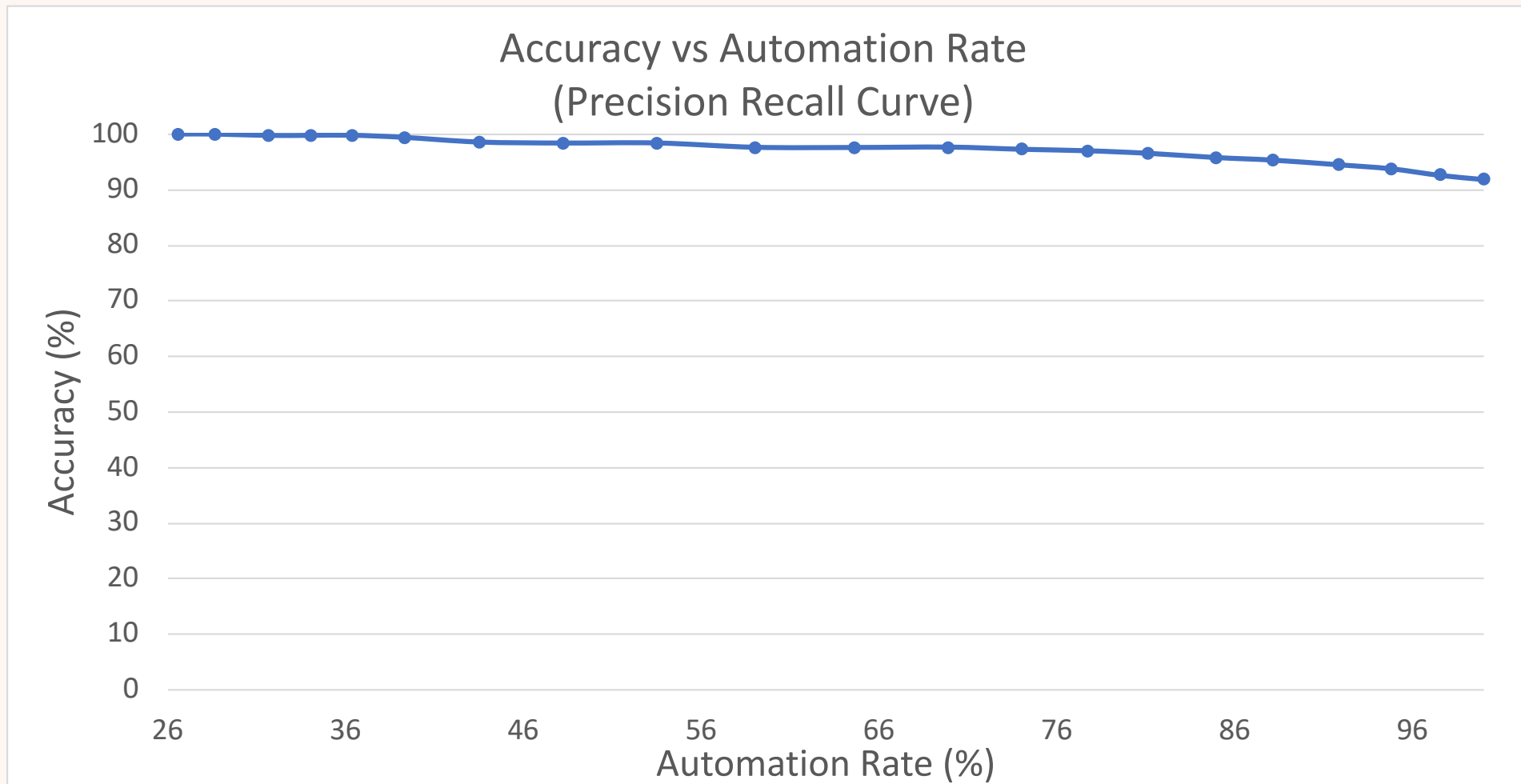
Designing & testing in small steps

- The algorithm was designed:
 - Using three fully connected layers and three convolutional layers that use a rectified linear unit (RELU) activation function
 - Using stochastic gradient descent with cross entropy loss
- Testing in 3 stages:
 1. Include only two categories: COVID and Healthy
 - a) Optimize the learning rate and number of epochs to improve accuracy
 - b) Generate precision recall curves
 2. Extract and independently analyze “confusing” images
 3. Test algorithm on all four categories

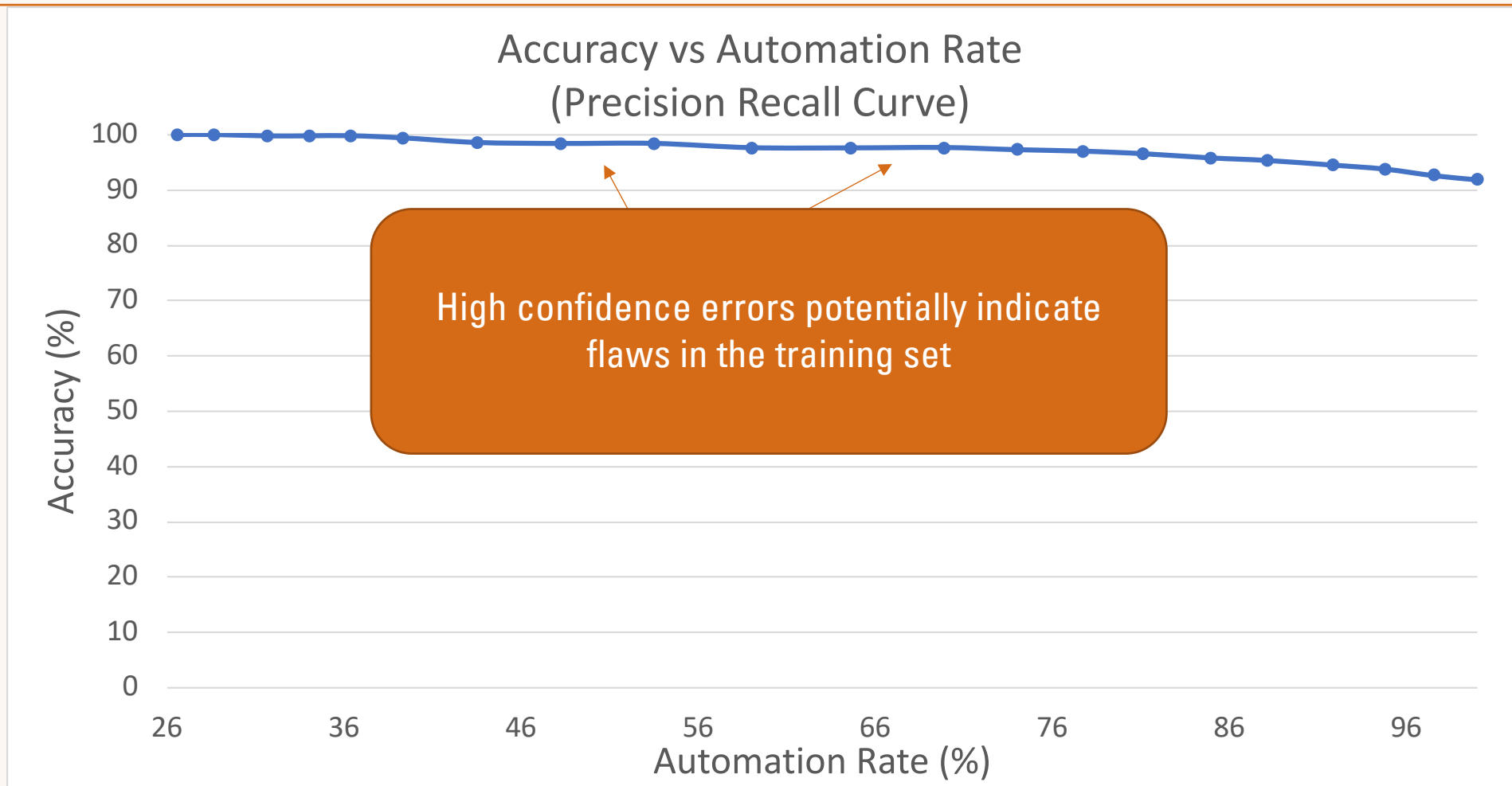
With only two categories (Healthy & COVID),
learning rate at 0.001 and 9 epochs was sufficient



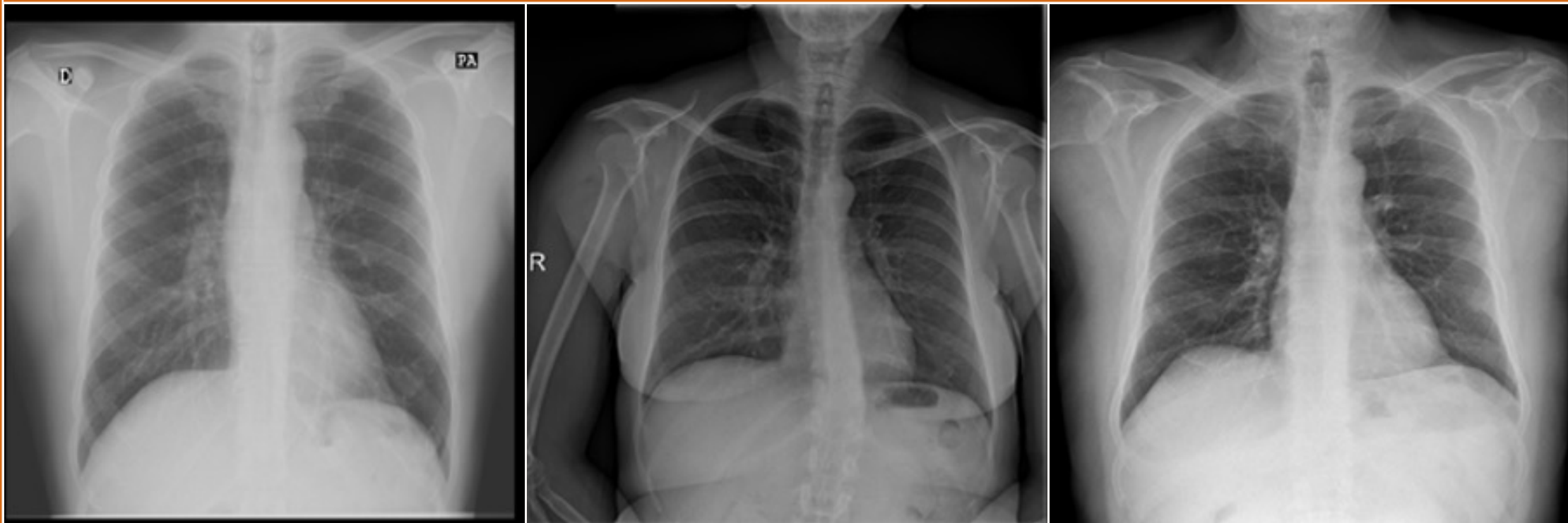
The precision recall curve shows increased accuracy by excluding low-confidence images



The precision recall curve also raises questions around dataset quality

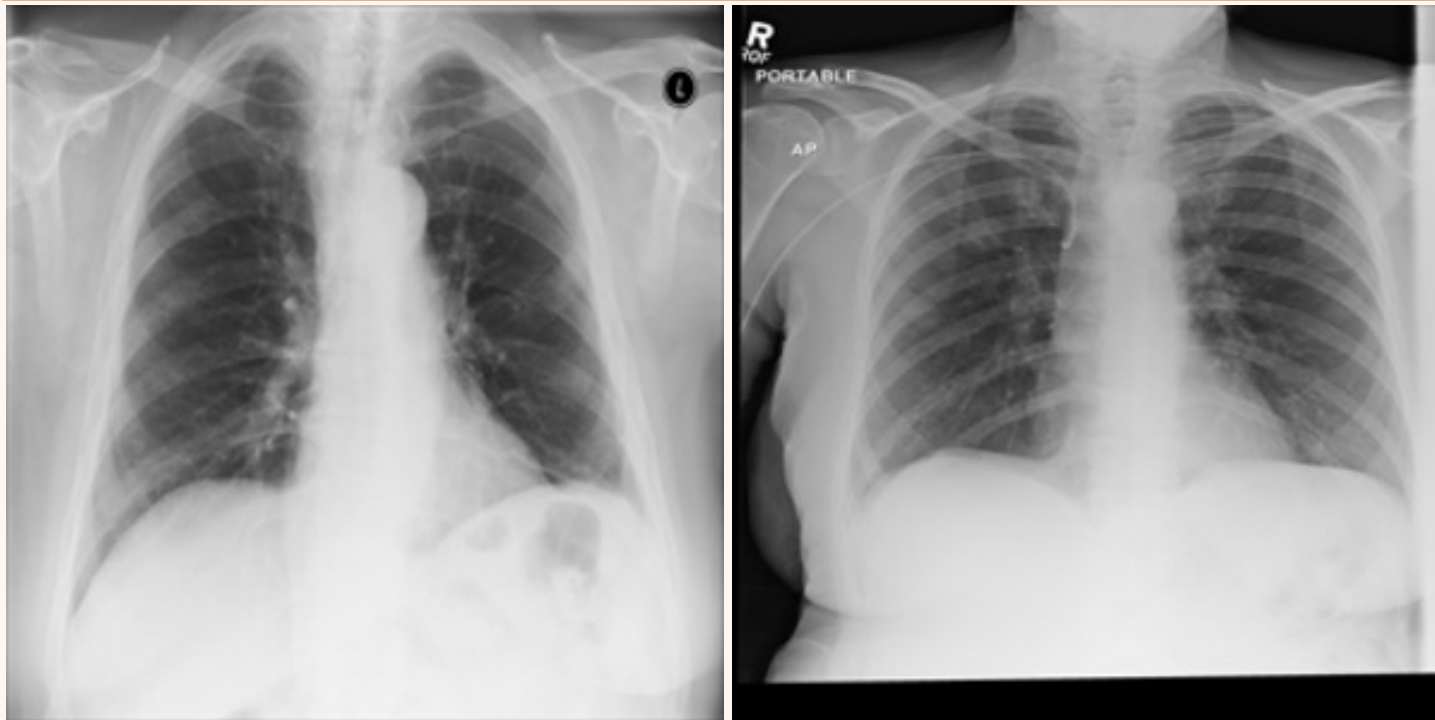


Images that are confusing to the algorithm seem to be edge cases in pathology



The pulmonologist says these look healthy, in agreement with algorithm, but the dataset has them labeled as COVID

In these confusing edge cases, the X-ray image is often retaken



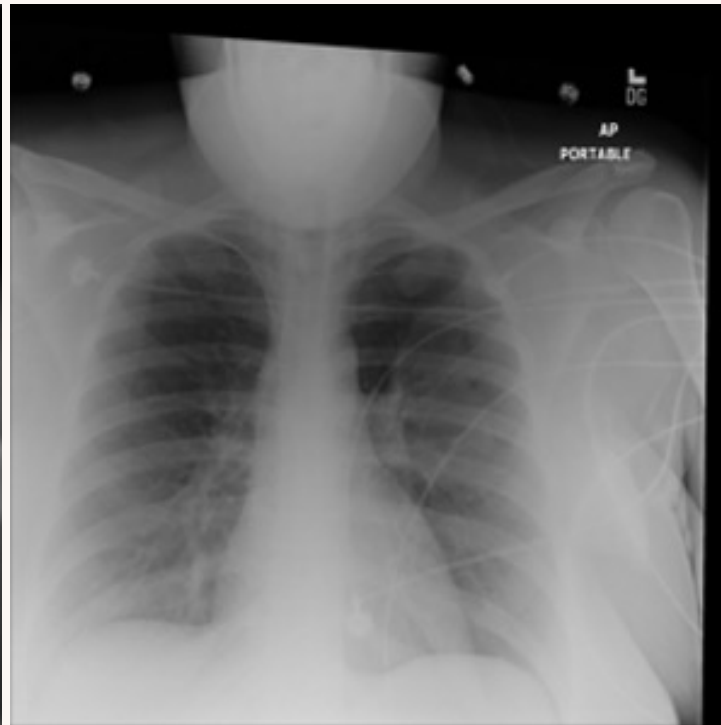
In practice, these would be retaken

These show some “ground glass opacity” and consolidation, common signs of COVID

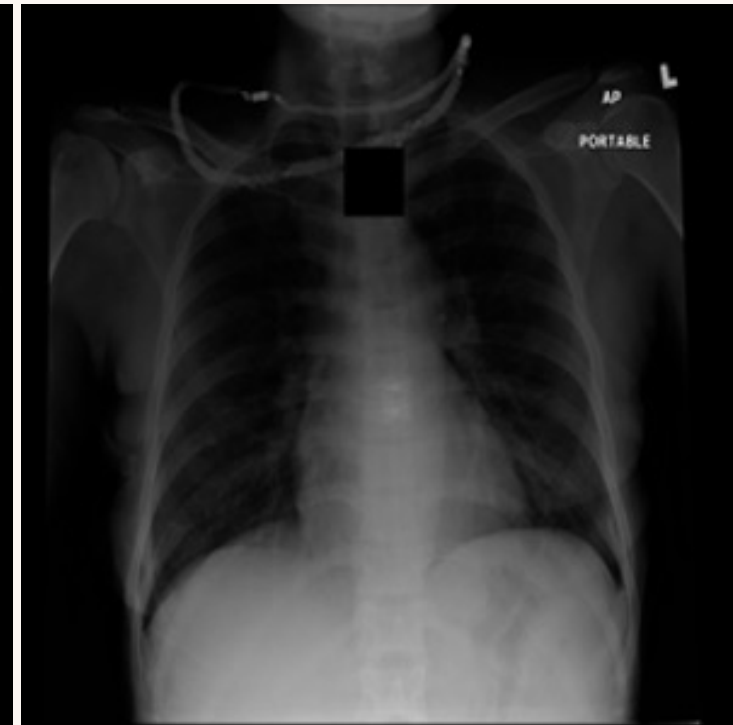
Sometimes, low quality images would be discarded altogether



Angle doesn't show much of the lungs



Off center

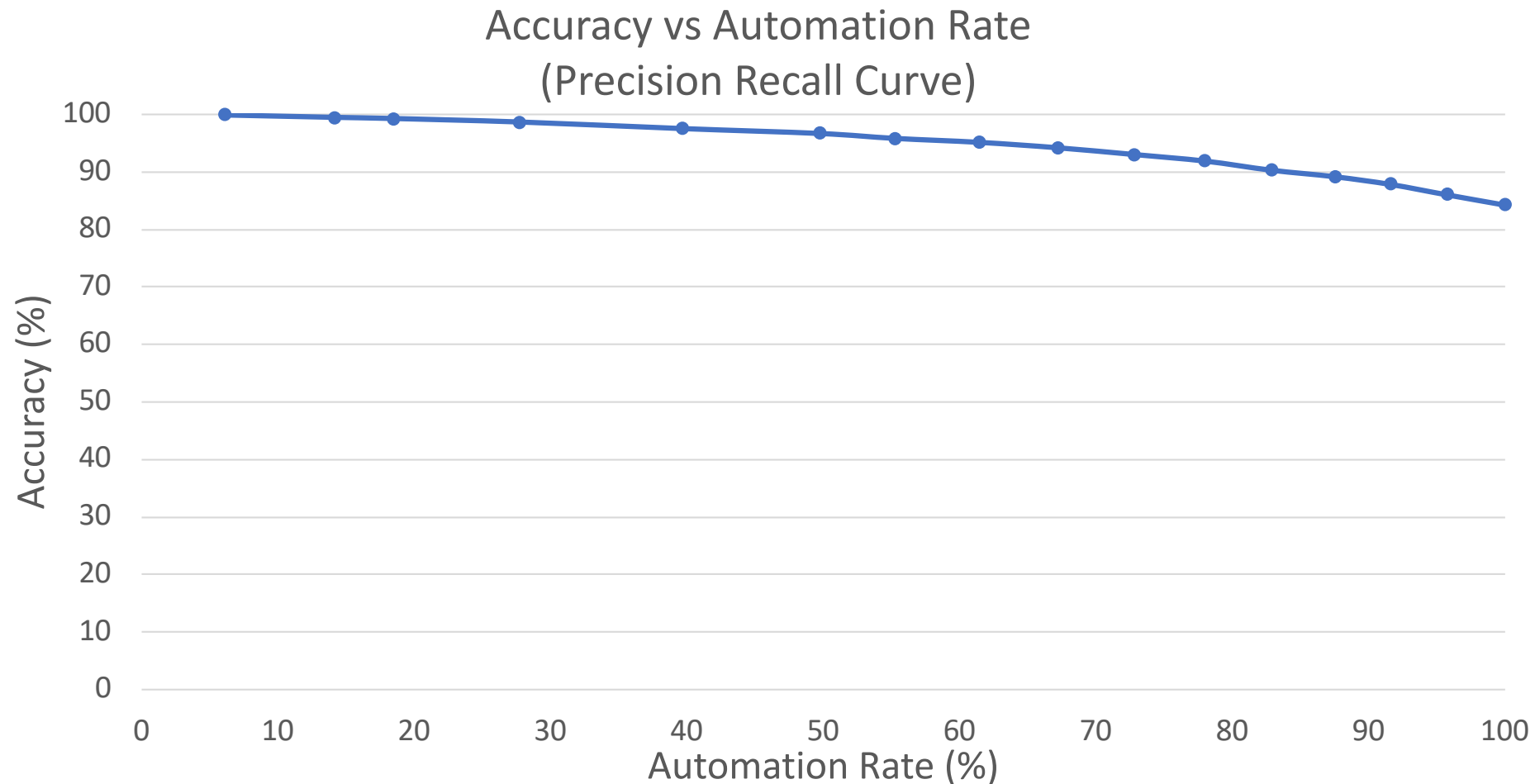


Too dark

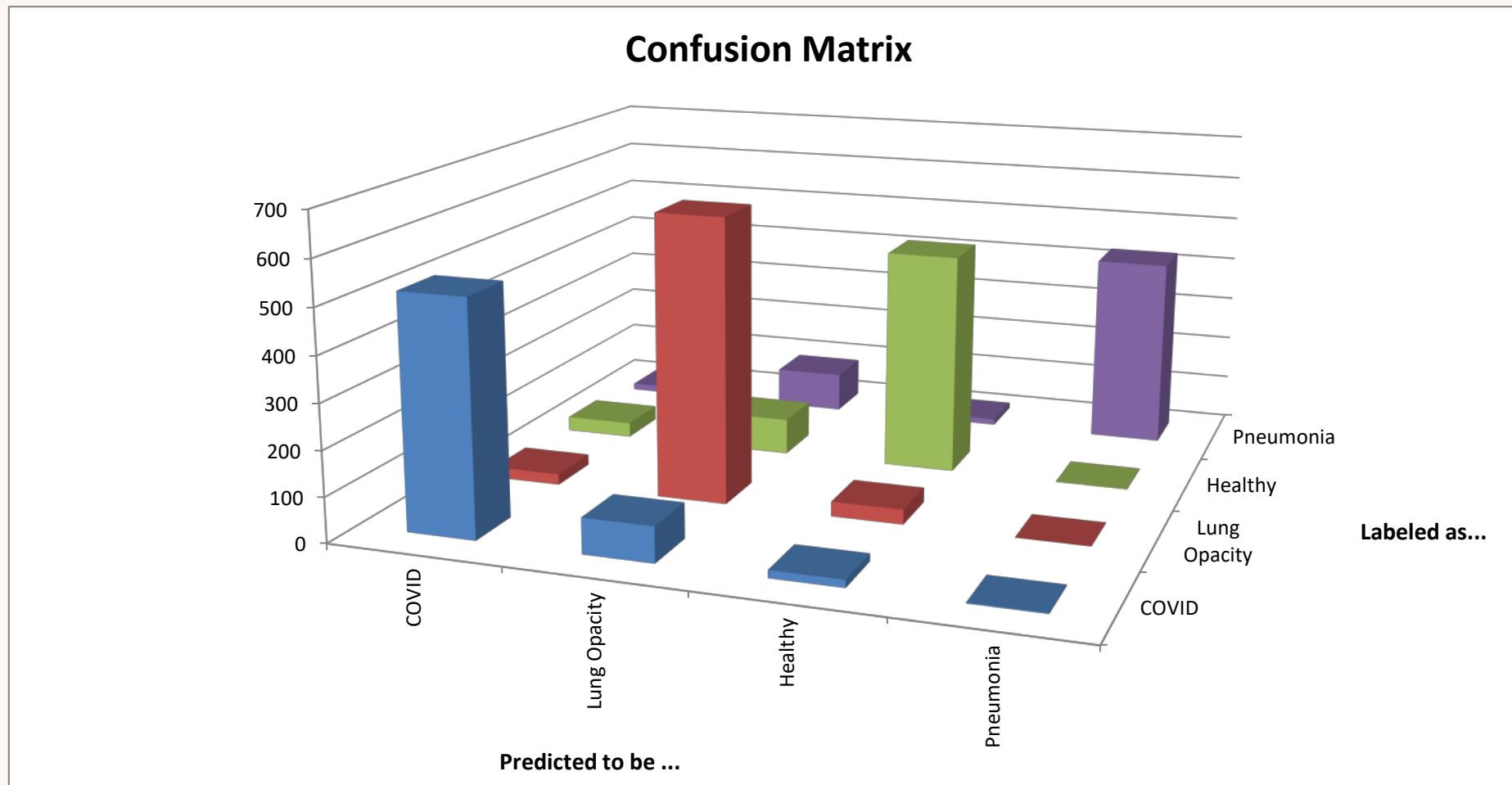
There is a lot of complexity in parsing COVID/non-COVID for X-rays because they are less sensitive than CT scans

- Portable vs in radiology lab, where the patient is in a different position (standing vs laying down in bed)
- Angle of X-ray beams
- Penetration, determined by calibration
- If they are inspiratory films (the patient is breathing in, since otherwise it's hard to see the base of the lungs)
- It's hard to read certain areas like behind the heart apex

When testing with all 4 categories, the more challenging goal results in slightly lower performance



The Confusion Matrix informs areas where further data refinement may be required



Next steps involve data refinement and ML training improvements

- Perform N-Fold Cross-Validation of original data set to identify suspect images
- Test hypotheses on sources of error
 - Angle / obstruction issues
 - Severity of pathology (e.g., early stage COVID may not be visible in X-rays)
- Refine approach to address sources of error
- Seek additional support for algorithm improvement, access to medical specialists
- ALTERNATIVELY, consider an approach that focuses on Ultrasound instead of X-Ray images to reach even more patients

Support & References

I would like to express my sincerest appreciation to

- Lali Devadas & Alex Henzinger, MIT PhD students and my PRIMES mentors this year
- Dr. Domingo Franco, Specialist in Critical Care Medicine, Pulmonary Disease, Henry Ford Health Systems

<https://iris.paho.org/handle/10665.2/34003?locale-attribute=es>

<https://iris.paho.org/handle/10665.2/55846>

<https://borgenproject.org/healthcare-in-paraguay/>

https://tigerprints.clemson.edu/all_theses/3886/