**FINAL PROJECT: SEGMENT\_1**

**PRESENTATION**:

**Selected Topic**:

In this study, we present the **prognostic/severity prediction model for COVID-19 from frontal chest X-ray/CT images of the patients**. The need to streamline patient management for COVID-19 has become more pressing than ever. Chest X-rays provide a non-invasive (potentially bedside) tool to monitor the progression of the disease. Such a tool can gauge severity of COVID-19 lung infections (and pneumonia in general) that can be used for escalation or de-escalation of care as well as monitoring treatment efficacy, especially in the ICU.

**Reason why selected topic:**

As the world grapples with COVID-19, every ounce of technological innovation and ingenuity harnessed to fight this pandemic brings us one step closer to overcoming it. Artificial intelligence and machine learning are playing a key role in better understanding and addressing the COVID-19 crisis. Machine learning technology enables computers to mimic human intelligence and ingest large volumes of data to quickly identify patterns and insights.

In the fight against COVID-19, organizations have been quick to apply their machine learning expertise in several areas: scaling customer communications, understanding how COVID-19 spreads, and speeding up research and treatment. In this scenario, it is both relevant and useful to understand **chest X-rays/CT images data to understand and develop an AI-based approach to predict and understand COVID-19.**

**Description of the source of data:**

This dataset is a database of COVID-19 cases with chest X-ray or CT images. It contains COVID-19 cases as well as [MERS](https://en.wikipedia.org/wiki/Middle_East_respiratory_syndrome), [SARS](https://en.wikipedia.org/wiki/Severe_acute_respiratory_syndrome), and [ARDS](https://en.wikipedia.org/wiki/Acute_respiratory_distress_syndrome). It can be found on Kaggle.com. It is one of the initial COVID-19 open image data collection. It was created by assembling medical images from websites and publications and currently contains 123 frontal view X-rays.

**Kaggle Link:**<https://www.kaggle.com/bachrr/covid-chest-xray>

**Kaggle Link:** <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>

**The primary site for the dataset**: <https://github.com/ieee8023/covid-chestxray-dataset>

Objective of this Kaggle Dataset was to build a public open dataset of chest X-ray and CT images of patients which are positive or suspected of COVID-19 or other viral and bacterial pneumonias ([MERS](https://en.wikipedia.org/wiki/Middle_East_respiratory_syndrome), [SARS](https://en.wikipedia.org/wiki/Severe_acute_respiratory_syndrome), and [ARDS](https://en.wikipedia.org/wiki/Acute_respiratory_distress_syndrome).). Data here is being collected from public sources as well as through indirect collection from hospitals and physicians. All images and data are being released publicly in this GitHub repo.

This project is approved by the University of Montreal's Ethics Committee #CERSES-20-058-D.

**Following is the list of each metadata field, with explanations**

* Patientid (internal identifier, just for this dataset)
* offset (number of days since the start of symptoms or hospitalization for each image, this is very important to have when there are multiple images for the same patient to track progression while being imaged. If a report says "after a few days" let's assume 5 days.)
* sex (M, F, or blank)
* age (age of the patient in years)
* finding (which pneumonia)
* survival (did they survive? Y or N)
* view (for example, PA, AP, or L for X-rays and Axial or Coronal for CT scans)
* modality (CT, X-ray, or something else)
* date (date the image was acquired)
* location (hospital name, city, state, country) importance from right to left.
* filename
* doi ([DOI](https://en.wikipedia.org/wiki/Digital_object_identifier) of the research article)
* url (URL of the paper or website where the image came from)
* license
* clinical notes (about the radiograph in particular, not just the patient)
* other notes (e.g. credit)

**Question we hope to be answered through data**:

Using input of chest X-ray and CT images, we intend to make the following Prognostic/severity predictions:

o Does the patient have Covid?

o Would the patient survive?

o Would the patient need intubation?

o Would the patient need supplemental oxygen?

This would give physicians an edge and allow them to act with more confidence while they wait for the analysis of a radiologist by having a digital second opinion confirm their assessment of a patient's condition. Also, these tools can provide quantitative scores to consider and use in studies.

**Tasks and Technologies used**:

·         STEP1\_ Data cleaning and analysis:

o    Panda and NumPy: We would be using these python libraries for cleaning the data before applying ML model as required.

O TorchXrayVision: this is a library of chest X-ray datasets and models including pre-trained models. This code is still under development.

·         STEP2: Saving cleaned data in MongoDB

o    MongoDB: We would use MongoDb to save cleaned data since our dataset has Chest X-ray/CT images.

o    Flask: We would be using integration with flask to display this data.

·         STEP3\_Applying Machine Learning Model

o    ScikitLearn lib: We would use these machine learning libraries to take the cleaned data from MongoDB as input and create classifiers, training /testing setup as required.

o    Imbalanced-learn lib: This library would be used as we have imbalanced data where the COVID-19 cases would be way more than [MERS](https://en.wikipedia.org/wiki/Middle_East_respiratory_syndrome), [SARS](https://en.wikipedia.org/wiki/Severe_acute_respiratory_syndrome), and [ARDS](https://en.wikipedia.org/wiki/Acute_respiratory_distress_syndrome).

o    We would figure out an algorithm/model for analysis based on decent accuracy score ad confusion matrix.

·         STEP4\_Dashboard: Tableau for visualization

o    We intend to use advanced features of tableau to develop effective storytelling and intuitive dashboards top present our analysis. There would be 2 sets of visualizations for Covid and other diseases analysis.

**To check SEQUENCE:**