MISSION: PREDICTABLE

X-RAY IMAGE CLASSIFICATION TO FIGHT **COVID-19**
AND **ANTIBIOTIC RESISTANCE**

BY TEAM 6 - ADELE GOLDBERG

YASAMIN ABBASZADEGAN
RAVI BANSAL
SOFIA PIGNATARO
MARIE-CLAIRE TRAORE

Image source: Pixabay

Meet the team

Yasamin Abbaszadegan
Machine Learning Engineer interested in the field of CV and NLP

Sofia Pignataro
Data Scientist passionate about improving people's lives with data

Ravi Bansal
Engineering graduate passionate to apply technologies for social good.

Marie-Claire Traore
Software Developer and Recent Information Technology Graduate

Special thanks for medical consultations:

Shreya Badhrinarayanan
Physician and MedTech Enthusiast

Problem statement

- In 2013, the World Health Organization (WHO) declared that **antibiotic resistance** is a global threat to public health.
- It is also an economic burden for society, with expenses on healthcare and pharmaceutical research.
- Currently, the number of deaths due to drug-resistant diseases is 700,000 each year around the globe.
- The World Bank alerted that these deaths would rise to 10 million by 2050, if no action is taken.

And COVID-19 may worsen the problem

- One key tool to tackle resistance is optimizing antibiotic use, **avoiding unnecessary prescription**.

- However, during the COVID-19 pandemic, many patients are receiving antibiotics without presenting signs of bacterial infection.

- A review of studies on COVID-19 patients identified that

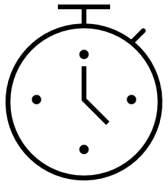
72% received antimicrobial therapy

but only

8% demonstrated superimposed bacterial or fungal co-infections.

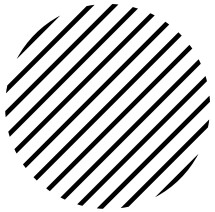
“I think, undoubtedly, we've seen more antibiotic use, because sometimes clinicians are unsure…and there really are no effective treatments for COVID-19”

statement by a medical director



“You're not going to wait 5 days to give a patient antibiotics, so patients were being treated for bacterial pneumonia just in case”

statement by a hospitalist



Statements’ source: CIDRAP (Center for Infectious Disease Research and Policy) News – University of Minnesota

Why is this happening?

COVID-19 increases the sense of **uncertainty** and **urgency** in health care services, since:

- There is no specific treatment.
- Its traits are difficult to distinguish from bacterial infection.
- Tests might not be readily available or have delayed results.
- Clinical symptoms can get worse quickly.

Available diagnosis tools

- While COVID-19 diagnosis can be made on clinical grounds, **chest imaging** can be used to assist the diagnosis, helping to identify or exclude pulmonary complications.
- It might also help to distinguish between a COVID-19 infection from a Bacterial Pneumonia.

"In COVID cases, you see very serious pathological signs. For example, **patchy shadowing in the lower chest area** is quite pathognomonic of COVID. Whereas Bacterial Pneumonia cases tend to **have consolidation or airspace shadowing**."

Shreya Badhrinarayanan - Full statement in the appendix

Source: https://doi.org/10.1016/j.cell.2018.02.010

- Previous studies have also shown successful applications in classifying bacterial versus viral pneumonia (a broader category for COVID-19 pneumonia).

Our analytical solution

- To solve the stated problem, we designed a tool for **healthcare professionals**.
- It is an easy-to-use app that provides additional information to assist the diagnosis. It shows the likelihood of patients presenting either COVID-19, Bacterial Pneumonia or Normal conditions, through a **deep learning model** for X-ray image classification.
- The app also contains information regarding antibiotic resistance (AR). It can be adapted according to local antimicrobial stewardship guidelines.

Workflow

1. Defining Problem
2. Data Preparation
3. Modeling
4. Hyperparameter Tuning
5. Deployment

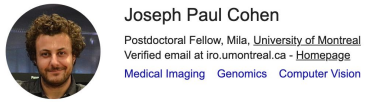
Data Collection

Two Resources were used to collect Covid19, Pneumonia Bacterial and Normal Chest X-ray Images

1- IEEE8023 GitHub Account

Author: Joseph Paul Cohen

https://github.com/ieee8023/covid-chestxray-dataset

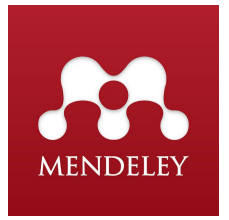


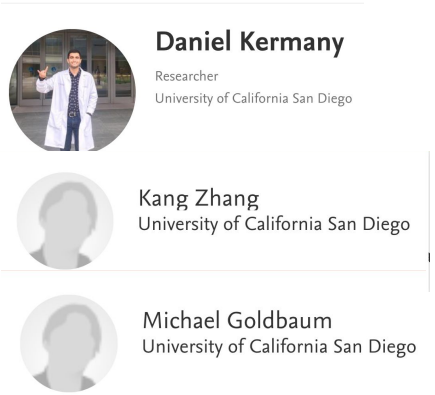
The **Covid X-Ray dataset** used is taken from an open source X-ray dataset. The dataset is created from images from research publications from China and other countries around the world.

2- Mendeley

Authors: Daneil Kermany, Kang Zhang, Michael Goldbaum

https://data.mendeley.com/datasets/rscbjbr9sj/3



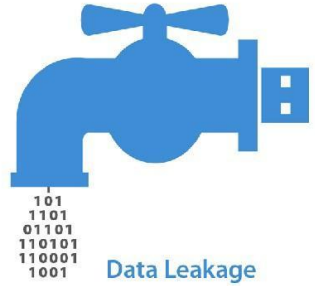


The **Pneumonia** and **Normal X-Ray dataset** used is taken from The Mendeley Public dataset.

Chest X-ray images were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center, Guangzhou. All chest X-ray imaging was performed as part of patients' routine clinical care.

Data Preparation

We fixed some problems during the data splitting.

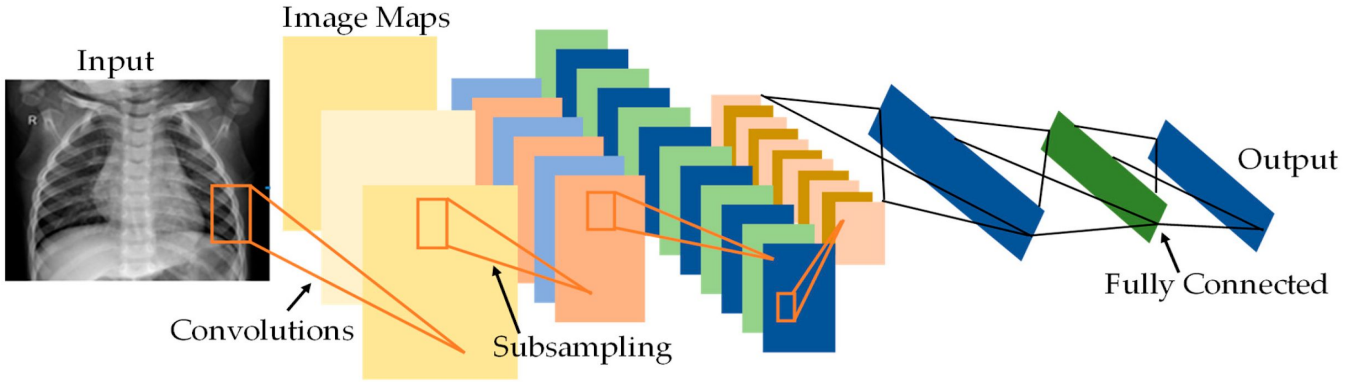


Data Leakage: In order to prevent over-optimistic test set performance, we made sure that there is no patient overlap.

Data Mismatch: In the COVID-19 data source, the files came from different websites. We used stratified random sampling to ensure that each dataset followed a similar distribution.

	COVID-19	Normal and Bacterial Pneumonia
Final number of images	479	490, for each class
Training	382 (80%)	392 (80%)
Validation	48 (10%)	49 (10%)
Test	47 (10%)	49 (10%)

Model Development



https://www.mdpi.com/2076-3417/10/9/3233/htm

- Used **Transfer Learning** and adapted a pretrained model to our X-ray image classification task.
- The model outputs class probabilities for 3 classes: **COVID-19, Normal and Bacterial Pneumonia**
- Used **DenseNet121** (*Densely Connected Convolutional Networks*) pretrained on **NIHCC X-Ray dataset.**
- Made the **last seven layers** trainable to fine tune on our dataset
- Added custom top layers
    - A **GlobalAveragePooling2D** layer to get the average of the last convolution layers from DenseNet121
    - A **Dense** layer of three outputs with **sigmoid activation** to get the prediction for each of our classes.

Model Training

- The model was generated using the **AWS SageMaker** service.
- Trained on ml.p2.xlarge instance with following hyperparameters:
  - Optimizer: Adam
  - Learning rate: 0.0001
  - Batch size: 32
  - Input Shape: (256,256,3)
  - Number of epochs: 25
- The values of hyperparameters were tuned to a separate holdout validation set.

Data Augmentations

- Data augmentations were used through Keras' ImageDataGenerator.
- It was used to reflect real world variations in X-ray data, in order to make the model generalize well.
- Augmentations were selected such that they do not change the original label.
- Augmentations used:
  - Adding rotation (0-3 degrees)
  - Adjusting brightness
  - Slightly shifting x-ray with respect to image (about 0.05 of image size)

Evaluation Metrics

Evaluation metrics needs to be carefully chosen in case of medical applications due to their high impact in decisions.

We looked at following metrics besides accuracy:

- **Sensitivity:** Probability that model predicts positive if a patient actually has the disease.
- **Specificity:** Probability that model predicts negative if a patient actually does not have the disease.
- **Positive Predictive Value (PPV):** Probability that the patient actually has the disease given the model predicted positive.
- **Negative Predictive Value (NPV):** Probability that patient is actually disease free given the model predicted negative.

Model Performance

- The model development was done to optimize the above mentioned metrics as these makes the algorithm more reliable.
- The following performance results were obtained with the final model:
  - Overall model accuracy on training set: 0.9989
  - Overall model accuracy on validation set: 0.9795
  - Overall model accuracy on training set: 0.9655
- Our model had a high level of accuracy for images not previously seen.

Model Performance

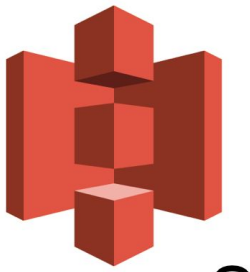
The detailed model performance on test set is being summarized in the following table:

	Prevalence	**Accuracy**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
COVID-19	0.324	1.000	1.000	1.000	1.000	1.000
NORMAL	0.338	0.966	1.000	0.948	0.907	1.000
BACTERIAL PNEUMONIA	0.338	0.966	0.898	1.000	1.000	0.950

The results show that the model is in good shape and able to generalize well on images that it has never seen before.

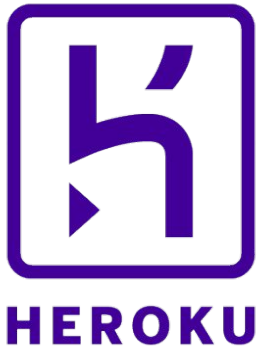
Modeling and Deployment Tools

Modeling



Amazon S3

WebAPI and Deployment



Streamlit

Uploaded and Shared

https://github.com/WWC-CovidHackathon

Final Application

- The deployed diagnostic application can be found at https://covid-wwc-hackathon.herokuapp.com/
- A demo of our application can be found the video below

Expected results

- We hope that our app helps healthcare professionals make **better informed decisions**, decreasing the already high degree of uncertainty in their workplace.

- We also hope that our tool helps professionals **not forget about antibiotic resistance threat** while fighting COVID-19.

Source: WHO Campaigns
World Antibiotic Awareness Week 2018

Appendix

Consultation with a medical physician

Shreya Badhrinarayanan
Physician and MedTech Enthusiast

“Pneumonia means an infection that is divided to bacterial and viral. Viral specifically could contain Covid which is a serious form of a viral pneumonia. In **Covid cases you see very serious pathological signs**, for example **patchy shadowing in the lower chest area** is quite Pathognomonic of covid. Whereas **Bacterial Pneumonia cases** tend to have **consolidation or airspace shadowing**.

It is important to note that x-rays are merely used to assist in our diagnosis rather than using it to confirm it.

For diagnosis of any patient that comes in with symptoms, medical practitioners use holistic approaches. For instance to confirm a diagnosis we look at series of things such as sputum samples which is a saliva sample, the history of a patient as well as if they have any inflammatory markers in their blood and other related tests to diagnose their condition.”

References

- COVID-19 Image Data Collection: Prospective Predictions Are the Future. Joseph Paul Cohen and Paul Morrison and Lan Dao and Karsten Roth and Tim Q Duong and Marzyeh Ghassemi. arXiv:2006.11988
- Dall, Chris. “COVID-19 presents antibiotic stewardship with challenges, opportunities.” *CIDRAP News*, May 19, 2020.
- Daniel S. Kermany, Michael Goldbaum, Wenjia Cai, Carolina C.S. Valentim, Huiying Liang, Sally L. Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, Justin Dong, Made K. Prasadha, Jacqueline Pei, Magdalene Y.L. Ting, Jie Zhu, Christina Li, Sierra Hewett, Jason Dong, Ian Ziyar, Alexander Shi, Runze Zhang, Lianghong Zheng, Rui Hou, William Shi, Xin Fu, Yaou Duan, Viet A.N. Huu, Cindy Wen, Edward D. Zhang, Charlotte L. Zhang, Oulan Li, Xiaobo Wang, Michael A. Singer, Xiaodong Sun, Jie Xu, Ali Tafreshi, M. Anthony Lewis, Huimin Xia, Kang Zhang. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell, Volume 172, Issue 5, 2018.
- Haileyesus Getahun, Ingrid Smith, Kavita Trivedi, Sarah Paulina & Hanan H Balkhyb. Tackling antimicrobial resistance in the COVID-19 pandemic. Bull World Health Organ 2020;98:442–442A. Published 2020.
- Huttner A, Harbarth S, Carlet J, et al. Antimicrobial resistance: a global view from the 2013 World Healthcare-Associated Infections Forum. *Antimicrob Resist Infect Control*. Published 2013 Nov 18.
- Interagency Coordination Group on Antimicrobial Resistance. No time to wait: securing the future from drug-resistant infections. Report to the secretary-general of the United Nations. Published 2019.
- Kermany D, Goldbaum M, Cai W et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell. 2018; 172(5):1122-1131. doi:10.1016/j.cell.2018.02.010.
- McKenna, Maryn. “Covid-19 May Worsen the Antibiotic Resistance Crisis.” *Wired*, April 23, 2020.
- World Bank. “Drug-Resistant Infections: A Threat to Our Economic Future.” Published 2017.

Project in Github: https://github.com/WWC-CovidHackathon/covid-detection