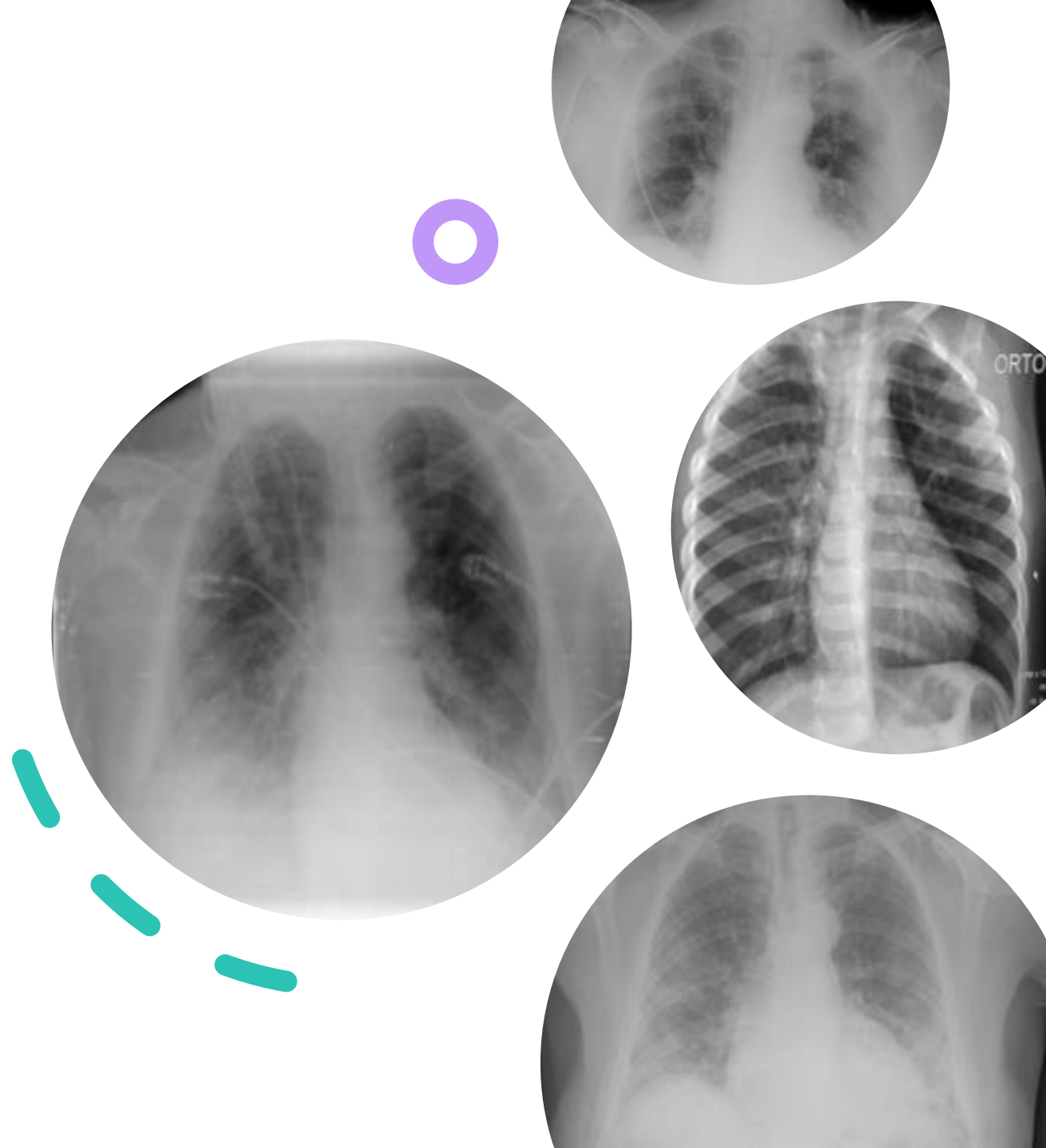


Imaging's Potential to Assist in COVID-19 Crisis

- CSIC 5011 / MATH 5473 Project 2 -

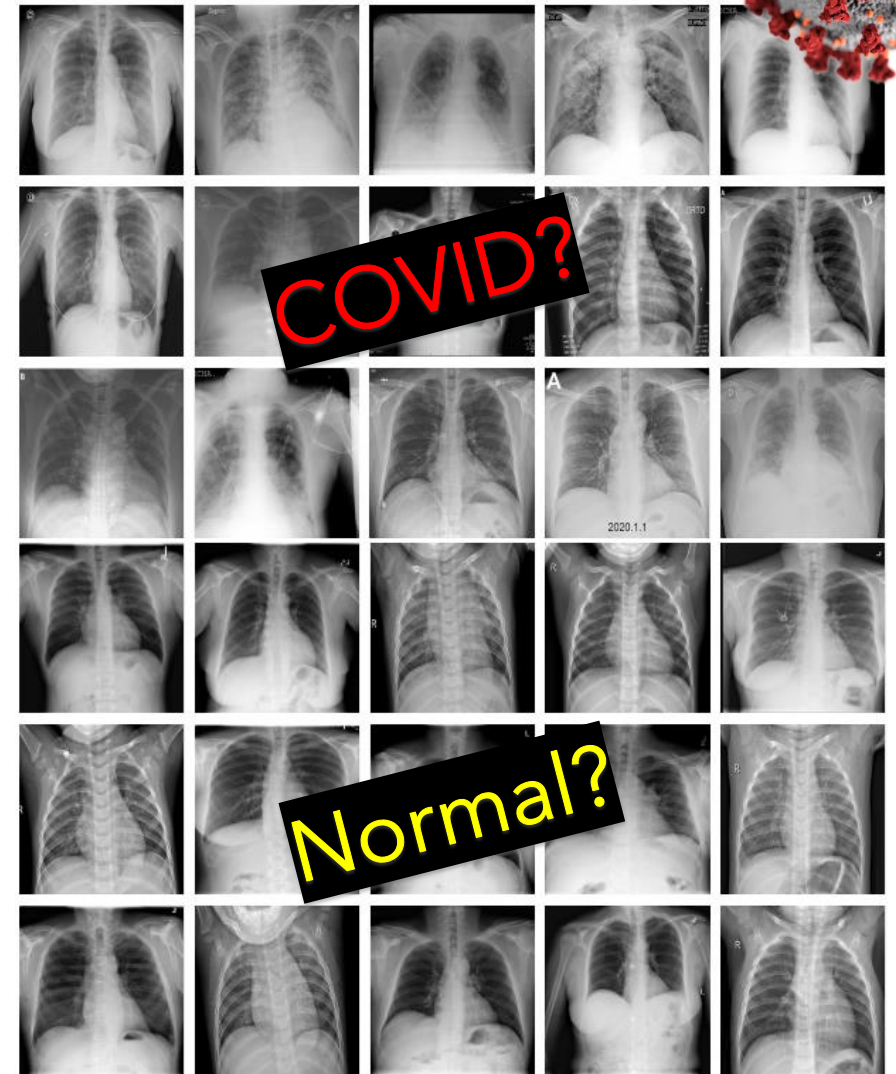
Yingshu CHEN¹, Wing Hei SUM², Ho Pan IP³, Zipeng WU⁴

Department of Computer Science and Engineering¹, Industrial Engineering and Decision Analytics², Mathematics³, Physics⁴



Motivation

- Severe COVID-19 pandemic
 - How to detect COVID without detection kit?
 - How to speed up detection?
- Chest X-ray Image Classification!
 - Issue: lack sufficient data
- Data augmentation via synthesis!



Outline

- Data and observation
- Data augmentation using generative model
- Data distribution
- Classification results

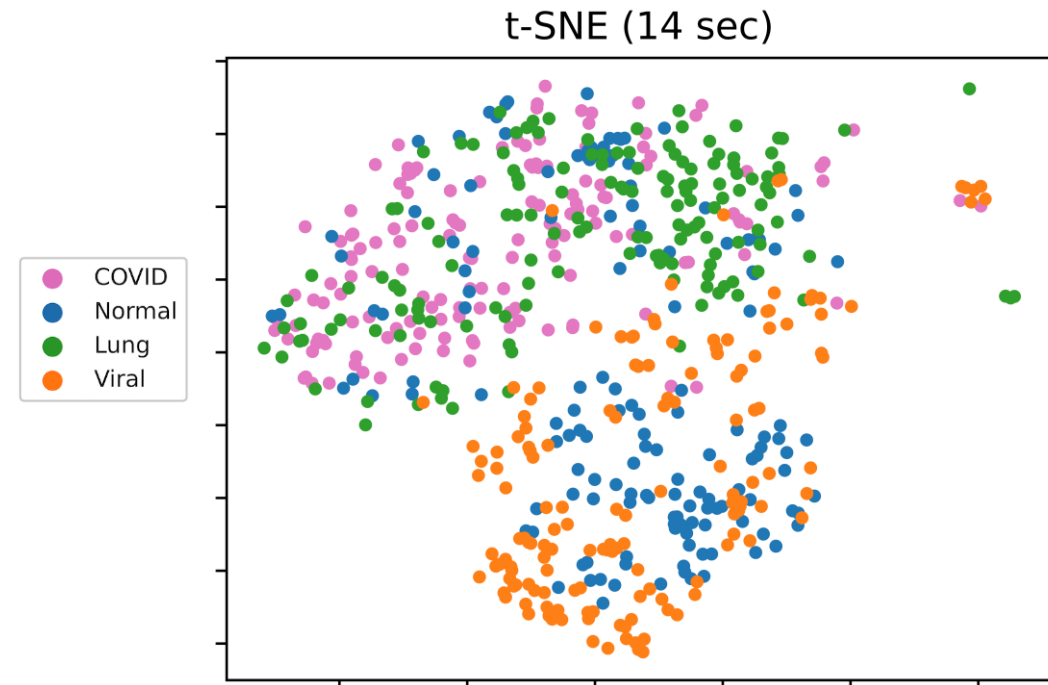
Data Observation

- Data* Statistics
 - Small original dataset
 - 20k images
 - Viral PNA only 1k
- Real Data visualization
 - Normal is scattered
 - Normal and Viral are close
 - COVID and Lung are close

Table 1: Data Statistics

Class	Normal	COVID-19	Lung Opacity ^a	Viral Pneumonia ^b
Train	9,942	3,366	5,762	1,095
Test	250	250	250	250
Total	10,192	3,616	6,012	1,345
Blended Train	9,942	7,000	5,762	7,000

^{a,b} Non-COVID lung infection.



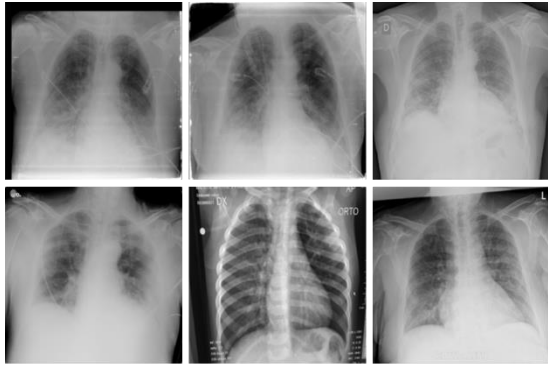
*Data source: <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>

Data Augmentation

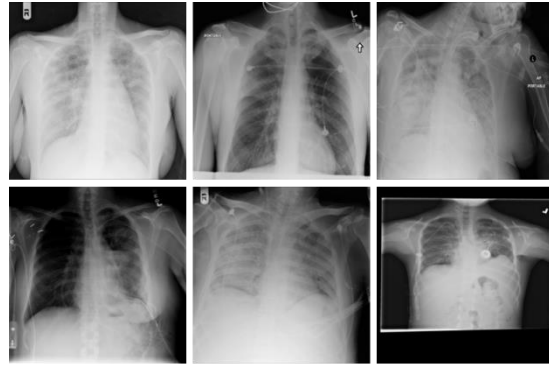
- Generative Model - SOTA conditional GAN
 - *stylegan2-ada* model with limited data (thousand scale)
 - Photorealistic synthesis
 - Unconditional vs Conditional GAN

Data Augmentation

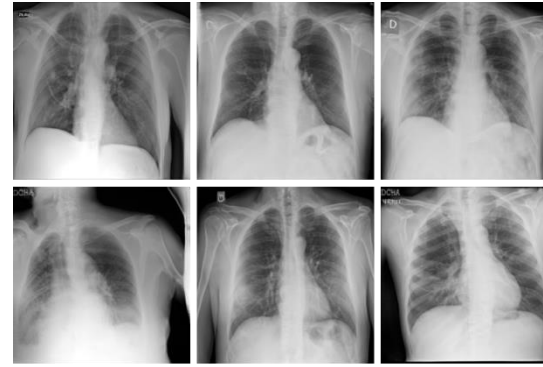
- Photorealism of Synthesis



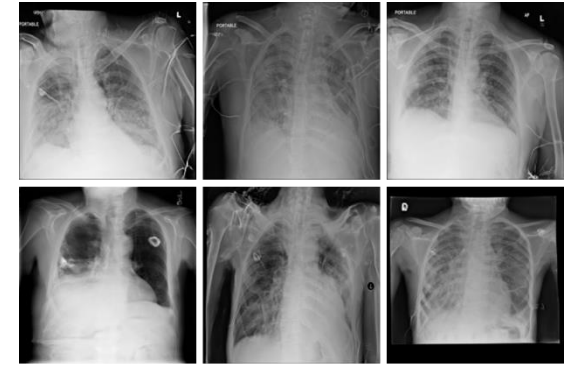
Real COVID-19 CXRs



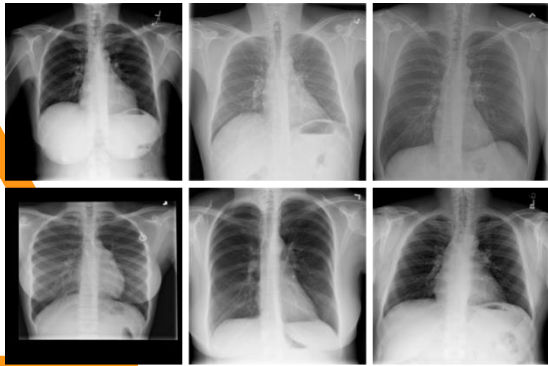
Real Lung Opacity CXRs



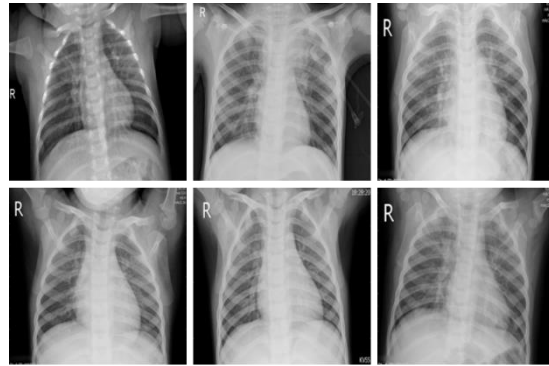
Synthetic COVID-19 CXRs



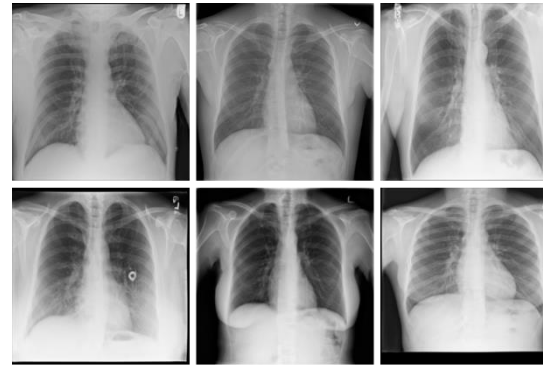
Synthetic Lung Opacity CXRs



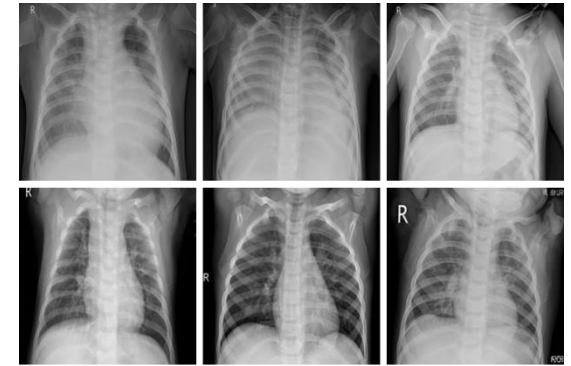
Real Normal CXRs



Real Viral PNA CXRs



Synthetic Normal CXRs



Synthetic Viral PNA CXRs

Data Augmentation

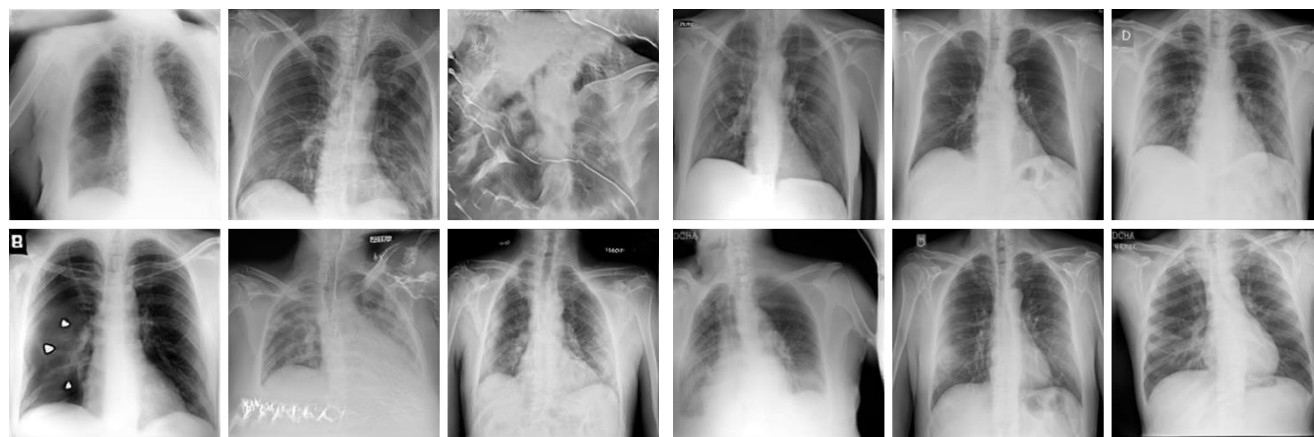
- Generative Model
 - SOTA conditional GAN
 - *stylegan2-ada* model with limited data (thousand scale)
 - Photorealistic synthesis
 - Unconditional vs Conditional GAN
 - Expected to assist in data augmentation

Table 3: Unconditional vs. Conditional Generative Model

Model*	Unconditional				Conditional			
	Normal	COVID-19	Lung Opacity	Viral PNA	Normal	COVID-19	Lung Opacity	Viral PNA
FID↓	18.28	33.48	28.74	28.25	16.39	25.24	24.22	26.85

*Four unconditional models and one conditional model with 4 labels

Lower FID means closer to real data distribution

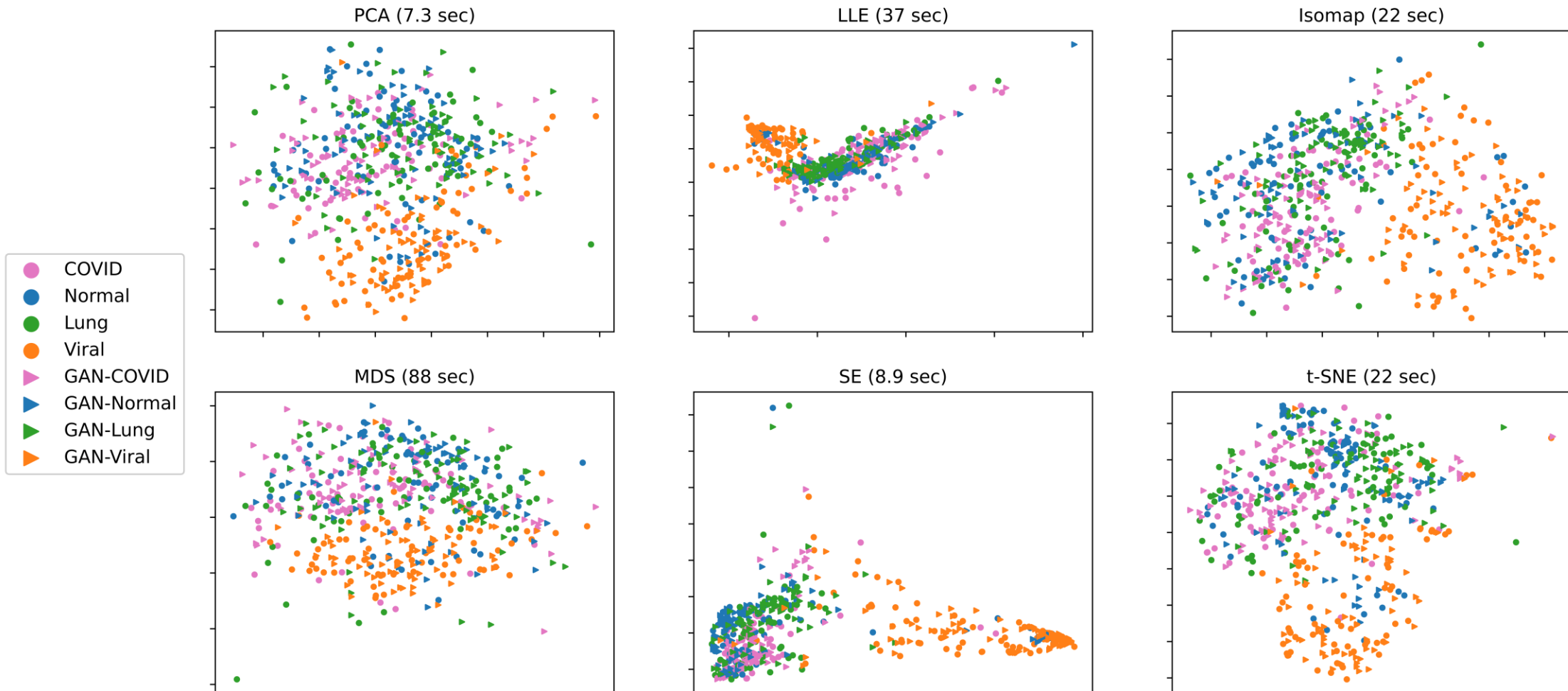


Unconditional - COVID-19

Conditional - COVID-19

Blended Data Distribution Visualization

- Generative model to some extent reproduce real data distribution



Classification

- Method description
 - Convolutional Neural Network (CNN)
 - Layers: 13
 - Architecture: Conv2D, Dense, Dropout, MaxPooling2D, Flatten
 - Activation functions: ReLU and Softmax
 - Early stopping to avoid over-fitting

Table 2: Hyper-parameter settings

Optimizer	Loss	Metrics	Batch Size	Epochs
Adam	sparse_categorical_crossentropy	Accuracy	32	20

• Blended Data

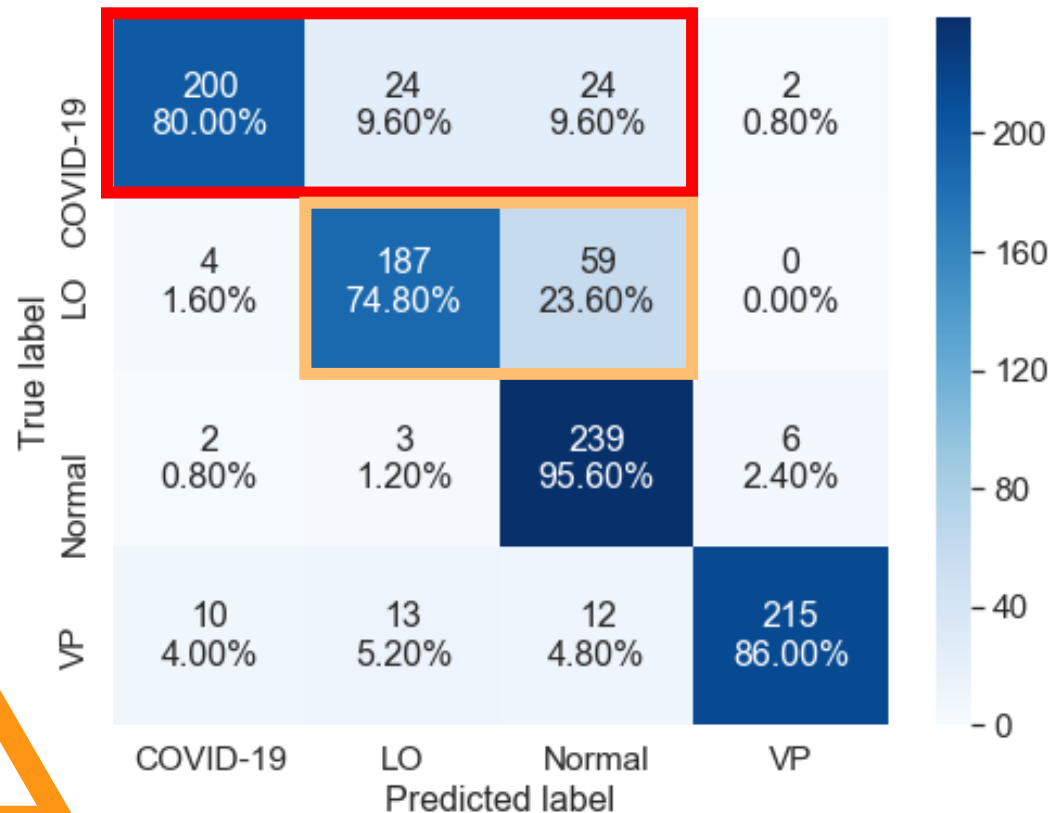
Table 5: Experiments on Data Combination

Data	Validation Acc ^a	Prediction Acc per Class ^b			
		COVID	Lung	Normal	Viral
Real only	86.41%	92.4%	75.2%	82.4%	92.0%
Real + 3,634 COVID synthesis	88.04%	90.4%	82.4%	89.2%	95.2%
Real + 1,238 Lung synthesis	86.05%	88.8%	81.2%	88.8%	90.4%
Real + 5,905 Viral synthesis	89.09%	94.4%	80.04%	87.6%	94.8%
Real + 3,634 COVID & 5,905 Viral synthesis	90.35%	95.2%	80.0%	91.6%	96.4%
Real + 3,634 COVID&1,238 Lung&5,905 Viral synthesis	89.35%	94.4%	81.2%	89.2%	97.2%

^{a, b} Different from Sec.4.1, here we use $\#correct/\#prediction$ in validation set^a and testing set^b as accuracy.

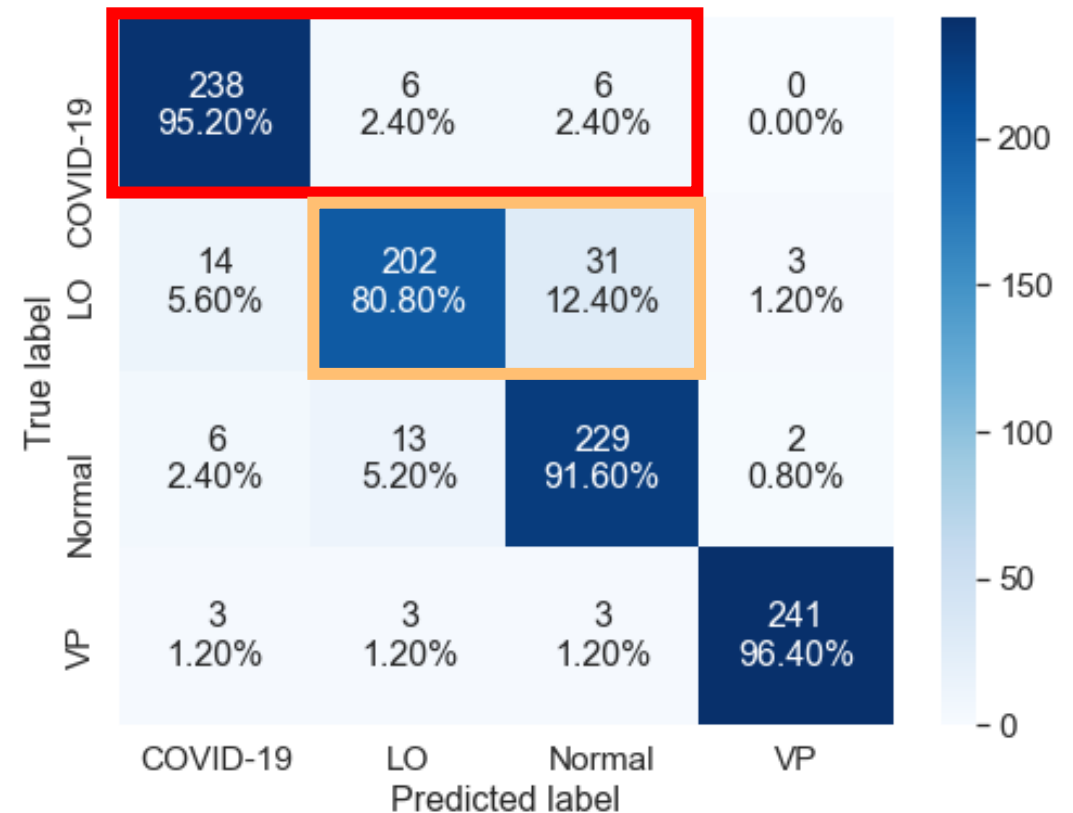
Evaluation on the Classification

Real Data



Overall Accuracy: 84.1%

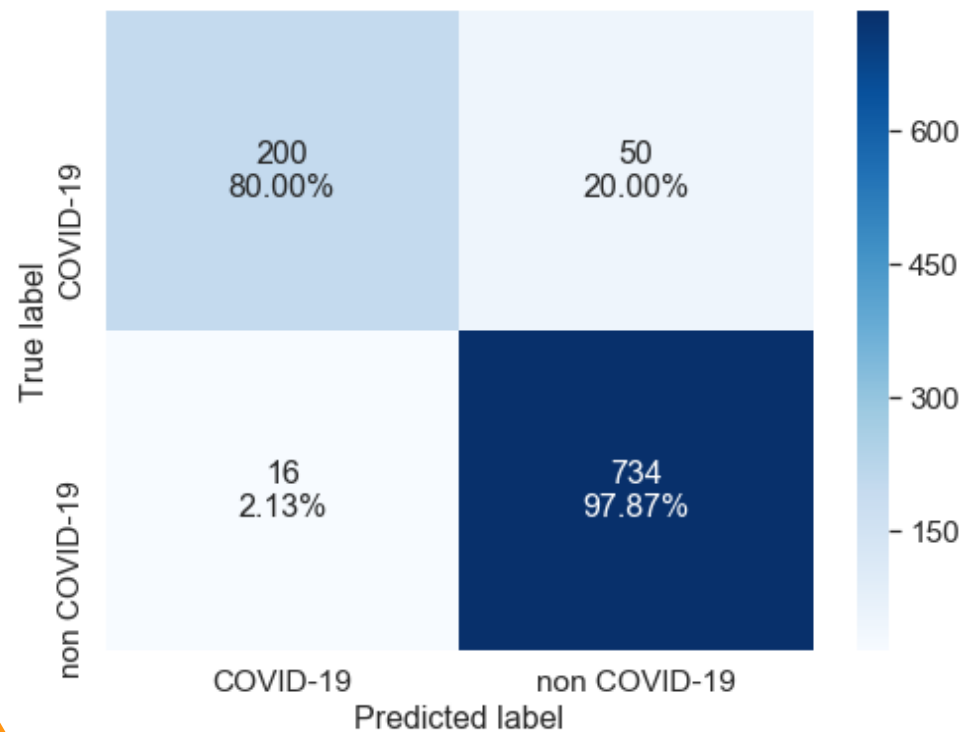
Real Data + Synthetic Data (COVID-19 & Viral)



Overall Accuracy: 91%

Evaluation on the Classification

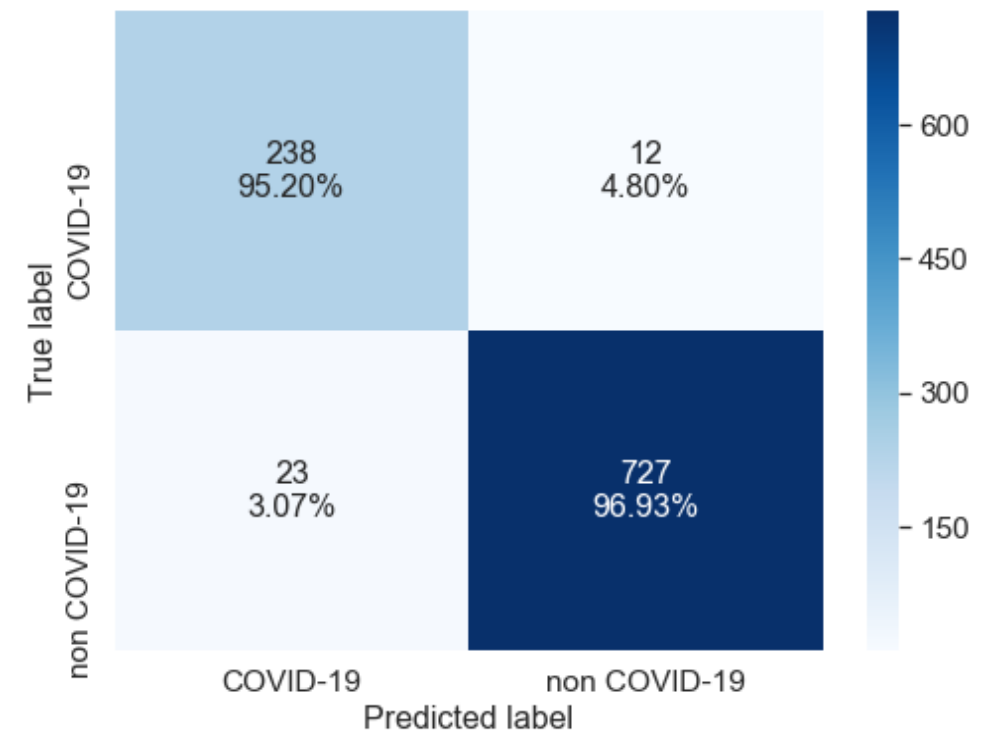
Real Data



Accuracy=0.934
Precision=0.936
Recall=0.979
F1 Score=0.957



Real Data + Synthetic Data (COVID-19 & Viral)



Accuracy=0.965
Precision=0.984
Recall=0.969
F1 Score=0.976

Summary

- Dimensionality reduction methods can work as a visualization and evaluation means to visualize data distribution and get data distribution observation.
- High-quality chest X-ray images synthesis with limited data can be synthesized via a conditional GAN.
- We transfer the COVID-19 CXR detection problem to a 4-class classification problem considering other non-COVID lung infectious CXRs as well.
- We validate the effectiveness of synthetic images for medical data augmentation and visual classification enhancement.
- Qualitatively and quantitatively verify the potential of image generative model to assist in new respiratory infectious disease detection.