COVID-19 and Pneumonia Chest X-Ray Classifier Web Application

Ananto J., Brandon S. B., Fernando L. M., William M., Wilson W.

Background

Problems

- COVID-19 Pandemic
- Importance of COVID-19 testing
- Limited access of RT-PCR
- **Alleged** manipulation of COVID-19 test results

Solutions

- Utilize chest x-rays (price)
- Use Deep Learning and Computer Vision (accuracy)
- Deploy on a web application (accessibility)

Related Works

Feasibility of COVID-19 CXR Classification

- Radiography experts are able to distinguish features in COVID-19 infected chest x-rays (Jacobi et al., 2020)
- Radiologist recommends the use of neural networks to improve the robustness of COVID-19 detection (Sarvamangala & Kulkarni, 2020)

Dataset

- J.P. Cohen
- GAN-based dataset by Loey et al.
- SIRM, BIMCV-COVID19+

Studies using pre-trained CNNs

- Chowdhury et al. 3-class classification with 99% accuracy, precision, sensitivity
- Jain et al. used Xception, ResNeXt, Inception V3

CNN from scratch

- Yoo et al. ResNet18 + decision tree
- COVID-Net by Wang et al. trained using COVIDx dataset
- Khan et al. modified Xception to create CoroNet

CXR Enhancements

- Degerli et al. compiled the largest dataset with 119316 CXR images to a pre-trained CNN model
- Qi et al. and Rahman et al. use enhanced CXR images are able to boost the performance of CNN’s when detecting COVID-19 infected lungs.

Methodology

Product Backlog

Role	Task	Sprint	Size	Priority	Status
Visitor	Register	1	5	Medium	Finished
User	Login	1	5	Medium	Finished
User	Upload Image	2	10	High	Finished
User	Classify Chest X-Ray Images	2	20	High	Finished
User	Find Nearest Hospitals	3	8	Medium	Finished
User	Learn More About the Classification Algorithm	3	3	Low	Finished
User	Get Next Steps and Suggestions	3	3	Medium	Finished

Effort-Hour Capacity

Person	Days Available	Days for Other Scrum Activities	Hours per Day	Available Effort-Hours
Ananto	9	2	4-7	28-49
Brandon	7	2	4-6	20-30
Fernando	7	2	5-6	25-30
William	8	2	4-7	24-42
Wilson	9	2	4-6	28-42
Total				125-193

Data Preparation

- Used the COVID-19 Radiography Database
  - SIRM COVID-19 Database
  - BIMCV-COVID19+ dataset
- Dataset Contents
  - 10192 normal CXR
  - 3616 COVID-19 CXR
  - Lung Opacity & Pneumonia CXR were excluded
- Split the data with a set seed
  - 80% training (8154 normal, 2892 COVID)
  - 10% validation (1019 normal, 362 COVID)
  - 10% test (1019 normal, 362 COVID)

Data Preparation (cont.)

- Preprocessing or Transformations
  - scaled the images down to 224 x 224 pixels
  - 0.2 shear range
  - 0.2 zoom range
  - rotation up to 10°
- We did not resample to balance the dataset, since it did not pose significant problems

Model Selection

- VGGNets: VGG-16, VGG-19
- ResNets: ResNet50, ResNet101, ResNet152
- Inception: InceptionV3, InceptionResNetV2
- DenseNets: DenseNet121, DenseNet169, DenseNet201
- Xception

Training Pipeline

- Unfreeze all layers
- Concatenated Layers
  - Global Average Pooling layer
  - 2-unit Dense output layer with sigmoid activation function
- Optimizer: Adam
- Loss Function: binary cross-entropy
- Learning rate schedule: Exponential Decay
- Implemented Early stopping with 10 epoch patience

Training Pipeline (cont.)

- Save the best model
- Evaluate the model using the test set
- Record validation and test accuracy precision recall
- Repeat for all the batch size and learning rate hyperparameters
  - Batch size 16, Learning rate 1e-4
  - Batch size 32, Learning rate 1e-4
  - Batch size 16, Learning rate 1e-6
  - Batch size 32, Learning rate 1e-6

Results

Results (1/4)

- In general, all of the models perform well in this classification task.
- ResNet101 performs outstandingly well in the classification task.
- It doesn't seem to overfit, as observed from the test set accuracy.

TABLE I
BATCH SIZE 32, LEARNING RATE 1E-4

Model Name	Num of Epochs	Validation			Test		
		Acc	*Prec*	*Rec*	*Acc*	*Prec*	*Rec*
ResNet50	29	99.64	99.64	99.64	99.28	99.28	99.28
ResNet101	39	**100.0**	**100.0**	**100.0**	99.06	99.06	99.06
ResNet152	41	99.86	99.86	99.86	99.42	99.42	99.42
Xception	21	99.78	99.78	99.78	99.35	99.35	99.35
VGG16	28	99.13	99.2	99.13	99.20	99.28	99.20
VGG19	22	99.13	99.13	99.13	98.41	98.41	98.41
InceptionV3	40	99.71	99.71	99.71	99.57	99.57	99.57
InceptionResNetV2	25	99.57	99.49	99.57	99.42	99.35	99.49
DenseNet121	31	99.71	99.71	99.71	**99.64**	**99.64**	**99.64**
DenseNet169	25	99.64	99.64	99.64	99.49	99.57	99.49
DenseNet201	39	99.71	99.71	99.78	99.49	99.49	99.49

**The best result of each metric is bolded*

Results (2/4)

- Changes in batch size also provided no visible difference in performance.
- Here we see that ResNet101 achieved the best results yet again.
- However, notice how the smaller architectures also have exceptional results.

TABLE II
BATCH SIZE 16, LEARNING RATE 1E-4

Model Name	Num of Epochs	Validation			Test		
		Acc	*Prec*	*Rec*	*Acc*	*Prec*	*Rec*
ResNet50	42	99.71	99.71	99.71	99.71	99.71	99.78
ResNet101	50	**99.78**	**99.78**	**99.78**	**99.93**	**99.93**	**99.93**
ResNet152	19	98.48	98.55	98.48	92.83	92.71	92.98
Xception	20	99.64	99.64	99.64	99.57	99.57	99.57
VGG16	39	99.49	99.49	99.49	99.13	99.13	99.20
VGG19	36	99.57	99.57	99,57	99.28	99.28	99.28
InceptionV3	34	**99.78**	**99.78**	**99.78**	99.35	99.35	99.35
InceptionResNetV2	32	99.64	99.64	99.64	99.35	99.35	99.35
DenseNet121	30	99.49	99.49	99.49	99.64	99.64	99.64
DenseNet169	27	99.67	99.67	99.67	98.19	98.19	98.19
DenseNet201	43	99.64	99.64	99.64	99.49	99.42	99.49

**The best result of each metric is bolded*

Results (3/4)

- Performance dips marginally when we decrease learning rate from 1e-4 to 1e-6.
- Also notice how ResNets always seem to have a slight edge over other architecture family.

TABLE III
BATCH SIZE 32, LEARNING RATE 1E-6

Model Name	Num of Epochs	Validation			Test		
		Acc	*Prec*	*Rec*	*Acc*	*Prec*	*Rec*
ResNet50	36	97.97	97.90	97.97	**98.19**	**98.19**	**98.19**
ResNet101	31	98.26	97.29	98.33	97.03	97.03	96.89
ResNet152	27	**98.41**	**98.19**	**98.41**	97.83	97.83	97.90
Xception	48	95.08	94.79	94.93	93.64	93.64	93.64
VGG16	47	97.18	97,17	96.89	96.52	96.66	96.52
VGG19	38	96.67	96.74	96.74	95.94	96.02	96.02
InceptionV3	38	95.15	95.08	95.08	94.86	94.80	95.08
InceptionResNetV2	36	95.22	95.01	95.08	93.56	93.50	93.70
DenseNet121	33	96.74	96.87	96.45	96.60	96.95	96.52
DenseNet169	47	97.61	97.54	97.47	97.25	97.60	97.10
DenseNet201	46	96.52	97.12	97.54	97.97	97.97	98.04

**The best result of each metric is bolded*

Results (4/4)

- Here, using a smaller batch size seems to help.
- However, 1e-4 still seems to be the more appropriate learning rate.

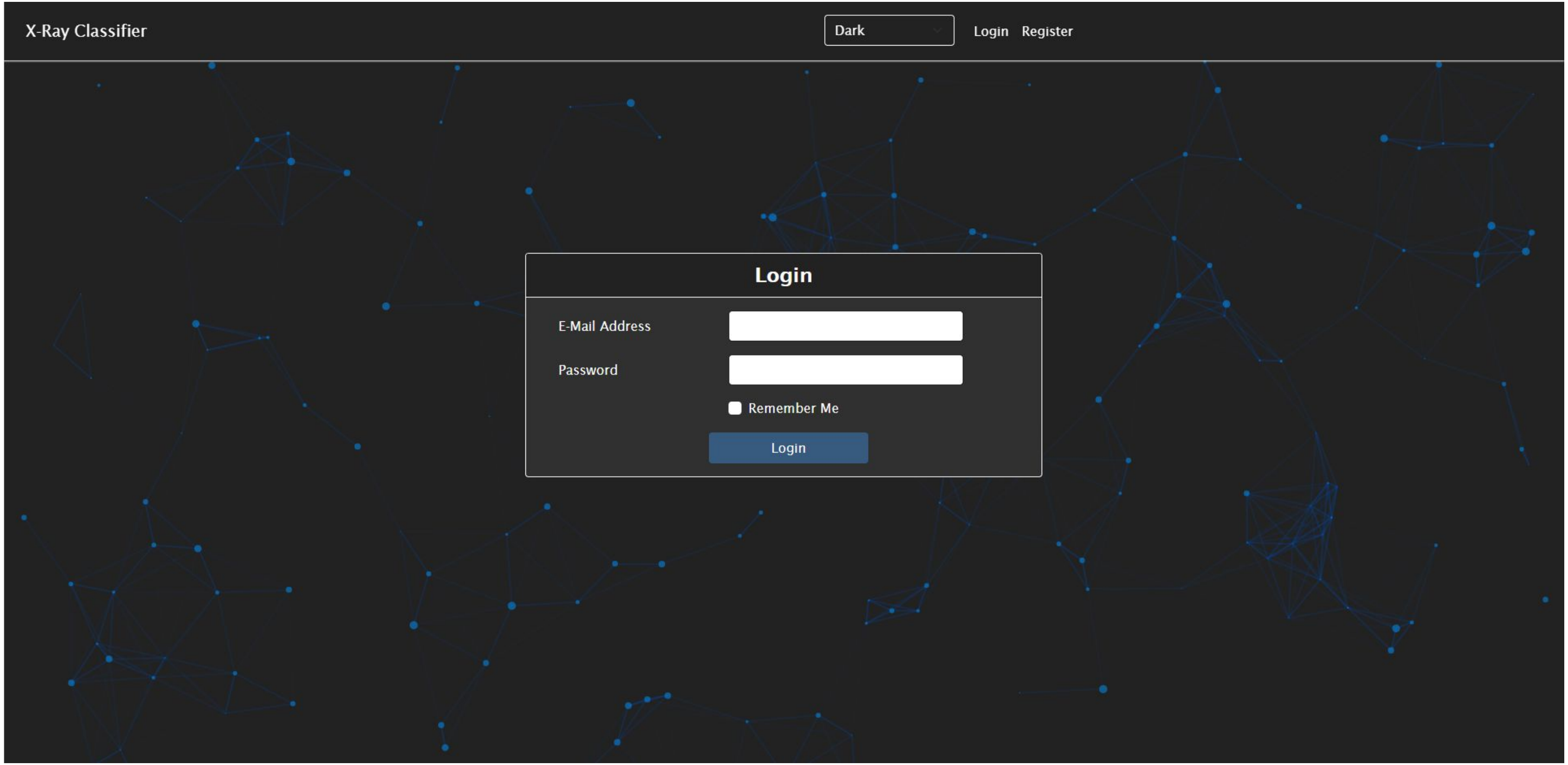
TABLE IV
BATCH SIZE 16, LEARNING RATE 1E-6

Model Name	Num of Epochs	Validation			Test		
		Acc	*Prec*	*Rec*	*Acc*	*Prec*	*Rec*
ResNet50	36	98.48	98.40	98.26	97.90	97.76	97.97
ResNet101	39	**98.77**	**98.48**	**98.70**	98.12	98.05	**98.26**
ResNet152	34	98.55	98.34	98.48	**98.19**	**98.11**	97.97
Xception	43	97.1	96.76	97.25	96.02	95.95	96.02
VGG16	48	97.47	97.25	97.25	96.67	96.53	96.81
VGG19	43	97.61	97.68	97.47	97.25	97.10	97.03
InceptionV3	29	95.73	95.59	95.73	94.13	94.19	93.99
InceptionResNetV2	50	96.16	95.96	96.23	94.71	94.92	94.64
DenseNet121	29	97.18	96.83	97.25	97.61	97.53	97.39
DenseNet169	27	97.54	97.39	97.18	97.83	97.68	97.76
DenseNet201	32	98.55	98.41	98.48	98.04	98.04	98.04

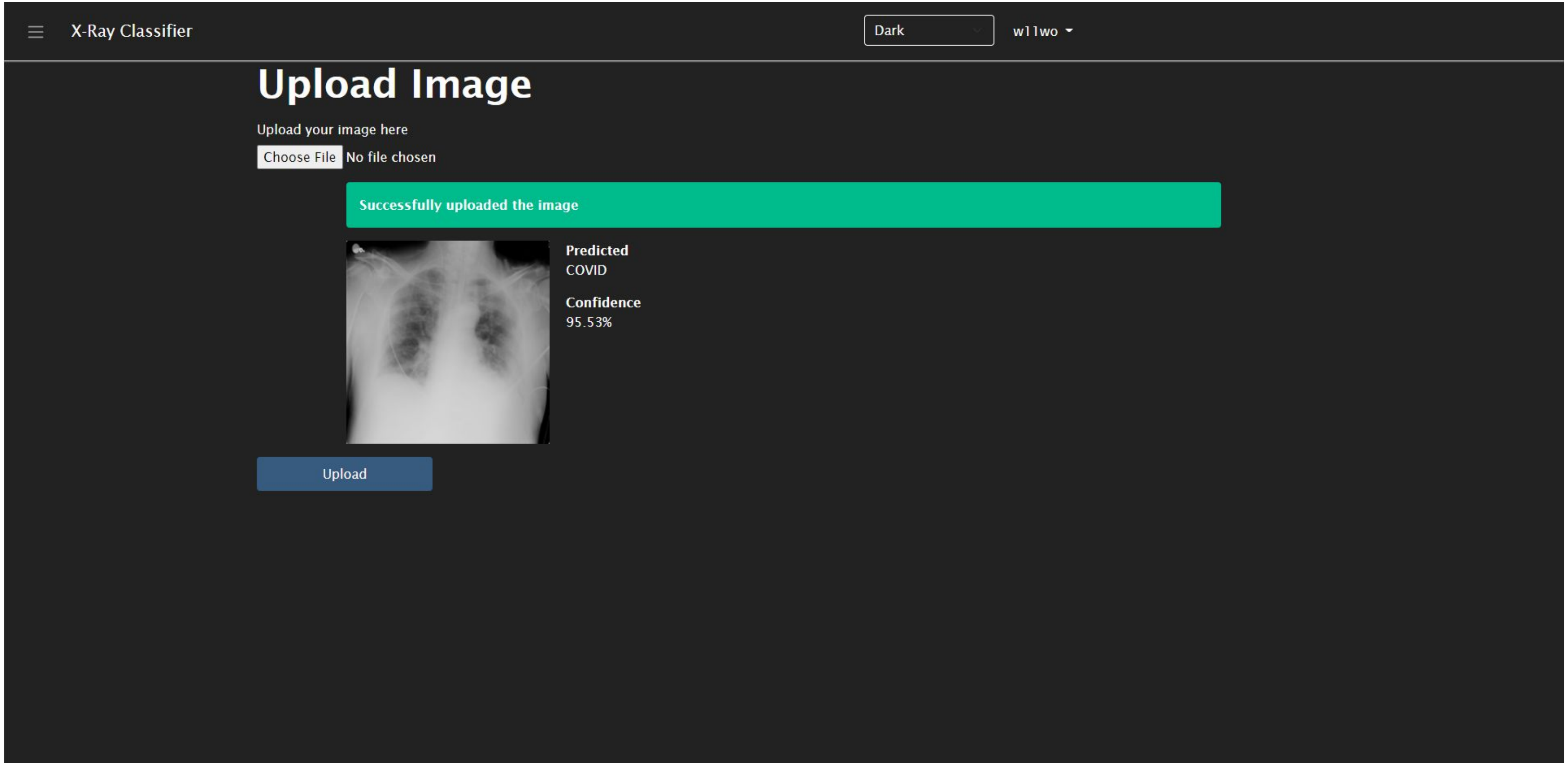
**The best result of each metric is bolded*

Deployment

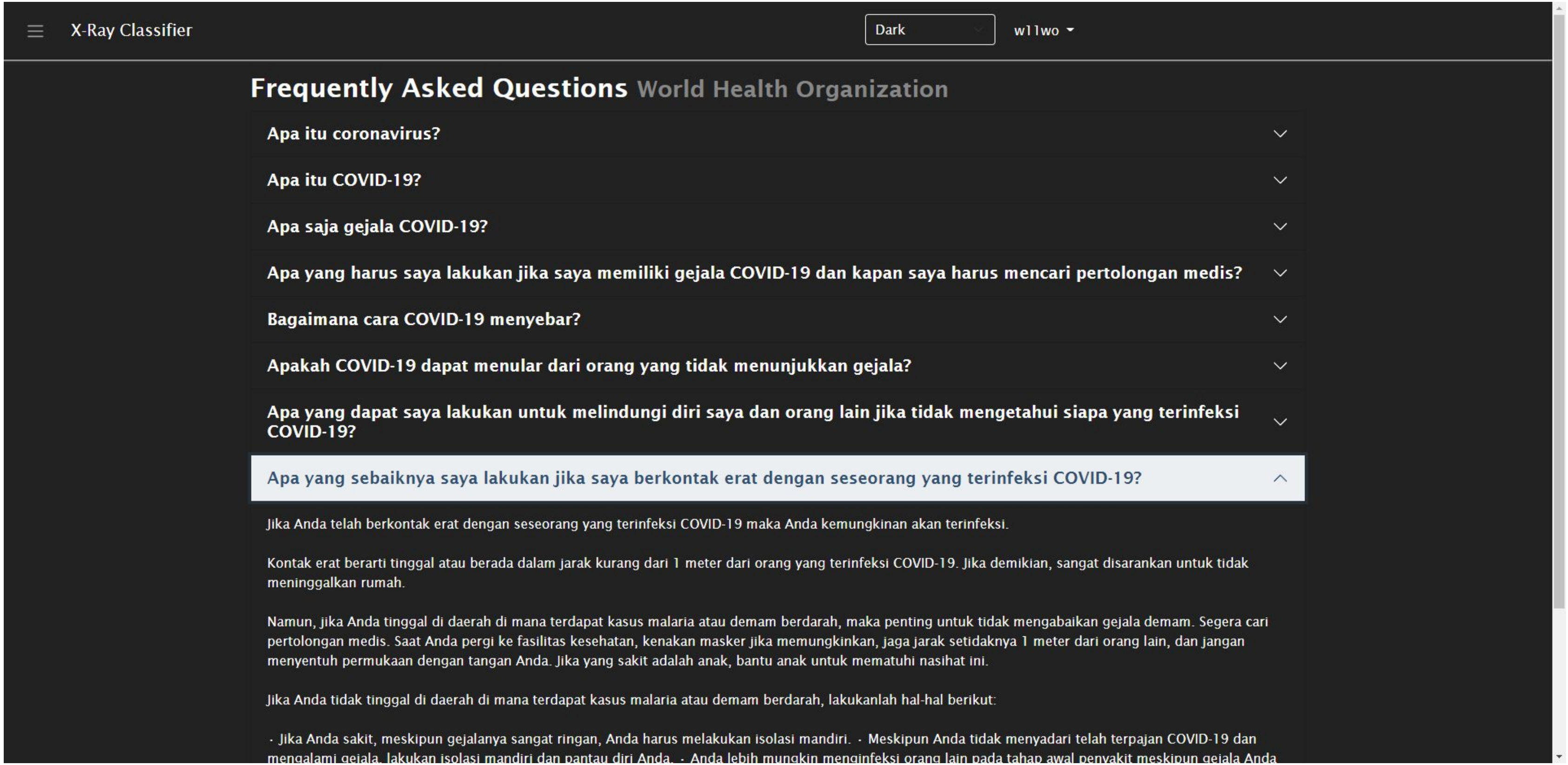
Login & Register



Upload & Predict



Get Next Steps and Suggestions



Conclusion

Conclusion

- We trained multiple models for differentiating COVID-19 CXRs from healthy/normal CXRs. Our experiments also found that smaller or simpler architectures, such as Xception, were able to perform to a similar caliber as the more complex pre-trained CNNs, like DenseNet201.
- We are then able to deploy the CNN model to a web application via Laravel to provide an accessible solution to our intended users.

Thank you