CNN Model for COVID-19 Detection Using CT Scans

Adèle Collin, Kyla Gabriel, Chuck Lin, Ritvik Raina, Sofía Rojas

BMI 707 - Deep Learning for Biomedical Data

1.Context

The COVID-19 pandemic significantly impacted global health, exposing the limitations of traditional diagnostic methods like PCR and rapid antigen tests, which are either time-consuming or offer moderate sensitivity [1].

This project extends the research evidencing the efficacy of computed tomography (CT) scans in identifying COVID-19-related anomalies in the respiratory system, such as ground-glass opacification and bilateral involvement, which are critical markers of the infection [2, 3]. CT scans provide a rapid and accurate alternative for diagnosing COVID-19, crucial in managing the high volume of cases. This study leverages advanced machine learning techniques to address the challenges posed by the scale of the pandemic and to enhance the diagnostic processes.

2. Research Question

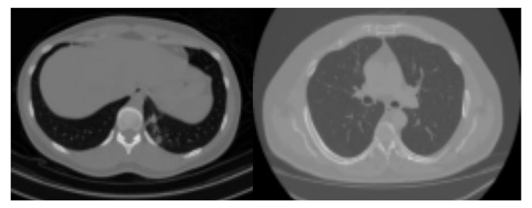
How can deep learning be effectively applied to CT scans to develop a diagnostic tool for COVID-19?

Specifically, this project aims to develop and evaluate a neural network architecture tailored for CT-based diagnosis of COVID-19, enhancing early detection capabilities crucial for controlling the spread of the virus.

3. Data Collection

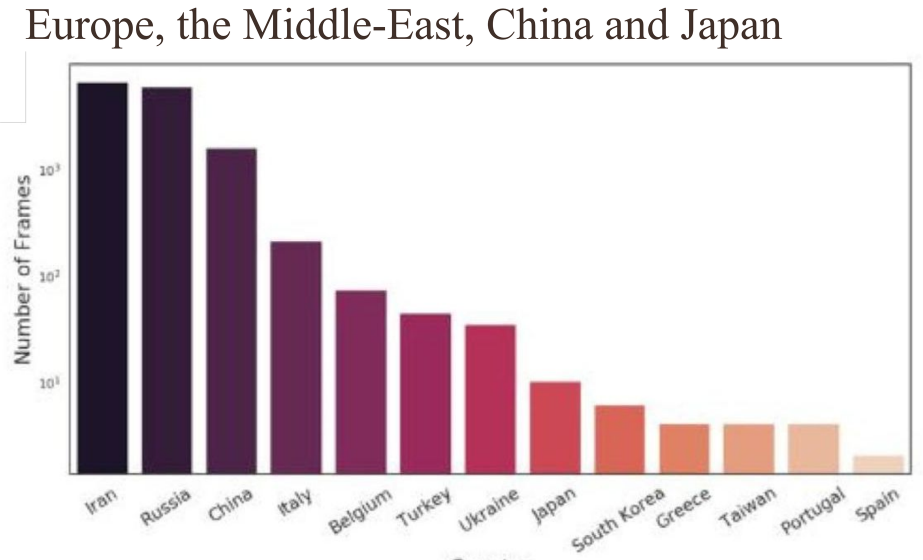
1) What

- COVID-19 CT scans of chest



2) Where

- Europe, the Middle-East, China and Japan



3) Who

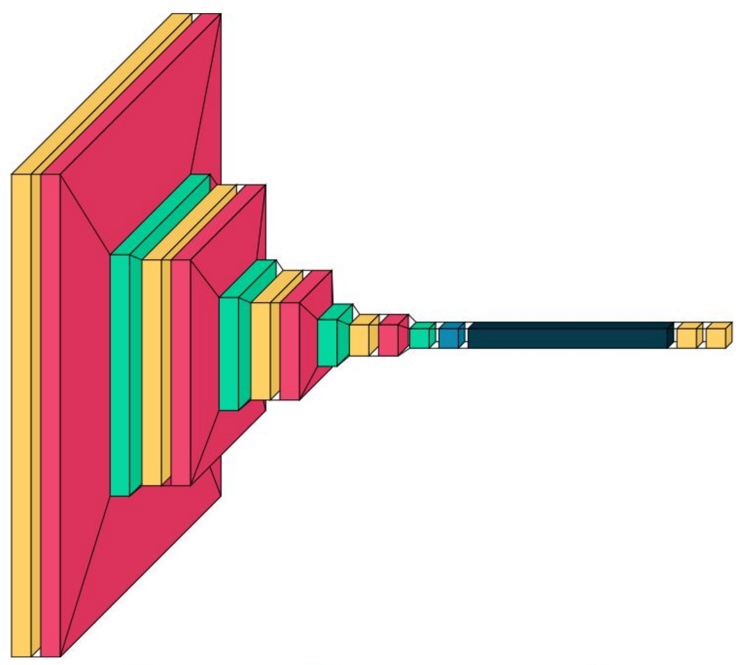
- 7,593 COVID-19 Images
- 6,893 Control images
- Sex:
  - 59% Male
  - 32% Female
  - 9% unknown
- Average age: 53

4) How

- The dataset is composed by Integrating 7 databases of CT scans

4. Methodology

1) Custom CNN



2) ResNet50

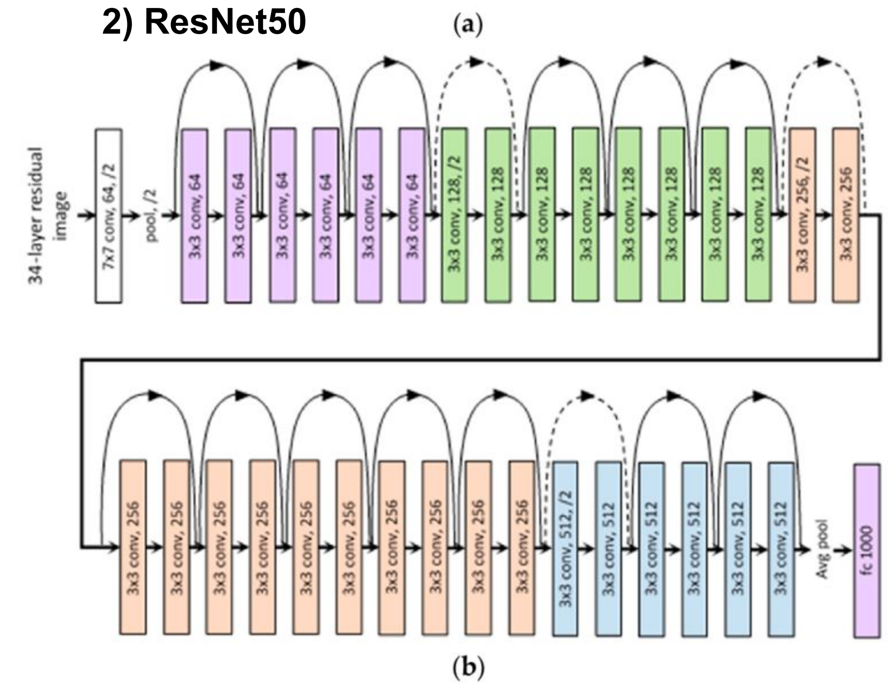


Image Source: Siddhesh Bangar [4]

3) Logistic Regression

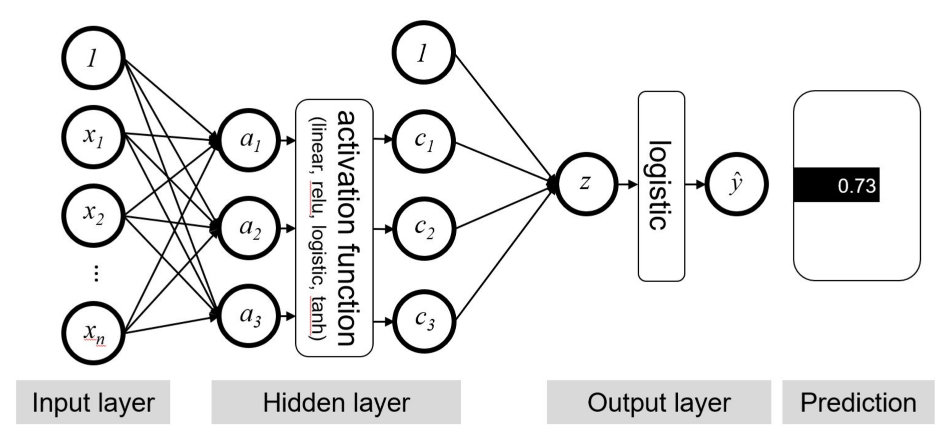


Image Source: Gitter Lab [5]

4) Random Forest

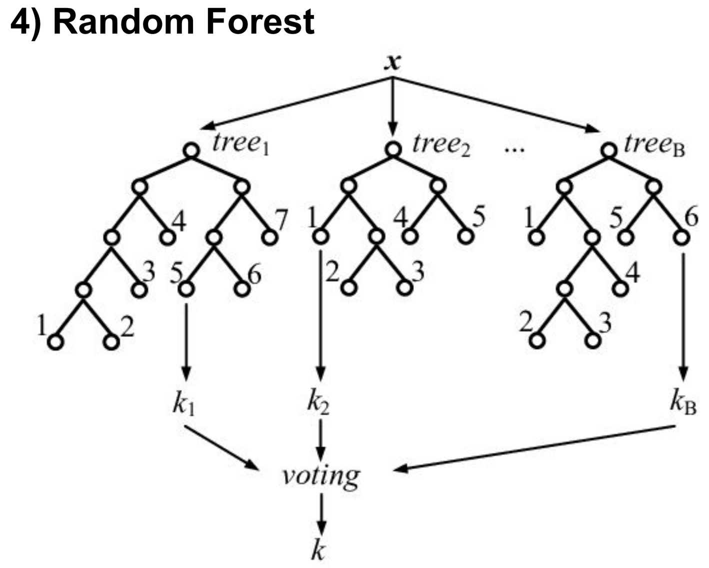


Image Source: Evaldas Vaiciukynas [6]

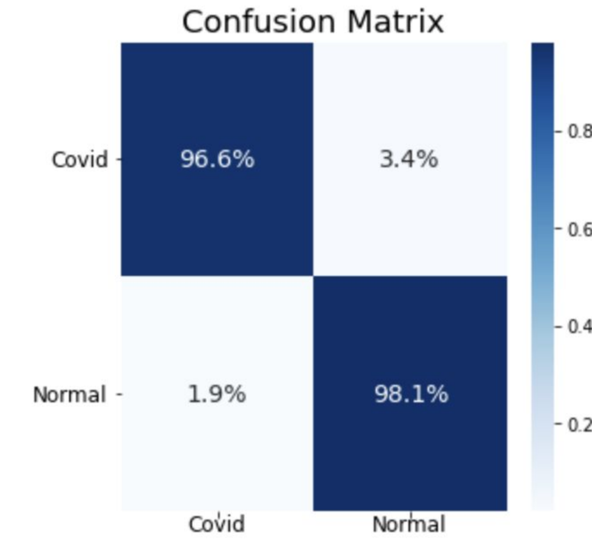
Summary of Model Hyperparameters

Model	Hyperparameter	Value
Custom CNN	Optimizer	*Adam*
	Learning rate	*0.001*
	Batch size	*64*
	Epochs	*20*
ResNet50	Optimizer	*Adam*
	Learning rate	*0.001*
	Batch size	*64*
	Epochs	*20*
Logistic Regression	Penalty	*l2*
	C	*1.0*
	Tol	*1e-4*
	Solver	*lbfgs*
	Max_iter	*100*
Random Forest	n_estimators	*50*
	Criterion	*gini*
	min_samples_split	*2*
	min_samples_leaf	*1*
	min_weight_ fraction_leaf	*0.0*
	max_features	*sqrt*
	max_leaf_nodes	*None*
	min_impurity_decrease	*0.0*
	bootstrap	*True*

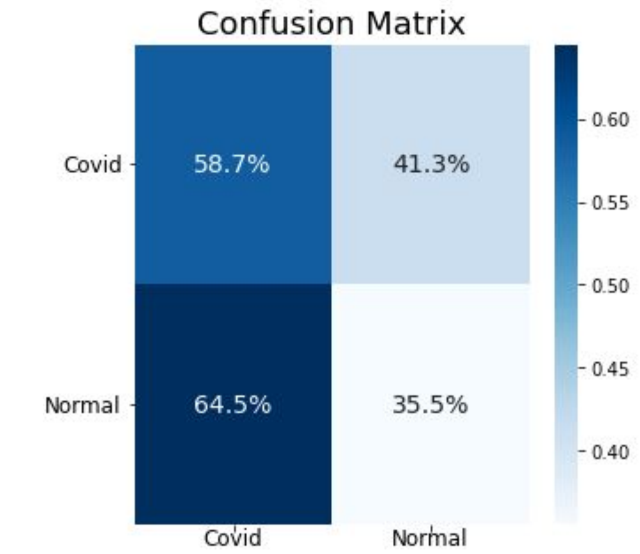
5. Results

Model	Metrics	Value
Custom CNN	Precision	*0.98*
	Recall	*0.97*
	F1	*0.97*
	Accuracy	*0.97*
ResNet50	Precision	*0.50*
	Recall	*0.59*
	F1	*0.54*
	Accuracy	*0.48*
Logistic Regression	Precision	*0.91*
	Recall	*0.93*
	F1	*0.92*
	Accuracy	*0.93*
Random Forest	Precision	*0.92*
	Recall	*0.98*
	F1	*0.95*
	Accuracy	*0.95*

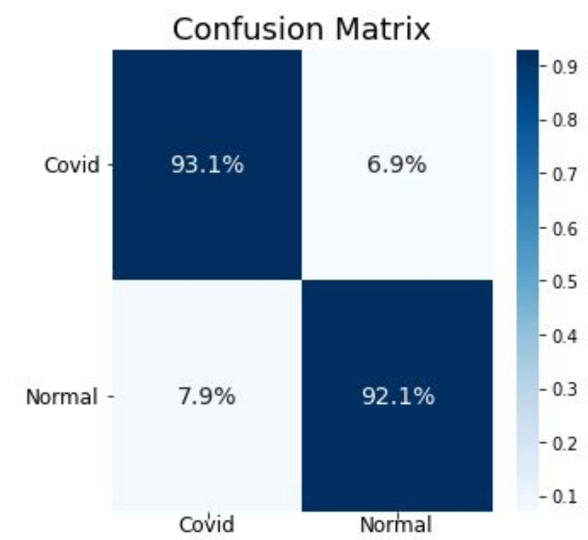
1) Custom CNN



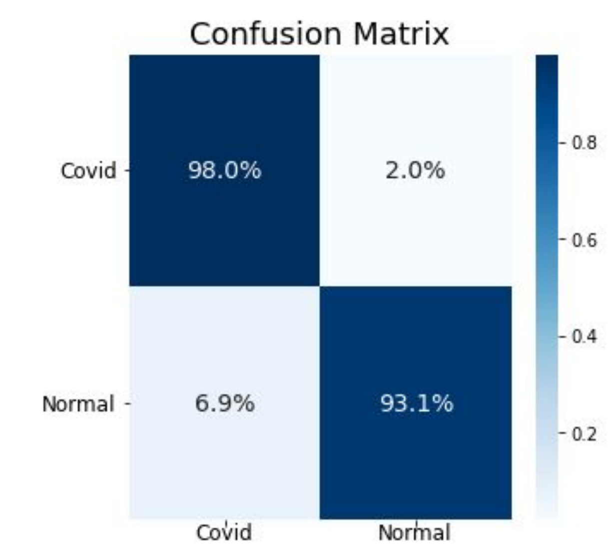
2) ResNet50



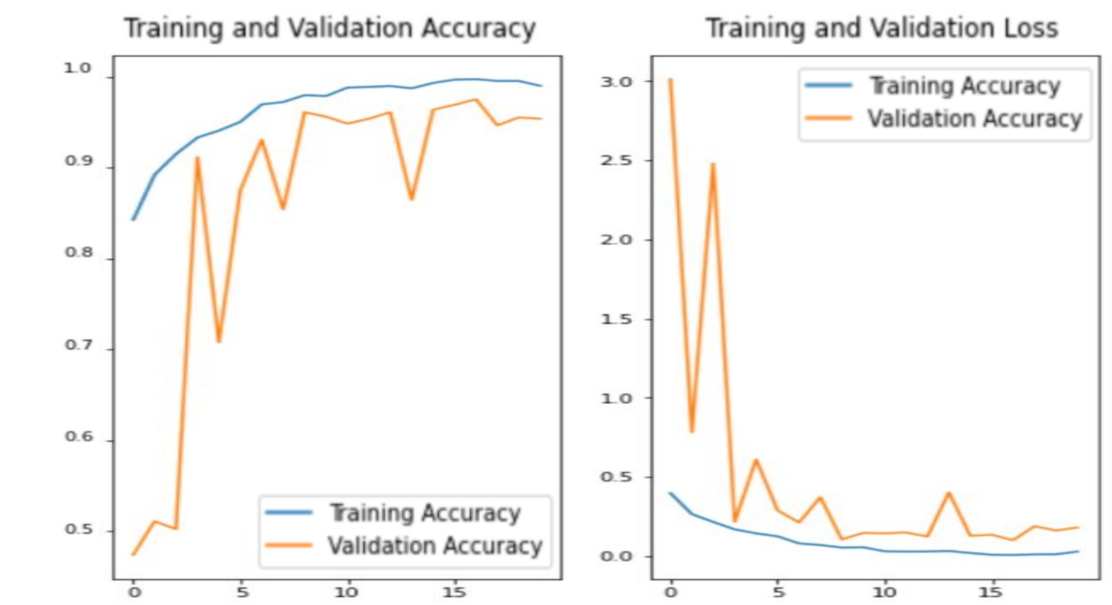
3) Logistic Regression



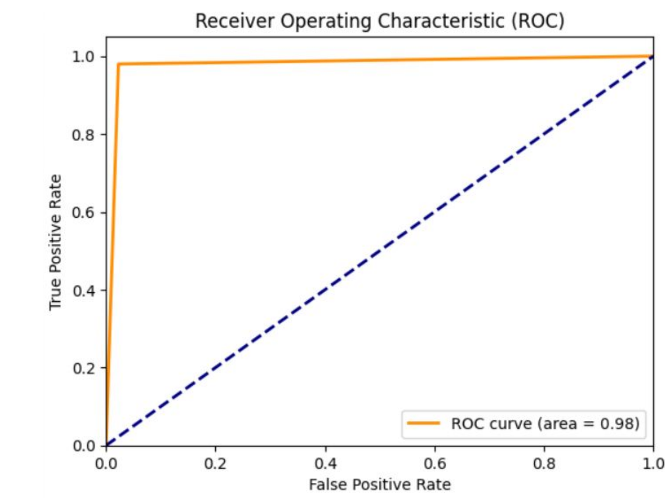
4) Random Forest



6. Results [Cont.]



Custom CNN Training and Validation Accuracy & Loss



Result of the ROC curve and the AUC of the custom CNN model

7. Conclusion

- **Overview**
  Our CNN model most effectively predicted COVID-19 diagnosis with a F1 score of 0.97 using chest X-rays of the lungs. We tested the effectiveness of our CNN model against other machine learning methods and found that the CNN was the best performing in terms of differentiating between COVID and Normal cases.
- **Challenges Faced**
  Initial model training due to computational limitations was dealt with by converting all images into Numpy arrays.
- **Limitations**
  Potential biases may exist because of variations in image acquisition. Also, the dataset used may not be representative of how COVID-19 manifests in different demographic populations.
- **Future Steps**
  We would like to perform further validation by testing our model on the dataset collected by Shakouri et al. And further validate Transfer Learning results with a shallower model.

8. References

Please feel free to reference our sources by scanning this QR code.

BLAVATNIK INSTITUTE
HARVARD MEDICAL SCHOOL