Diagnosing COVID-19 Using CT-Scan Images

Hui Zeng, Weibo Xu, Jiteng Fu, Yu-Chien Lee, Zhuoyuan Li , Matthew A. Lanham

Purdue University, Krannert School of Management

zeng162@purdue.edu; xu1069@purdue.edu; fu229@purdue.edu; lee2925@purdue.edu; li2952@purdue.edu; lanhamm@purdue.edu

ABSTRACT

The COVID-19 pandemic has been a huge threat to human health. This study investigates machine learning and deep learning techniques to predict if a patient has COVID-19 or not based on CT scans. We posit this model can be beneficial in numerous ways that are almost covered from emergency to post-treatment. By using our model, researchers will be able to provide and analyze the CT-scan to diagnose accurately and more quickly than other tests. This research project was part of the INFORMS 2020 QSR-Data Challenge. Here we learned how to process CT-scan images and properly predict and evaluate sophisticated predictive models that might help in diagnosing COVID-19. We believe our approach in this research can be extended to other projects where images are used for prediction purposes.

INTRODUCTION

We predict if someone has COVID-19 based on a patient's grey scale CT scan image. As Sharma (2020) stated, CT scan images may be an alternative solution to diagnose COVID-19 infection. Below are two CT scan figures, one for non_COVID and one for COVID. Our predictive analytics approach would allow hospitals to identify the possibility of a COVID-19 positive patient in a more efficient way using Deep Learning methods. We believe this study of using predictive analytics on images could be extended into other practical areas of business where a firm is tasked to identify or classify products, customers, or defects using deep learning techniques.

Fig 1. CT_COVID

Fig 2. CT_Non_COVID

Research questions:

- Could a hospital successfully diagnose a patient having COVID-19 using their CT scan?
- How effective are deep learning methods at COVID-19 diagnosis of CT scanned images?

LITERATURE REVIEW

We compare our model with other similar studies by distinguishing the dataset size and the different methods we used in data processing and model building.

Study	Dataset Size	Preprocess Method	Model
(2020) EED Hemdan	Small	Image-pixel conversion	VGG19
(2020) Song Ying	Small	Hounsfield Unit	DRE-Net
(2020) Chuansheng Zheng	Large	Hounsfield Unit	DeCoVNet
(2020) Harsh Panwar	Large	Image-pixel conversion	CNN (nCOVnet)
Our Study	Large	Image-pixel conversion	Deep Learning, Neural Network

Our model is advanced because (1) we included a larger dataset; (2) we used a more general preprocess method that is widely used in image prediction.

METHODOLOGY

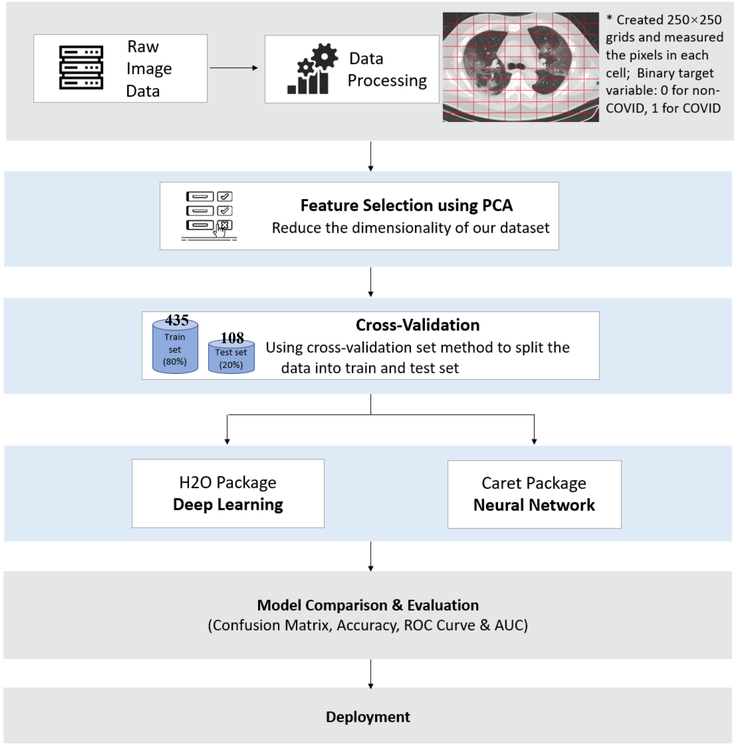

Fig 3. Study Design

STATISTICAL RESULTS

As our problem has been defined as classification-type, the output results in both models (deep learning model using H2O package and neural network model using caret package) include same output parameters. The comparison between the results from both models are presented as figures below.

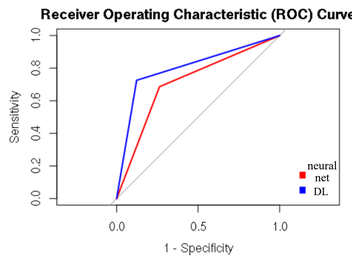

Fig 4. ROC Curve Results

	DL Model	Neural Net Model
Auccuracy	0.8056	0.713
95% CI	(0.71383, 0.8754)	(0.618, 0.7959)
Sensitivity	0.8409	0.7368
Specificity	0.7812	0.6863
AUC	0.8013	0.7116

Fig 5. Performance Metrics Comparison

The deep learning model using H2O package has a higher accuracy and larger area under the curve than the neural network model using the caret package. Therefore, we chose the deep learning model as our best model.

EXPECTED BUSINESS IMPACT

By using our final model, we will be able to predict the likelihood of COVID-19 of a patient based on the CT-scan result. This model can be a practical assistance for patient who needs fast examination.

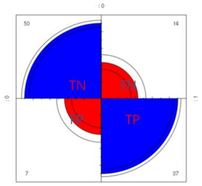

Fig 6. The Confusion Matrix

Our Assumptions

Cost Matrix		Actual Class	
		NO	YES
Predicted Class	NO	0	1000
	YES	200	-230

Fig 7. The Cost Matrix

	With DL Model	Without DL Model	Improvement
Sensitivity	84.09%	58.88%	25 points
Specificity	78.12%	61.44%	17 points
Accuracy	80.56%	60.23%	20 points

Fig 8. Potential Improvement on Predicting Performance

Cost Comparison (With Model vs. Without Model)			
TN	FN	TP	FP
↑ 13.85%	↓ -7.41%	↑ 6.48%	↓ -12.92%
-	Save $1000/patient	Gain $230/patient	Save $200/patient

Fig 9. Cost Comparison

Doctors can only determine general viral pneumonia but could hardly diagnose specific diseases manually. Compared with manual doctor diagnosis, the predicting performance of using our model **improves around 20%**. With the cost comparison, the average saving per prediction with our model is **$91.58**

CONCLUSIONS

The whole model is based on deep learning and neural network methods, which can be used for image recognition. This model can successfully diagnose COVID-19 by learning the pixel values extracted from the CT scan images and classify the patients into two levels: 0 for non COVID and 1 for COVID. This model can be used by the hospitals to do quick analysis with the following process:

Input the scan image -> model processing -> suspicious patients/confirmed patients

The statistical results of 80% accuracy show the effectiveness of our model and the implement of PCA can shorten the run time to a large extent.

Hence, the benefits that our model could bring include:

- Allowing hospitals to diagnose virus more accurately
- Making CT-scan as a fast and effective method for doctor to use

ACKNOWLEDGEMENTS

We'd like to thank Professor Matthew Lanham, Xinyu Wang, Theo Ginting, graduate student mentors Rassul Yeshpayev and Khalid Sinada for their guidance on this project.