COVID19 Diagnosis using AutoML from 3D CT scans

Talha Anwar

Independent Researcher

ICCV 2021 Workshop: MIA-COV19D

Introduction

- Coronavirus disease (COVID-19) is an infectious disease that primarily affects lungs.
- Standard way to diagnose COVID is to use PCR test
- Researchers at John Hopkins University studied that On the day of symptom onset, the median false-negative rate was 38%. This decreased to 20% on day 8 (3 days after symptom onset) then began to increase again [1].
- To overcome this false negative ratio, an aided method is required.
- This can be an imaging technique such as Xray or CT scan that can provide detailed image of chest

Problem statement

- Unlike PCR, diagnosing COVID from CT-scan requires expert radiologists or pulmonologists.
- Due to huge number of COVID patients and a smaller number of expert doctors, it is not possible to examine CT scans rapidly.
- Further, due to over workload, the doctor may misdiagnose the disease.

Motivation

- To make radiologist and pulmonologist work automatic and rapid, an artificial intelligence based method is needed.
- AI-enabled Medical Image Analysis Workshop and Covid-19 Diagnosis Competition (MIA-COV19D) organized a competition to bring the most accurate solution to diagnose COVID from CT scan by proving a huge data labelled by four specialists, two radiologists, and two pulmonologists.

Competition source: https://mlearn.lincoln.ac.uk/mia-cov19d/

Material and Method

- Dataset
  - Dataset is split into 3 sets; training, validation and test.
  - Each 3D CT scan is composed of 50 to 700 2D slices.
  - Test set comprised of **3455** CT scans.
  - External data: An external data is also used to validate the trained model. It comprised of comprised of 9776 non-COVID and 2282 COVID slices of different CT scans.

Class	**2D/3D**	**Train**	**Validation**
COVID	2D slice	153681	35016
	3D scan	690	165
Non COVID	2D slice	181991	40516
	3D scan	780	209

Training and validation data distribution

Material and Method

- Deep Learning
  - Deep learning is a branch of artificial intelligence that extract features from data by passing it through different layers and then classify it according to corresponding labels.
  - There are different deep learning models and frame work
- Auto machine learning framework (AutoGluon) is used to carry out experimentation.
  - It saves human effort from pre-processing and make things completely automatic.

Material and Method

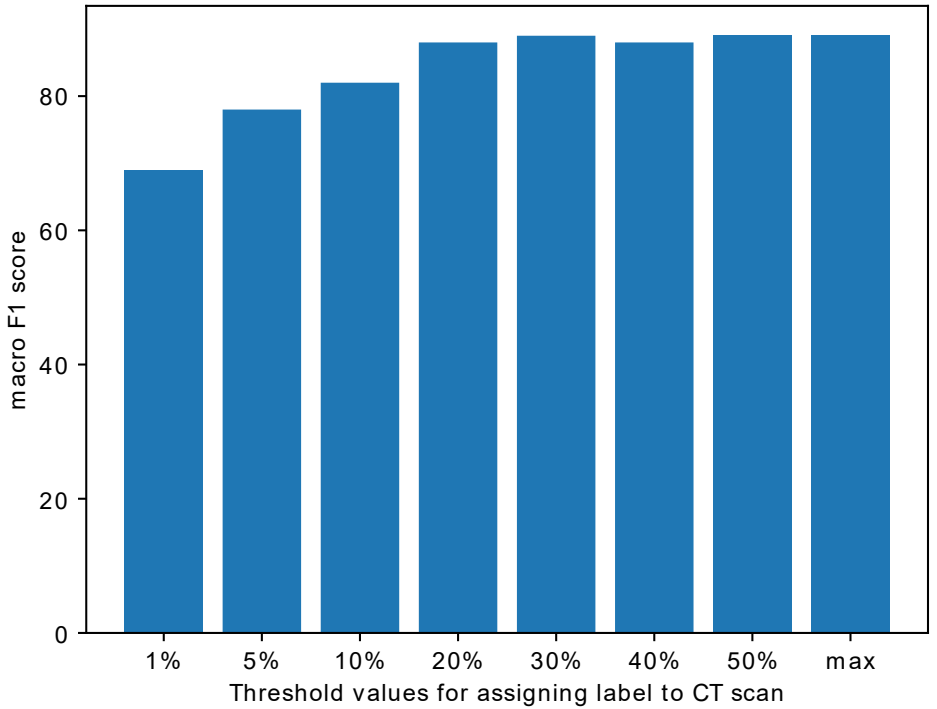
- Seven different deep learning architectures are used
  - VGG19 with batch normalization
  - ResNet152
  - DenseNet201
  - ResNest14
  - ResNext50
  - SeResNext50
  - MobileNetV3
- All the models are trained on 2D slices and ensembled together.

Material and Method

- Conversion of predictions from 2D to 3D
- As 2D models are used for training, so predictions are also slices wise.
- But predictions are need for 3D scans
- So different thresholding techniques are used
  - If x% of prediction contain COVID prediction, mark all the slices of a scan as COVID. The value of x used is 1, 5, 10,20,30,40, and 50
  - Label the 3D scan based on maximum occurrence of class.
- For final submission, 10%,20%, 30%, maximum class occurrence and ensemble of all technique is used.

Results

- Figure shows the macro F1 score using different threshold techniques on validation data.
- Similar F1 score is achieved using 20%,30%,40% ,50 and by selecting maximum occurrence of class.



Results

	2D CT-scan slice				3D Volumetric Image			
Model name	**F1-score**	**Precision**	**Recall**	**Accuracy**	**F1-score**	**Precision**	**Recall**	**Accuracy**
ResNet152	80%	82%	80%	81%	85%	89%	85%	86%
DenseNet201	81%	83%	81%	82%	84%	88%	84%	86%
ResNest14	**85%**	**86%**	**85%**	**85%**	**88%**	**90%**	**88%**	**89%**
ResNext50	83%	83%	83%	83%	88%	89%	89%	88%
Se_ResNext50	84%	84%	83%	84%	87%	89%	87%	88%
MobileNetV3	80%	81%	88%	81%	86%	88%	86%	87%
VGG19_bn	84%	84%	84%	84%	87%	89%	87%	88%

Results on the validation set using maximum occurrence of class technique in 3D CT scan

Results

- Table shows that model size after training, the time for training and time for inference on test set.

Model name	**Model Size**	**Training Time**	**Inference Time**
ResNet152	472 MB	7 hours 41 minutes	3 hours 15 minutes
DenseNet201	166 MB	5 hours 21 minutes	2 hours 48 minutes
ResNest14	**93 MB**	7 hours 24 minutes	2 hours 22 minutes
ResNext50	203 MB	6 hours 31 minutes	2 hours 7 minutes
Se_ResNext50	222 MB	6 hours 10 minutes	2 hours 49 minutes
MobileNetV3	63 MB	6 hours 21 minutes	2 hours 37 minutes
VGG19_bn	1 GB	7 hours 56 minutes	2 hours 42 minutes

Results

- On external data macro F1 score of 76% and accuracy of 84% is achieved.
- On test set F1 score of 78.78% is achieved using maximum occurrence of class technique.

Threshold	**Macro F1**	**F1 COVID**	**F1 Non COVID**
10%	61.8	44.85	78.75
20%	74.4	59.25	89.54
30%	80.88	68.3	93.45
Maximum class	**87.77**	**78.78**	**96.75**
Ensemble of above	79.36	66.07	92.64

Leaderboard Score

Ranking	Author	F1-score	Technique
1	Miron	90.06 %	Inflated 3D ResNet50
2	Hsu	88.74 %	CT scan-Aware Transformer
3	Tan	88.22 %	BERT and MLP
4	**Our Approach**	**87.77 %**	**AutoML**
5	Liang	78.86 %	SE CNN and transformer
6	Trinh	78.13 %	Res Dense Net
7	Qi	71.83 %	3D-RegNet
8	Teli	70.86 %	TeliNet (custom CNN)
9	Gao	70.5 %	COVID-ViT
10	Kollias	70%	CNN RNN

Link for leaderboard

Conclusion

- CT-scan coupled with AI can be used as a quick diagnostic method to categorize patients into “probably positive” and “probably negative” cohorts.
- **Limitations**:
  - The methodology would be more generalizable if source of training and test data is different. It mean that train data is collected from different group of CT scans machine, hospitals and countries, where as test data should be from different group. This will help to explain that these method can be used every where round the globe or not. As a participant, no such information is available.

Acknowledgement

- I would like to specially thanks to Dr Dimitrios Kollias for organizing this competition.
- I would also like to thanks Nick Erickson and Joshua Z. Zhang from AutoGluon for resolving my queries and issue regarding AutoGluon.

References

[1] Kucirka, Lauren M., et al. "Variation in false-negative rate of reverse transcriptase polymerase chain reaction–based SARS-CoV-2 tests by time since exposure." *Annals of internal medicine* 173.4 (2020): 262-267.

[2]. Mohammad Rahimzadeh, Abolfazl Attar, and Seyed Mo-hammad Sakhaei. A fully automated deep learning-basednetwork for detecting covid-19 from a new and large lungct scan dataset. Biomedical Signal Processing and Control,68:102588, 2021.

Thank you