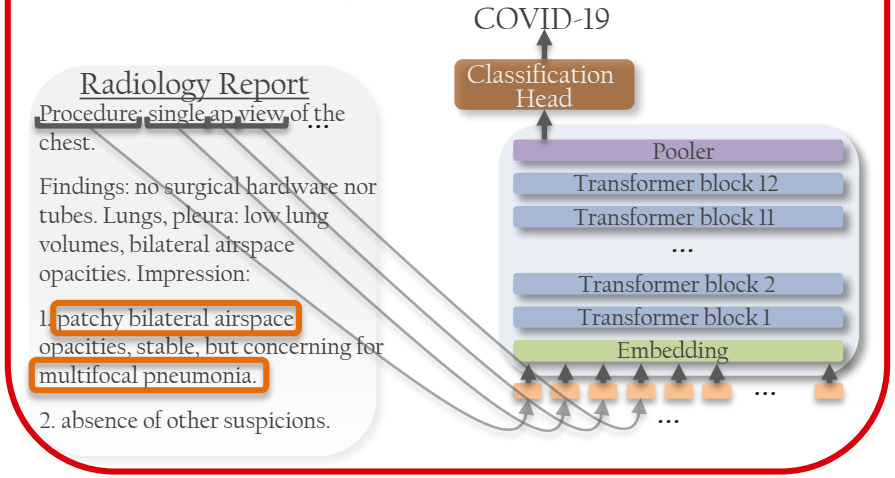
Improved Fine-tuning of In-domain Transformer Model for Inferring COVID-19 Presence in Multi-institutional Radiology Reports

Pierre J. Chambon, Tessa S. Cook, Curtis P. Langlotz

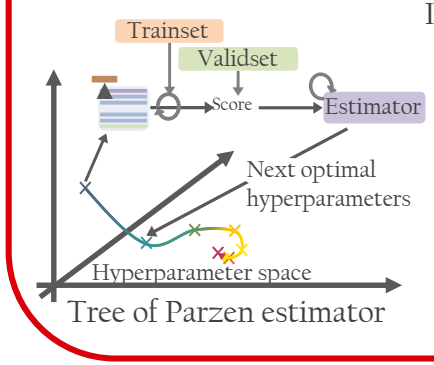
The Problem

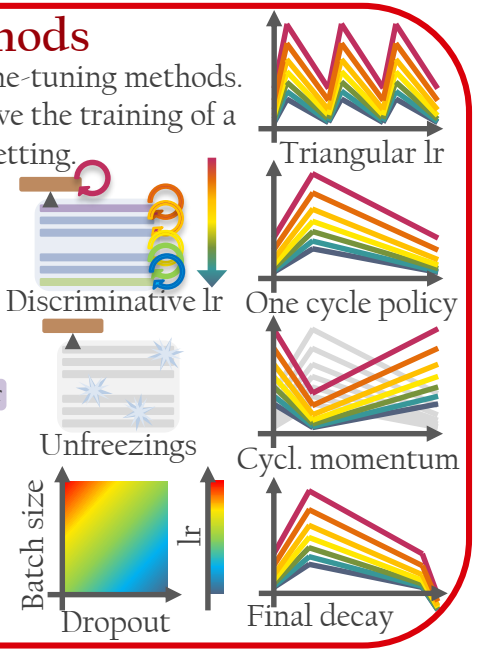
- Detect COVID-19/uncertain/no COVID-19 in radiology reports (X-rays, CTs). Distinguish from other concurrent lung diseases such as focal pneumonia, atelectasis, ...



Methods

- Various pre-training and fine-tuning methods.
- Evaluate which ones improve the training of a transformer and in which setting.
- Comparison with existing models (BioBERT, BlueBERT, ...).





Results

- Models are tested on 1387 X-rays (in green) and on 267 CTs (in blue), and compared using F1-score.

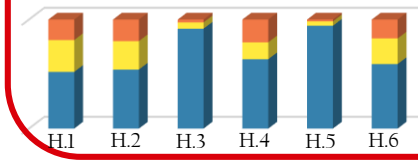
Method	COVID-19	Uncertain COVID-19	No COVID-19	Macro avg.
Pre-training				
BERT	88.6 / 81.1	86.6 / 59.8	95.1 / 91.6	90.1 / 77.5
BioBERT	87.8 / 79.0	87.2 / 57.1	95.2 / 91.6	90.1 / 75.9
BlueBERT	87.1 / 82.5	86.9 / 54.4	95.6 / 90.8	89.9 / 75.9
Ours	89.1 / 83.9	87.1 / 61.7	95.3 / 92.5	90.5 / 79.4
Fine-tuning				
Standard	86.6 / 80.9	85.9 / 51.5	95.0 / 88.6	89.2 / 73.7
Ours	89.1 / 83.9	87.1 / 61.7	95.3 / 92.5	90.5 / 79.4

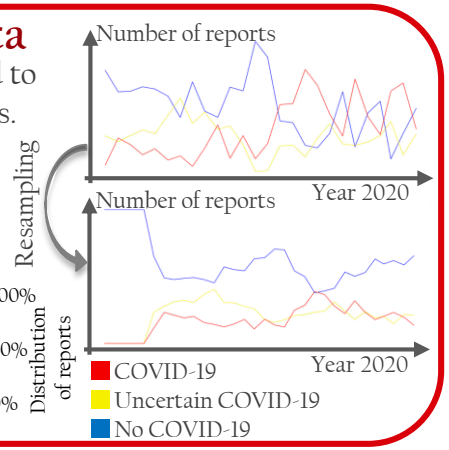
- Most errors are due to misclassifications between COVID-19 and uncertain COVID-19 reports.

Data

- 3 labels on patients suspected to have COVID-19, by physicians.

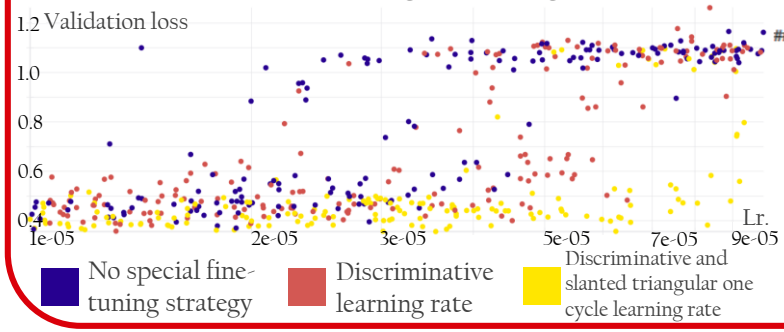
Pre-training	Fine-tuning
4 million reports	20k reports+labels
1 institution	6 hospitals



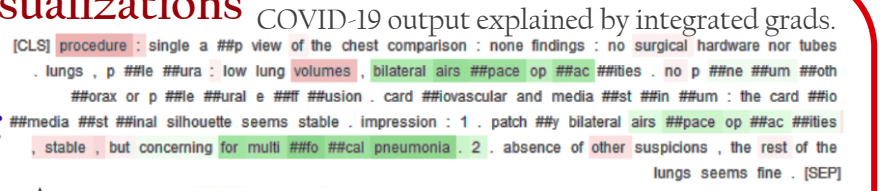


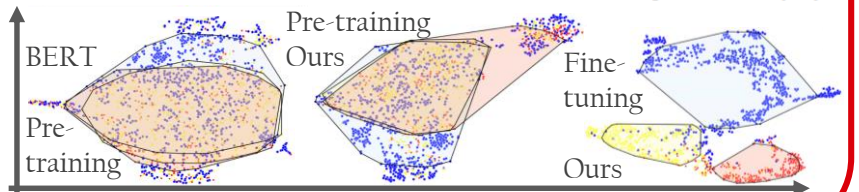
Visualizations

500 random samples of the hyperparameter space. Yellow runs are more stable even at higher learning rate values.



COVID-19 output explained by integrated grads.





2D projection of the transformer hidden-states, on test reports.