
Using deep learning on chest CT to track COVID-19 patients

000
001
002
003
004
005
006
007
008
009
010
011
012
013
014
015
016
017
018
019
020
021
022
023
024
025
026
027
028
029
030
031
032
033
034
035
036
037
038
039
040
041
042
043
044
045
046
047
048
049
050
051
052
053
054

Anonymous Authors¹

Abstract

We developed COVID3D, a 27-layer convolutional neural network that uses the entire chest computed tomography (CT) volume to automatically predict COVID+ from COVID- pneumonia on two external hold-out institutions with comparable performance from two board-certified radiologists. Next, we used COVID3D to describe the trajectory of the patient's disease course. We finetuned COVID3D on follow-up patients to predict prognosis based on length of hospitalization using 1 scan only and 2 serial scans (1 prior and 1 follow-up). This study used 397 CT scans or 84,806 image slices from 3 hospitals.

1. Introduction

Coronavirus disease 2019 (COVID19) caused by severe acute respiratory syndrome corona virus 2 (SAR-CoV-2) has inflicted a global health crisis and was declared a pandemic in March 2020(1). The high transmission rates that can lead to respiratory distress and multiple organ failures, requisite critical care resources, and rising mortality(2; 3; 4) have prompted an urgent need for early detection, accurate diagnosis, and predictive tools.

Real-time reverse-transcription polymerase chain reaction (RT-PCR) is the primary method for SAR-CoV-2 diagnosis. However, RT-PCR has shown variable sensitivity and specificity(6; 7; 27) either due to insufficient viral load, sample collection methods, or lack of definitive reference standards(8; 9). Studies have reported characteristic imaging features of COVID19 pneumonia(10) and proposed chest CT to either complement RT-PCR or serve as the initial workup in highly suspected cases given the potential for false negative RT-PCR(11; 12; 13) and to gauge disease severity(14; 15).

As the pandemic expands to global regions with limited

access to nucleic acid detection kits, chest CT may play a greater diagnostic role for COVID19 and disease monitoring, highlighting a need for automated or quantitative analytics. Recently, studies have reported success in deep learning methods with 2D CT slices as inputs for COVID19 classification(18) or segmented outputs to quantitate lung opacification and correlate disease severity(20). Machine learning can capitalize on large-scale, high-dimensional image data and offers the opportunity to optimize a framework for COVID19 evaluation, including prognostic models that stratify risk groups. This study goal was to develop a deep learning model to perform two tasks: **Task 1:** classify COVID19 pneumonia from non-COVID19 pneumonia, and **Task 2:** predict disease course (prognosis) from length of hospitalization.

1.1. Prior Works

Prior COVID+ pneumonia prediction studies have used either individual 2D slices or combined 2D slices to form a 2.5D model(18; 17). Li(18) et al. used a 2.5D model where 2D lung-segmented slices are forward-passed into a 2D ResNet-50 N times (one for each of the N slices), which then concatenates these features to form a prediction. However, given the 2D nature of data input and training, 2D or 2.5D models either distort or lose important spatial features related to depth contiguity. Recently published 3D models(19) show promise but rely on another segmented lung model for preprocessing. Furthermore, there is a lack of work in prognosis or outcome. To our knowledge, this is the first 3D model without lung segmentation for use in both classification and tracking prognosis across follow-up scans.

2. Methods

Dataset

We conducted this multi-center retrospective study with institutional review board approval. Eighty-three consecutive COVID+ and 83 consecutive COVID- pneumonia patients who met the inclusion criteria at Hospital A between January to February 2020 at Hospital A were recruited for the study. Inclusion criteria were patients who presented with clinical symptoms suspicious for COVID-19 pneumonia, obtained at least two confirmatory real time RT-PCR tests

¹Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

055 to determine their COVID-19 status, and obtained diagnostic
 056 quality chest CT exams. Twenty consecutive COVID+
 057 and 20 COVID- pneumonia who met the same above
 058 inclusion criteria between January and February 2020 at an
 059 independent institution, Hospital B, served as an external
 060 dataset. Dataset from Hospital C served as a second external
 061 dataset and comprised 21 consecutive patients (12 COVID+;
 062 9 COVID- pneumonia).

063 Model

064 The COVID3D model consists of a 27-layer I3D-Inception
 065 feature extractor, 3D spatial average pooling, and 1 fully-
 066 connected layer. The Inception model(33) was pretrained
 067 on Kinetics-600, a video dataset(34). In task 1, to train
 068 COVID3D as a classifier, binary cross-entropy was used as
 069 the training loss using Adam optimization(35); we evaluated
 070 the model on the internal cross-validation set every epoch.
 071 Ten-fold cross-validation was used on hospital A. The model
 072 with the highest AUC over the cross-validation folds was
 073 chosen for evaluation on Hospitals B and C.

074 In task 2, we deployed COVID3D on successive follow-up
 075 scans to measure COVID3D features over time. The goal
 076 was to track the progression of each patient's individual
 077 disease status in an unsupervised manner. Higher feature
 078 scores for a given scan mean that the scan appears more
 079 COVID-like, and lower feature scores mean less COVID-
 080 like. In Fig. 2, we compute features denoted as $s(t)$ for
 081 49 follow-up patients from Hospital A and 4 patients from
 082 Hospital C without any additional supervision from task 1.

083 We finetune the original COVID3D classifier to predict pa-
 084 tient prognosis. We quantify prognosis by the length of
 085 hospitalization time (in days) measured from when the scan
 086 was imaged to time at discharge. All discharge times are
 087 uncensored except for 1 patient in Hospital C who is still
 088 hospitalized. A longer hospitalization time is indicative
 089 of worse prognosis. Rather than predicting the time using
 090 regression, we treat this problem as a binary classification
 091 problem to classify whether the patient will stay hospital-
 092 ized for longer than 7 days (median) given the presenting
 093 scan at any given time of the patient's disease course. We
 094 define 7 days or longer to indicate a subjective "high-risk"
 095 prognosis and below 7 days to indicate a "low-risk" sta-
 096 tus. We perform classification instead of regression. Cox
 097 proportional-hazards loss was not considered because 1) we
 098 designed this task as a classification problem to compare
 099 and interpret easily to radiologists, 2) we had no censored
 100 data with the exception of a single patient not used in the
 101 test set, and 3) asking radiologists to predict the number of
 102 days is not common clinical practice.

Human experts

The two radiologists are board-certified and each have over 10 years of experience in clinical radiology practice. For task 1, we asked each reader to differentiate COVID+ from COVID- (e.g. non-COVID19 pneumonia). For task 3, we asked each reader to first rate the prognosis of the following patient as either "high-risk" or "low-risk" after reviewing one scan and then rate again using both the prior scan and the follow-up. The readers were told that the prognosis is measured by length of hospitalization.

	Hospital B	Hospital C
2D ResNet-50	0.65 ± 0.01	0.69 ± 0.02
COVID3D	0.78 ± 0.01	0.84 ± 0.02
COVID3D* (tuned)	0.81 ± 0.01	0.87 ± 0.02

Table 1. Performance of COVID3D on 2 external institutions against a baseline 2D model trained over individual 2D slices. After evaluation, we noticed slight improvements to the AUC by very slightly modifying the lung windows of B and C to match that of the training set of Hospital A. However, we do not use this tuned model for any experiments beyond this table.

	Rad1-B	Rad2-B	COVID3D-B
Sensitivity (%)	95	70	83
Specificity (%)	77	75	70
	Rad1-C	Rad2-C	COVID3D-C
Sensitivity (%)	83	66	80
Specificity (%)	95	90	78

Table 2. The sensitivity and specificity in the classification task on hospitals B and C by 3 radiologists and from the COVID3D model. Reader #1 evaluated on hospital B is denoted as Rad1-B.

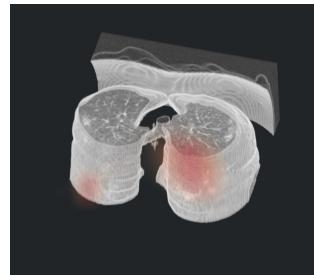


Figure 1. 3D view of a model-generated 3D Grad-CAM superimposed on the CT of a COVID+ case with bilateral peripheral ground glass opacities and consolidation. The map was generated from only 1 forward and 1 backward pass of 1 example in the test set. Note that because it's a 3D model, the focus of attention is contiguous in all spatial dimensions.

110 **3. Results**

111 **Task 1: classifying COVID+ from COVID- pneumonia**

112 COVID3D achieves an AUC of 0.78 ± 0.01 on hospital B
 113 and 0.84 ± 0.02 on hospital C. ROCs and AUCs are shown
 114 in Table 1. COVID3D on the internal validation sets of Hos-
 115 pital A achieved an average AUC of 0.92 ± 0.01 across the
 116 10 validation folds. Our volumetric-based approach is also
 117 far superior to a 2D approach using a ResNet-50 pretrained
 118 on ImageNet. AUC bounds were computed by evaluating
 119 COVID3D across 100 different image windows across 1 std.
 120 dev. of variability (Hounsfield Units) during test-time. This
 121 study is necessary to ensure that our model performance is
 122 robust and reproducible across a large diversity of scans. We
 123 ensured that slight mismatch in windows led to only modest
 124 and graceful degradation.

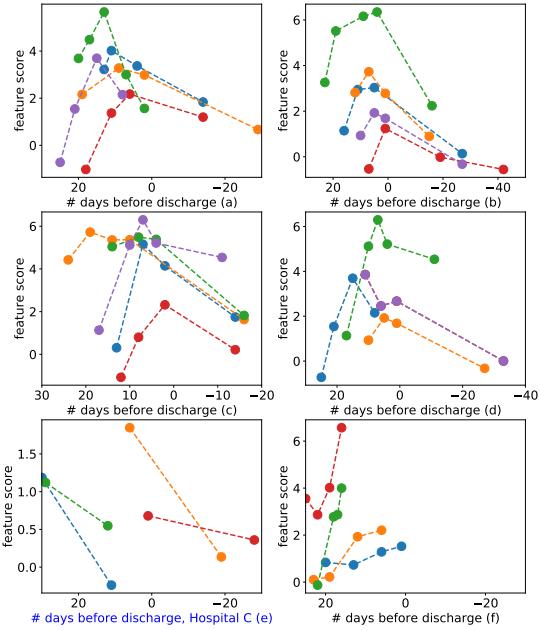
125 We compared COVID3D’s diagnosis prediction to that of
 126 radiologists in Table 2 using specificity and sensitivities
 127 as metrics. We observed that COVID3D performance is
 128 comparable to that of our radiologists, and also found large
 129 inter-observer variability consistent with the findings of
 130 Li(18) et al. We used COVID3D to generate Grad-CAMs
 131 (32) on our external cohorts. In Fig. 1, we illustrate exam-
 132 ples of where the COVID3D features activate strongly to
 133 certain regions of support in the lungs. Across the COVID+
 134 patients, we see almost all the ground glass opacity lighting
 135 up. In COVID- patients, we observe insignificant activation
 136 in scans from COVID- viral pneumonia.

137 **Task 2: Disease trajectory over time**

138 We deployed COVID3D on successive follow-up scans to
 139 measure COVID3D features over time. The goal was to
 140 track the progression of each patient’s individual disease
 141 status. Higher feature scores for a given scan mean that the
 142 scan appears more COVID-like, and lower feature scores
 143 mean less COVID-like. In Fig. 2, we compute features
 144 denoted as $s(t)$ for 49 follow-up patients from Hospital
 145 A and 4 patients from Hospital C without any additional
 146 supervision from task 1.

147 We show in Fig. 2(b) that virtually all patients appeared
 148 COVID+ to the model at initial presentation. However, the
 149 follow-up scans in about a week to 10 days from initial
 150 scan show an increase in severity. Using COVID3D fea-
 151 tures, we saw that virtually all patients peaked at around
 152 a week to 10 days from initial scan or roughly 10 to 15
 153 days from onset of initial symptoms. Finally, at the end
 154 of their hospitalization, patients had recovered so dramati-
 155 cally that COVID3D even predicted COVID- in many of
 156 these cases. Hence, we observe a curve characteristic of
 157 virtually all of the follow-up patients. The curve first starts
 158 high, peaks in slightly longer than a week and a couple
 159 days before the time-of-discharge, and decreasingly tapers
 160

161 to $s = 0$. The next logical question to ask is: can we
 162 map the curve’s shape (values, slope, 1st-derivative, 2nd-
 163 derivative) to predict prognosis? In the next section, we
 164 aim to build a prognostic model between these high-level
 165 temporal features and disease course.

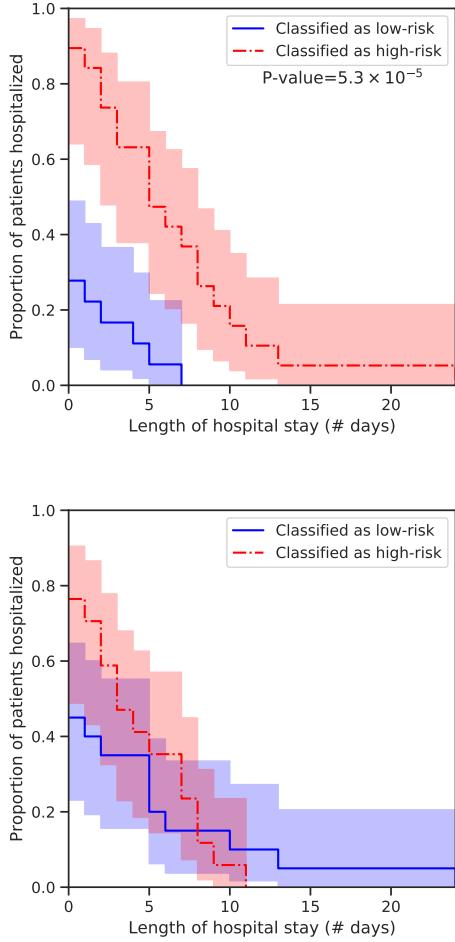


166 **Figure 2.** Features from COVID3D for the follow-up patients over
 167 time. Scans with high scores indicate high similarity from COVID+
 168 population. Many patients from initial scan to discharge show
 169 feature trajectories that start off as COVID-like but finish with an
 170 appearance characteristic of COVID- pneumonia. Four of Hospital
 171 A’s patients had no scan near their time of discharge.

172 **Finetuning to predict patient prognosis**

173 In task 2, our aim is to use COVID3D to predict the patient
 174 prognosis. We quantify prognosis by the length of hospital-
 175 ization (in days) measured from when the scan was imaged
 176 to time at discharge. A longer hospitalization window is
 177 indicative of worse prognosis. From our findings in Fig. 2,
 178 we expect that a large increase in COVID3D features over a
 179 short time window between 2 scans is indicative of a long
 180 hospitalization period and “high-risk” prognosis outcome.
 181 Similarly, a significant decrease in features over a short time
 182 indicates a low-risk prognosis. Features that grow in time
 183 but flatten out (concave down) may also indicate low-risk.
 184 Furthermore, predicting prognosis is difficult with one scan
 185 alone, and knowing two scans may not be enough to tell
 186 when the patient’s features will peak (i.e. curvature). Our in-

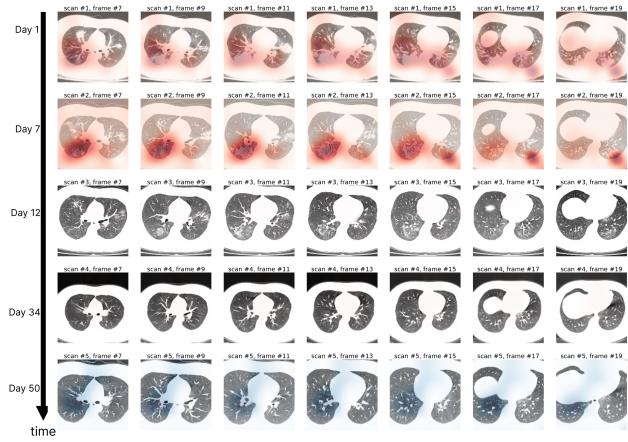
165 tuition tells us that radiologists and models trained to predict
 166 prognosis on hospitalization times can do better by looking
 167 at many sequential scans than just one scan. In the following
 168 experiments, we compared the prognostic performance of
 169 two scans (one prior and one follow-up) to one scan alone.
 170
 171
 172



204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 Figure 3. Kaplan–Meier plots on COVID disease course:
 COVID3D using 1 prior + 1 follow-up scans (top), and Radiologist
 1 using 1 prior + 1 follow-up scans (bottom).

The model stratifies the patients in the test set into high-risk or low-risk groups depending. We plot Kaplan–Meier (KM) plots for the two groups in Fig. 3. Using 2 sequential scans (one prior and one follow-up study) yielded better separation than using 1 scan alone (log-rank p-value of 5.3×10^{-5} using 2 scans versus 0.0033 using 1 scan) and clinical factors only (age and sex, p-value of 0.77). Two radiologists rated the prognosis of the test cohort as either high-risk or low-risk based on 1 scan and 2 sequential scans. The radiologists were informed that risk is measured with the length of hospitalization. Radiologist 1’s KM performance

is shown in Fig. 3(bottom). their stratification separations are not as compelling as that of COVID3D. Both radiologists had trouble stratifying patients who stayed hospitalized for longer than 9 days. Nonetheless, using two sequential scans modestly helped the reader (e).



203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 Figure 4. Case study using COVID3D on a follow-up patient
 (24 years old) with 5 scans. This patient was discharged on
 day 13. Gradient CAMs $H(x, y, z)$ are first scaled by $s(t)$ are
 superimposed onto the original CT, and color-coded (red for
 $H(x, y, z) > 0$ and blue for $H(x, y, z) < 0$). Severity predicted
 by COVID3D was highest on day 7 (as indicated by the visual
 difference between H on day 7 and day 1). On day 12, COVID3D’s
 $H(x, y, z, t = 12) \approx 0$ indicating significant recovery. In fact,
 the model predicts COVID+ with $\approx 50\%$ probability, indicating
 that this scan is indistinguishable from those of COVID- viral
 pneumonia cases.

4. Discussion

We present COVID3D, a single 3D model that diagnoses and tracks disease course over hospitalization without the aid of complex preprocessing. We finetune model parameters from classification to prognosis. Even without training on hospitalization times, we show that COVID3D from task 1 learned filters that correlate with the patient outcome. There are some limitations to our study. First, large sample size is always desirable, but we demonstrate model generalizability in two external sites. Creating any prognostic model has inherent challenges such as the existence of many complex clinical variables. To mitigate diversity in treatment protocols, we selected a cohort of patients who required hospitalization for COVID+ pneumonia but did not require intubation. Also, patients received mostly homogeneous therapy regimen including anti-HIV drugs.

220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279

References

- [1] Zhou P, Yan X-L, Wang Z-G et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579: 270-273. WHO Director-General's opening remarks at the media briefing on COVID-19. 2020. Available at: <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020>. Accessed March 22, 2020.
- [2] Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet* 2020 Jan 30. doi: 10.1016/S0140-6736(20)30211-7.
- [3] Wang D, Hu B, Hu C, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA* 2020.
- [4] Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *The New England Journal of Medicine* 2020.
- [5] Pan J, Wang MX, Ang IYH, et al. Potential Rapid Diagnostics, Vaccine and Therapeutics for 2019 Novel Coronavirus (2019-nCoV): A Systemic Review. *J Clin Med* 2020 Feb 26; 9 (3)
- [6] Liu R, Han H, Liu F et al. Positive rate of RT-PCR detection of SARS-CoV-2 infection in 4880 cases from on hospital in Wuhan, China, from Jan to Feb 2020. *Clin Chim Acta* 2020 Mar 7.
- [7] Wang W, Wu Y, Gao R et al. Detection of SARS-CoV-2 in different types of clinical specimens. *JAMA* 2020 Mar 11.
- [8] Pulia MS, O'Brien TP, Hou PC, et al. Multi-tiered Screening and Diagnosis Strategy for COVID-19: A Model for Sustainable Testing Capacity in Response to Pandemic. *Ann Med* 2020
- [9] Omer SB, Malani P, del Rio C. The COVID-19 pandemic in the US: a clinical update. *JAMA* 2020
- [10] Xu YH, Dong JH, And WM et al. Clinical and computed tomographic imaging features of novel Coronavirus pneumonia caused by SARS-CoV-2. *J Infect* 2020; 80: 394-400
- [11] Xie Z, Zong Z, Zhao W et al. Chest CT for typical 2019-nCoV pneumonia: relationship to negative RT-PCR testing. *Radiology* 2020; 200343
- [12] Ai T, Yang Z, Hou H et al. Correlation of chest CT and RT-PCR testing in Coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology* 2020 Feb 26
- [13] Yang W, Sirajuddin A, Zhang X et al. The role of imaging in 2019 novel coronavirus pneumonia (COVID-19). *Eur Radiol* 2020 April 15
- [14] Liu KC, Xu P, Lu WF, et al. CT manifestations of coronavirus disease-2019: a retrospective analysis of 73 cases by disease severity *Eur J Radiol* 2020;126:108941
- [15] Li K, Wu J, W F et al. The Clinical and Chest CT Features Associated with Severe and Critical COVID-19 Pneumonia. *Invest Radiol* 2020 Jun
- [16] Li J, Chen T, Yang H et al. Clinical and radiological changes of hospitalized patients with COVID-19 pneumonia from disease onset to acute exacerbation: a multicenter paired cohort study. *Computed Tomography* 2020 May 8
- [17] Song J, et al. End-to-end automatic differentiation of the coronavirus disease 2019 (COVID-19) from viral pneumonia based on chest CT, Under review
- [18] Li L, Qin L, Xu Z. et al. Artificial intelligence distinguishes COVID-10 from Community Acquired Pneumonia on Chest CT. *Radiology* 2020
- [19] Zhang K., Liu X., et al. Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements and Prognosis of COVID-19 Pneumonia Using Computed Tomography
- [20] Huang L, Han R, Ai T et al. Serial Quantitative Chest CT Assessment of COVID-19: Deep Learning Approach. *Radiology: Cardiothoracic Imaging* 2020
- [21] Chung M, Bernheim A, Mei X. et al. CT imaging features of 2019 Novel Coronavirus (2019-nCoV). *Radiology* 2020; 295:202-207
- [22] Saleh S, Abedi A, Balakrishnan S, et al. Coronavirus disease 2019 (COVID-19): a systematic review of imaging findings in 919 patients. *AJR Am J Roentgenol* 2020
- [23] Fang Y, Zhang H, Xie J et al. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. *Radiology* 2020
- [24] Kanne JP, Little BP, Chung JH et al. Essentials for Radiologists on COVID-19: An Update- Radiology Scientific Expert Panel. *Radiology* 2020.
- [25] Franquet T. Imaging of pulmonary viral pneumonia. *Radiology* 2011; 260:18-39

- 275 [26] Bernheim A, Mei X, Huang M, et al. Chest CT findings
276 in Coronavirus Disease-19 (COVID-19): relationship
277 to duration of infection. Radiology 2020
- 278 [27] Pan F, Ye T, Sun P, et al. Time course of lung changes
279 on chest CT during recovering from 2019 Novel Coron-
280 avirus (COVID-19) Pneumonia. Radiology 2020
- 281 [28] Wen Z, Chi Y, Zhang L, et al. Coronavirus disease
282 2019: initial detection on chest CT in a retrospective
283 Multi-center study of 103 Chinese subjects. RYCT 2020
- 284 [29] Bai HX, Hsieh B, Xiong MS et al. Performance of ra-
285 diologists in differentiating COVID-19 from viral pneu-
286 monia on chest CT. Radiology, 2020.
- 287 [30] Song F, Shi N, Shan F et al. Emerging Coronavirus
288 2019-nCoV Pneumonia. Radiology.
- 289 [31] Wang Y, Dong C, Hu Y et al. Temporal Changes of
290 CT findings in 90 patients with COVID-19 pneumonia:
291 a longitudinal study. Radiology 2020.
- 292 [32] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam,
293 R., Parikh, D. and Batra, D., 2017. Grad-cam: Visual
294 explanations from deep networks via gradient-based
295 localization. In Proceedings of the IEEE international
300 conference on computer vision (pp. 618-626).
- 301 [33] Tsang, S.H., 2018. Review: GoogLeNet (Inception
302 v1)—Winner of ILSVRC 2014 (Image Classification).
- 303 [34] Carreira, J. and Zisserman, A., 2017. Quo vadis, ac-
304 tion recognition? a new model and the kinetics dataset.
305 In proceedings of the IEEE Conference on Computer
306 Vision and Pattern Recognition (pp. 6299-6308).
- 307 [35] Kingma, D.P. and Ba, J., 2014. Adam: A
308 method for stochastic optimization. arXiv preprint
309 arXiv:1412.6980.
- 310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329