



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Deep learning analysis provides accurate COVID-19 diagnosis on chest computed tomography

D. Javor^a, H. Kaplan^b, A. Kaplan^b, S.B. Puchner^{a,*}, C. Krestan^c, P. Baltzer^a

^a Department of Biomedical Imaging and Image-guided Therapy, Medical University of Vienna, Vienna, Austria

^b Deepinsights Study Group for Artificial Intelligence, Vienna, Austria

^c Department of Radiology, Sozialmedizinisches Zentrum Süd – Kaiser-Franz-Josef Spital, Vienna, Austria

ARTICLE INFO

Keywords:

Computed tomography
Coronavirus disease 2019
Deep learning

ABSTRACT

Introduction: Computed Tomography is an essential diagnostic tool in the management of COVID-19. Considering the large amount of examinations in high case-load scenarios, an automated tool could facilitate and save critical time in the diagnosis and risk stratification of the disease.

Methods: A novel deep learning derived machine learning (ML) classifier was developed using a simplified programming approach and an open source dataset consisting of 6868 chest CT images from 418 patients which was split into training and validation subsets. The diagnostic performance was then evaluated and compared to experienced radiologists on an independent testing dataset. Diagnostic performance metrics were calculated using Receiver Operating Characteristics (ROC) analysis. Operating points with high positive (>10) and low negative (<0.01) likelihood ratios to stratify the risk of COVID-19 being present were identified and validated.

Results: The model achieved an overall accuracy of 0.956 (AUC) on an independent testing dataset of 90 patients. Both rule-in and rule out thresholds were identified and tested. At the rule-in operating point, sensitivity and specificity were 84.4 % and 93.3 % and did not differ from both radiologists ($p > 0.05$). At the rule-out threshold, sensitivity (100 %) and specificity (60 %) differed significantly from the radiologists ($p < 0.05$). Likelihood ratios and a Fagan nomogram provide prevalence independent test performance estimates.

Conclusion: Accurate diagnosis of COVID-19 using a basic deep learning approach is feasible using open-source CT image data. In addition, the machine learning classifier provided validated rule-in and rule-out criteria could be used to stratify the risk of COVID-19 being present.

1. Introduction

Since the outbreak of corona virus disease 19 (COVID-19) and its declaration by the WHO as a pandemic on March 11, 2020, various measures have been implemented worldwide in order to achieve containment of the disease [1–3]. One key factor has been proven to be the rapid and accurate identification of infected patients and their isolation among with social distancing for the general population [4,5]. While reverse-transcription polymerase chain reaction (RT-PCR) remains up to date the main diagnostic tool, the role of computed tomography (CT) of the chest as a complementary method or even a reliable alternative is increasingly recognized, given the vast numbers of potential carriers of the disease [6–10].

Artificial intelligence with the use of deep learning technology has

been sought to improve the capability of CT in terms of distinguishing typical COVID-19 features from other types of pneumonia [11,12]. The use of this kind of technology could expedite the identification of diseased patients and further improve the risk stratification in cases with indeterminate findings on CT in absence of other diagnostic methods.

Recently, a study was published reporting a high accuracy of a newly developed machine learning (ML) model for detecting COVID-19 on chest CT images [11] utilizing extensive preprocessing (lung segmentation) and a large dataset. The aim of the present study was to evaluate if a robust deep learning derived classifier can be built with a simplified programming approach without extensive preprocessing and a smaller open source dataset and to compare such a model to experienced radiologists in terms of diagnostic accuracy.

* Corresponding author at: Department of Biomedical Imaging and Image-guided Therapy, Medical University of Vienna, Waehringer Guertel 18–20, A-1090 Vienna, Austria.

E-mail address: stefan.puchner@meduniwien.ac.at (S.B. Puchner).

<https://doi.org/10.1016/j.ejrad.2020.109402>

Received 20 May 2020; Received in revised form 13 October 2020; Accepted 2 November 2020

Available online 4 November 2020

0720-048X/© 2020 Elsevier B.V. All rights reserved.

2. Methods

2.1. Patients

We assembled a dataset of 6868 images of CT chest exams from 418 patients from public sources (see supplemental material). 3102 images from 209 patients were labeled as COVID-19-positive according to the information provided with the images; only exams from patients that were unequivocally categorized and described as COVID-19-positive (indicating a positive RT-PCR result) by the providing source were included. 3766 images from 209 patients were labeled COVID-19-negative and included other lung pathologies, such as lobar bacterial pneumonias, atypical or viral pneumonias, lung cancer, organizing pneumonia, infectious bronchiolitis and other diseases. To avoid class imbalance within the training data resulting in over-classification of the majority group (non-Covid19) due to its increased prior probability [13], the datasets were intentionally balanced (50:50 ratio). The patient disease statistics are summarized in Table 1. All exams with inferior image quality or uncertain COVID-19-status were excluded.

From the entire dataset, an independent testing set was assembled containing 90 images of 90 patients. Forty-five COVID-19-positive patients were selected randomly and 45 negative patients were selected manually to ensure a similar distribution of diseases as in the training dataset and to avoid an insufficient amount of infectious diseases similar to COVID-19, such as virus pneumonias, due to randomisation. The training dataset (6778 images of 328 patients) was then further split for training the model and internal validation (20 % of the samples). The independent testing set was not used for training nor for internal validation. The validation set was used for internal validation but not for training.

2.2. Data processing

After selection, the validation and test datasets were not presented to the DL model in the training phase. The input sources originally were plain png and jpeg files. Out of these, 82.8 % were jpeg, while the rest (17.3 %) were png files. In order to homogenise the data and minimize any bias due to compression artefacts, all png files were converted to jpeg files.

Table 1
Distribution of COVID-19 Cases and non-COVID-19 Diseases.

| Training and Validation Dataset (n patients) | Disease | Test Dataset (n patients) |
|--|---|---------------------------|
| 165 | COVID-19 | 45 |
| 12 (7 %) | Bacterial Pneumonia | 8 (18 %) |
| 4 (2 %) | Pneumocystis Pneumonia | 2 (4 %) |
| 2 (1 %) | Viral Pneumonia | 5 (11 %) |
| 8 (5 %) | Other infectious diseases (Klebsiella pneumonia, Bronchiolitis, mycobacterium avium complex infection, fungal pneumonia, etc.) | 15 (33 %) |
| 1 (0,6 %) | Abscess | 5 (11 %) |
| 21 (13 %) | Lung cancer and metastasis | 9 (20 %) |
| 5 (3 %) | Idiopathic pulmonary fibrosis | |
| 8 (5 %) | Organizing Pneumonia | |
| 3 (2 %) | Lipoid Pneumonia | |
| 5 (3 %) | Diffuse pulmonary haemorrhage and trauma | 1 (2 %) |
| 5 (3 %) | Hypersensitivity pneumonitis | |
| 5 (3 %) | Pulmonary edema | |
| 2 (1 %) | ARDS | |
| 86 (51 %) | Other (Aspergillosis, Sarcoidosis, Histioplasmosis, Silicosis, eosinophilic pneumonia, Granulomatosis with polangitis, drug induced lung disease, Giant cell interstitial pneumonia, Cystic fibrosis, etc.) | |

2.3. Deep learning model

The Deep Learning model is able to extract 2D features from the submitted images. The Convolutional Neuronal Network was a plain, default ResNet50 [14]. Images were standardized to 448×448 pixels. The random seed was set to 43 and kept constant in order to reproduce our results. We used only the default transformations and augmentations offered by the fastai2 [15] Deep Learning framework. All hyper parameters are provided in supplemental material section II.

We used 17 epochs of training via the fastai2 library on a single Nvidia Tesla GPU with 16 GB of VRAM. Batch size was set to 32. As opposed to other approaches no prior lung segmentation or pre-processing was performed reducing the required computational time to a minimum and increasing significantly the feasibility.

To guarantee an open science approach, the trained model can be downloaded and used with a sample script (inference.py). The training python scripts are also released as open source on github.

To evaluate a possible bias from recognition of the various images sources, a separate ML model was developed with the same architecture and hyperparameters as the original model to identify the image sources.

2.4. Evaluation of radiologist interpretive performance

The trained ML classifier was compared to two radiologists, both having more than 15 years of experience (each), on the same independent test dataset. The testing dataset cases were presented to both radiologists independently in different reading sessions in random order. The readers were neither aware of the diagnoses nor of the prevalence of positive cases. They used diagnostic criteria known at the time of reading [16,17] and assigned a diagnosis (COVID-19 positive vs negative).

2.5. Statistics

The diagnostic performance of the ML model classifier was calculated using ROC statistics. The area under the ROC curve (AUC) was used as the measure of diagnostic accuracy for this study. The null hypothesis for diagnosis of COVID-19 was pure chance (an AUC of 0.5). To determine exploratory thresholds to either rule-in or rule-out COVID-19, operating points on the ROC curve of the ML classifier in the validation dataset were chosen that lead to a positive likelihood ratio (LR+) of >10 (rule-in criterion for COVID-19) or a negative likelihood ratio (LR-) of <0.01 (rule-out criterion for COVID-19). These thresholds were then applied to the independent testing dataset. The areas under the ROC curves (AUC) of the ML classifier and the two radiologists were compared with each other using the DeLong method. Further, sensitivity and specificity at these thresholds were compared to the classification results of both radiologists using McNemar tests. P-values <0.05 were considered significant. No alpha error correction procedure was applied as all statistics are exploratory in nature. To exemplify the effect of positive and negative test results, pre-test probabilities were mapped to post-test probabilities using the likelihood ratios in a Fagan nomogram. The nomogram can be used as a graphic calculator to map any given pre-test probability to a post-test probability using LR+ or LR- for a positive or negative test result [18].

3. Results

3.1. Open access ML classifier

The trained ML model is available to the general public (<https://labs.deep-insights.ai>). The server running the web page is an IBM x3550 M2, with an Intel(R) Xeon(R) CPU X5460 @ 3.16 GHz (an old CPU from 2007). Inference on this approx.10 year old, low powered system was fast: 1.2 s per image. The ML classifier trained to identify whether the

image source presented a source of bias resulted in a correct classification rate of the image sources of 61 % (with 50 % being chance), indicating that such a bias based on the source of the images can be largely excluded.

The diagnostic performance metrics of the ML model and the human readers (radiologists R1 and R2) for detection of COVID-19 are summarized in Fig. 1 and Table 2. The area under the ROC curve was numerical higher than that of both readers but did only prove statistically significant superiority compared to R1 (8.9 % difference, 95 %-CI 0.7–17 %, $p = 0.033$) but not compared to R2 (6.7 % difference, 95 %-CI -1.2–14.5 %, $p = 0.097$). R1 and R2 did not differ significantly from each other ((2.2 % difference, 95 %-CI -7.6–12.1 %, $p = 0.659$). While the radiologists test response was dichotomous, the metric character of the ML classifier results allowed the definition for rule-in and rule-out thresholds on the validation dataset which were then applied to the independent testing dataset (Fig. 1, Table 2). The rule-in threshold coincided with the automatically selected threshold based on the Youden index. Positive likelihood ratios >10 were present at this threshold (predicted probability of COVID-19 by the ML classifier of >40 %) for both validation and testing dataset (Table 2). The negative likelihood ratio at the rule-out threshold was <0.01 for both datasets (Table 2).

Comparison of ML classifier sensitivity and specificity with radiologists' performance

Dichotomous radiologists' performance in terms of sensitivity (Radiologist 1: 80 %, Radiologist 2: 82.2 %) and specificity (Radiologist 1: 91.1 %, Radiologist 2: 97.8 %) did not differ from the ML classifier (sensitivity: 84.4 %, specificity 93.3 %) at the rule-in threshold ($p > 0.05$, respectively) while sensitivity was significantly inferior and specificity significantly superior at the rule-out threshold (sensitivity and specificity of the ML classifier: 100 % and 60 %, $p < 0.05$, respectively, Table 2).

Prevalence independent estimation of ML classifier performance

The likelihood ratios calculated within the independent testing dataset can be mapped to post-test probabilities given pre-test probabilities based on the prevalence of COVID-19 within the literature [19] in a Fagan-Nomogram (example given in Fig. 2) for both rule-in and rule-out thresholds.

4. Discussion

In this study, we developed a novel deep learning derived classifier for detecting COVID-19 on CT-images. Although omitting the pre-processing step of lung segmentation and utilizing an open source

dataset that is rather small for deep learning methodology, the model achieved high accuracy with an AUC of 0.956 on an independent testing dataset. We compared the novel ML classifier with two radiologists, both having more than 15 years of experience (each), on the same test dataset and found a minor overall superior diagnostic accuracy in terms of the AUC. As opposed to the dichotomous decision by the radiologists, the continuous character of the ML classifier prediction results allowed us to identify thresholds that were highly sensitive or specific for diagnosis of COVID-19. This is of high clinical relevance since decision making requires a clear-cut decision between cases that have a high, medium or low level of suspicion. Using simple Bayes theorem in a Fagan nomogram, one can map various pre-test probabilities based on the local prevalence of COVID-19 to post-test probabilities using the likelihood ratios at the respective thresholds. By establishing explorative rule-in and rule-out thresholds for low and high predictive probabilities, this enables a "traffic light"-system for patient triage that could be combined with clinical parameters. If positive, the patient should be isolated as a consequence until (multiple) RT-PCR tests have confirmed or rejected the predicted diagnosis. If negative, the post-test probability is low enough to practically rule out COVID-19 as the suspected underlying lung disease.

RT-PCR to diagnose COVID-19 has its own limitations: the test is not universally available, turnaround times can be lengthy, and reported sensitivities vary [20]. The combination of chest CT and ML has the advantage of obtaining an immediate result – practically after image reconstruction and while the patient is still on the examination bed – and independent of the presence of the radiologist [21–24]. This scenario could potentially play a role in either minor peripheral hospitals or places with radiological personnel shortage due to various reasons including a high load of COVID-19 cases and could also act as a support for inexperienced radiologists during night-duty. In the case of staff shortages and overwhelming patient loads an initial reading from an AI-based tool could dramatically reduce waiting times and facilitate a rapid risk stratification. A specialist could follow up the AI system's reading with a more thorough diagnosis later [25]. Since our ML algorithm is published as open source and only open source images have been utilised to build it, easy access to a fast preliminary diagnosis can be offered.

This has various applications from low-income countries to acute pressing patient numbers or staff shortage.

Compared to a recently published study which was trained on a much larger dataset [11] our results are only slightly inferior in terms of diagnostic accuracy though the results cannot directly be compared due

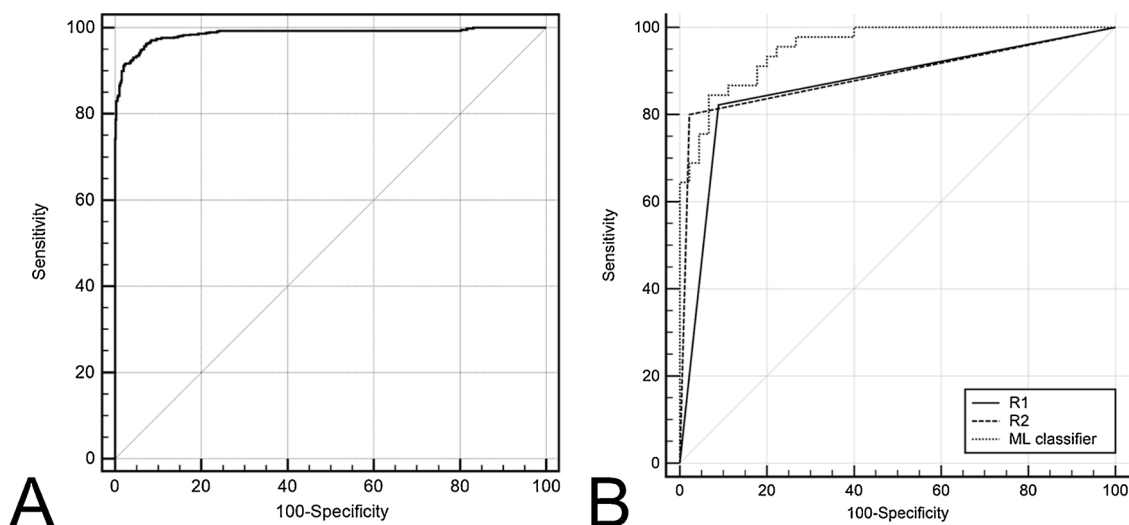


Fig. 1. ROC curves of the ML model classifier performance in the validation (A) and the independent testing (B) dataset. The dichotomous performance of both radiologists (R1, R2) in the test dataset is plotted as a comparator in B. Diagnostic performance metrics are given in Table 2.

Table 2
diagnostic performance indices of the ML model classifier in validation and testing datasets compared with two radiologists performance.

| | | AUC | Threshold | Sensitivity% | Specificity% | LR+ | LR- |
|--------------|------------------------------|---------------------|--------------------|--------------|--------------|------|-------|
| ML Model | Validation Dataset | 0.986 (0.978–0.992) | Rule-out (>0.0006) | 99.3 | 75.8 | 4.1 | 0.009 |
| | | | Rule-in (>0.4) | 92.2 | 96.3 | 25.0 | 0.081 |
| | Test Dataset | 0.956 (0.890–0.988) | Rule-out (>0.0006) | 100 | 60 | 2.5 | <0.01 |
| Radiologists | | | Rule-in (>0.4) | 84.4 | 93.3 | 12.7 | 0.17 |
| | Radiologist 1 (Test Dataset) | 0.867 (0.779–0.929) | n.a. | 82.2 | 91.1 | 9.25 | 0.2 |
| | Radiologist 2 (Test Dataset) | 0.889 (0.805–0.945) | n.a. | 80 | 97.8 | 36.0 | 0.2 |

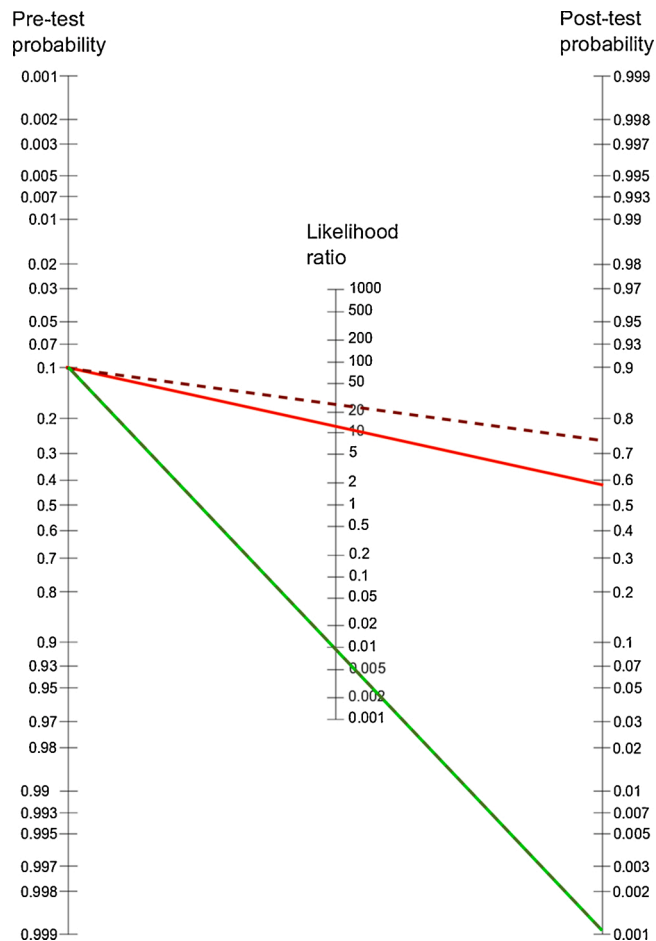


Fig. 2. Fagan nomogram illustrating the ML classifier performance in case of a pretest probability of 10 %. Dashed lines represent validation sample and straight lines independent test sample results. A positive result exceeding the rule-in COVID-19 threshold is indicated by the black lines while a negative result below the rule-out threshold for COVID-19 is indicated by the grey lines.

to different datasets. Nevertheless, the diagnostic accuracy was comparable and even slightly superior compared to experienced radiologists and in contrast to the above mentioned study our approach required no complex preprocessing with lung segmentation reducing the required time to a minimum and increasing the overall feasibility significantly.

Further developments could directly be compared to our ML classifier as all data used within this study is free of access and the ML model itself is also provided without access restrictions. Our results indicate that the complex preprocessing step of lung segmentation may be omitted while still achieving high sensitivity and specificity. ResNet-50 was used as the backbone residual neural network architecture. Experiments with architectures with more layers (ResNet-101) did not yield better results.

Although showing an acceptable accuracy, one must be well aware of the limitations of the method which combines CT and machine learning.

CT has shown a high sensitivity in detecting COVID-19 lesions [6]. On the other hand, COVID-19 shows a broad spectrum of possible CT features and demonstrates a heterogeneous CT appearance at different stages [26], from single or multiple scattered patchy or conglomerate ground glass opacities as the distinctive feature in the early stages to diffuse consolidation of the lungs of varying density and areas of fibrosis in severe advanced and dissipation stages. We have noted that most cases of false-negatives are cases of advanced-stage COVID-19 and most cases of false-positives are cases of non-covid pathologies with very similar features to COVID-19, such as ground-glass opacities (Fig. 3). Inui and colleagues [27] reviewed CT scans of 112 cases of RT-PCR-confirmed COVID-19 from the Diamond Princess cruise ship. Less than two-thirds (61 %) of cases had lung opacities on CT; 20 % of symptomatic patients had negative CT's, suggesting that the CT findings studied (e.g. ground-glass opacities, consolidations) are not specific for COVID-19. The limitations of CT are the limitation for such ML models in general since CT is providing the dataset. The general accuracy of ML models should always be interpreted with care as it reflects the average performance within the whole dataset and does not reflect difficulties in certain subsets. Likely, the diagnostic performance will be high if only early-stage images of COVID-19 patients are provided (with ground glass as the distinctive feature) in one group and only pathologies that have non-similar features, such as lobar pneumonias, tumors and fibrosis, in the control group. The more cases of advanced stage COVID-19 (with consolidation and fibrosis) and similar non-COVID-19 pathologies, especially those featuring ground-glass opacities, are added to the system the lower the accuracy will be, simply because some diseases can have the same features as COVID-19, e.g. some cases of atypical pneumonia, lung toxicity, organizing pneumonia, acute eosinophilic pneumonia and diffuse alveolar haemorrhage. In our study, we included such COVID-19-similar pathologies in the training, validation and test dataset.

However, it is not excluded that neural networks can extract features of COVID-19 and COVID-19-similar pathologies on CT images that are invisible to the human eye and that would permit a significantly better differentiation between the two groups. In order to test this hypothesis a much larger dataset would be necessary and further ML techniques such as multi-view fusion [28] could be applied.

The specific limitations of our study are on one hand the heterogeneous nature of the dataset given the multiple sources and the rather small number of images. On the other hand, the ML model currently does not analyze entire 3D datasets of DICOM images at the time of publication. The reason was that the majority of the open source data was available as non-DICOM images only containing the slices with pathological alterations of the lungs. Expanding the model to work with volumes however is well feasible with expectedly comparable results in terms of COVID-19 detection as a fully negative CT slice (no opacities/densities) is highly unlikely to harbor COVID-19. Nevertheless, in clinical practice 2D ML models are relevant with the application of assessing pathological alterations on a single slice regarding the probability of COVID-19.

Although our results indicate that the complex preprocessing step of lung segmentation may be omitted, the lack of lung segmentation can also lead to a possible bias due to extra thoracic objects, such as intubation tubes of intensive care patients. While this cannot be excluded

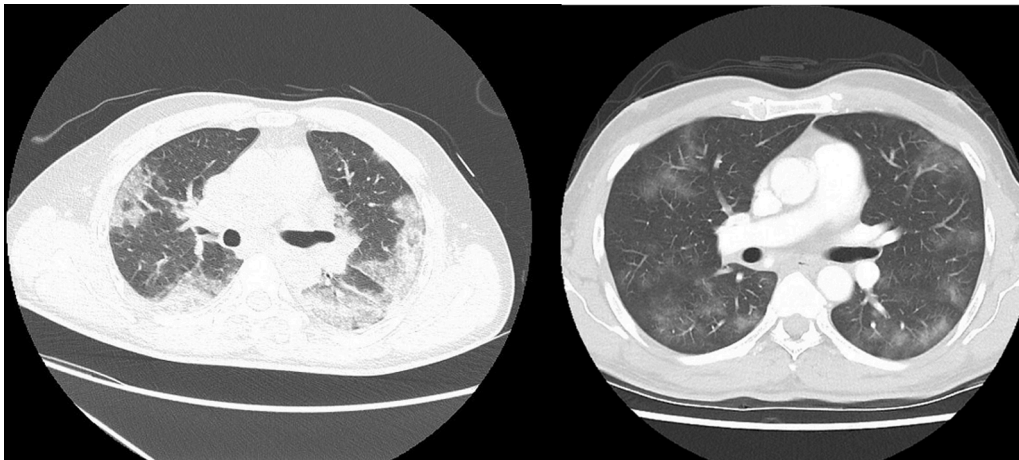


Fig. 3. Advanced stage COVID-19 with consolidations on the left which was falsely diagnosed as non-COVID-19 (false-negative); Pneumocystis carinii pneumonia on the right which was falsely diagnosed as COVID-19 (false-positive).

entirely, only a small percentage of patients in the COVID-19 and the control group showed any kind of extrathoracic objects.

Finally, the balanced training dataset with a high prevalence of COVID-19 cases yields sensitivity and specificity estimates that are potentially affected towards higher sensitivity and lower specificity [29]. While this does less likely affect the ML classifier but rather human readers, the effect of prevalence on predictive values is of a more direct mathematical nature [30]. Consequently, we determined the operating points to either rule-in or rule-out malignancy based on prevalence-independent likelihood ratios together with a Fagan-nomogram to adapt the ML classifier results to different clinical settings [31].

Detection of COVID-19 with the help of ML using CT images is feasible with an acceptable accuracy even with a relatively small amount of open-source data. Such a ML model can play a supporting role as an automated triage tool for a preliminary diagnosis although the limitations have to be kept in mind. With further open data sources appearing as COVID-19 progresses, the ML model can be improved even further.

CRedit authorship contribution statement

D. Javor: Conceptualization, Writing - original draft, Data curation, Methodology, Software, Visualization. **H. Kaplan:** Software, Methodology. **A. Kaplan:** Software, Conceptualization, Writing - original draft. **S.B. Puchner:** Writing - original draft. **C. Krestan:** Validation, Writing - review & editing. **P. Baltzer:** Writing - original draft, Writing - review & editing, Formal analysis, Visualization.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.ejrad.2020.109402>.

References

- [1] The Lancet, COVID-19: learning from experience, *Lancet* 395 (2020).
- [2] Y. Zhao, C. Xiang, S. Wang, C. Peng, Q. Zou, J. Hu, Radiology department strategies to protect radiologic technologists against COVID-19: experience from Wuhan, *Eur. J. Radiol.* 127 (2020).
- [3] R.C. Chen, T.T. Tan, L.P. Chan, Adapting to a new normal? 5 key operational principles for a radiology service facing the COVID-19 pandemic, *Eur. Radiol.* (2020).
- [4] J. Cohen, K. Kupferschmidt, Strategies shift as coronavirus pandemic looms, *Science* 367 (2020).
- [5] A. Wilder-Smith, D.O. Freedman, Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak, *J. Travel Med.* 27 (2020).
- [6] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, 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).
- [7] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, L.P. Pang, W. Ji, Sensitivity of chest CT for COVID-19: comparison to RT-PCR, *Radiology* (2020).
- [8] Q. Yang, Q. Liu, H. Xu, H. Lu, S. Liu, H. Li, Imaging of coronavirus disease 2019: a Chinese expert consensus statement, *Eur. J. Radiol.* (2020).
- [9] H. Wang, R. Wei, G. Rao, J. Zhu, B. Song, Characteristic CT findings distinguishing 2019 novel coronavirus disease (COVID-19) from influenza pneumonia, *Eur. Radiol.* (2020).
- [10] X. Chen, Y. Tang, Y. Mo, S. Li, D. Lin, Z. Yang, et al., A diagnostic model for coronavirus disease 2019 (COVID-19) based on radiological semantic and clinical features: a multi-center study, *Eur. Radiol.* (2020).
- [11] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, et al., Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT, *Radiology* (2020).
- [12] H.X. Bai, R. Wang, Z. Xiong, B. Hsieh, K. Chang, K. Halsey, et al., AI augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other etiology on chest CT, *Radiology* (2020).
- [13] J.M. Johnson, T.M. Khoshgoftaar, Survey on deep learning with class imbalance, *J. Big Data* 6 (2019) 27, <https://doi.org/10.1186/s40537-019-0192-5>.
- [14] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. Deep Residual Learning for Image Recognition. <https://arxiv.org/abs/1512.03385>.
- [15] J. Howard, S. Gugge, Fastai: a layered API for deep learning, *Information* 11 (2) (2020) 108, <https://doi.org/10.3390/info11020108>. February[Online]. Available: .
- [16] S. Zhou, Y. Wang, T. Zhu, L. Xia, CT features of coronavirus disease 2019 (COVID-19) pneumonia in 62 patients in Wuhan, China, *AJR Am. J. Roentgenol.* (2020).
- [17] S. Salehi, A. Abedi, S. Balakrishnan, A. Gholamrezaeezad, Coronavirus disease 2019 (COVID-19): a systematic review of imaging findings in 919 patients, *AJR Am. J. Roentgenol.* (2020).
- [18] T.J. Fagan, Nomogram for bayes theorem, *N. Engl. J. Med.* 293 (1975).
- [19] H. Kim, H. Hong, S.H. Yoon, Diagnostic performance of CT and reverse transcriptase-polymerase chain reaction for coronavirus disease 2019: a meta-analysis, *Radiology* (2020).
- [20] M.D. Hope, C.A. Raptis, A. Shah, M.M. Hammer, T.S. Henry, A role for CT in COVID-19? What data really tell us so far, *Lancet* (2020).
- [21] G. Chassagnon, M. Vakalopoulou, N. Paragios, M.P. Revel, Artificial intelligence applications for thoracic imaging, *Eur. J. Radiol.* 123 (2020), 108774, <https://doi.org/10.1016/j.ejrad.2019.108774>. February[Epub 2019. Dec 11].
- [22] J. Gong, J. Liu, W. Hao, S. Nie, B. Zheng, S. Wang, et al., A deep residual learning network for predicting lung adenocarcinoma manifesting as ground-glass nodule on CT images, *Eur. Radiol.* (2020).
- [23] M. Sandstedt, L. Henriksson, M. Janzon, G. Nyberg, J. Engvall, J. De Geer, Evaluation of an AI-based, automatic coronary artery calcium scoring software, *Eur. Radiol.* (2020).
- [24] X. Li, J.H. Thrall, S.R. Digumarthy, M.K. Kalra, P.V. Pandharipande, B. Zhang, et al., Deep learning-enabled system for rapid pneumothorax screening on chest CT, *Eur. J. Radiol.* (2019).
- [25] K. Hao, Doctors are using AI to triage covid-19 patients. The tools may be here to stay, *MIT Technology Review*.
- [26] W. Yang, A. Sirajuddin, X. Zhang, G. Liu, Z. Teng, S. Zhao, et al., The role of imaging in 2019 novel coronavirus pneumonia (COVID-19), *Eur. Radiol.* (2020).

- [27] S. Inui, A. Fujikawa, M. Jitsu, N. Kunishima, S. Watanabe, Y. Suzuki, et al., Chest CT findings in cases from the cruise ship "Diamond Princess" with coronavirus disease 2019 (COVID-19), *Radiology* (2020) (published online March 17.).
- [28] X. Wu, H. Hui, M. Niu, L. Li, L. Wang, B. He, et al., Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: a multicentre study, *Eur. J. Radiol.* 128 (May 5) (2020), 109041.
- [29] H. Brenner, O. Gefeller, Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence, *Stat. Med.* (1997).
- [30] G. Altman Douglas, Bland J. Martin, Statistics notes: diagnostic tests 2: predictive values, *BMJ* (1994).
- [31] T.J. Fagan, Letter: nomogram for bayes theorem, *N. Engl. J. Med.* (1975).