An externally validated fully automated deep learning algorithm to classify COVID-19 and other pneumonias on chest computed tomography

Akshayaa Vaidyanathan[1,2,7], Julien Guiot[3,7], Fadila Zerka[1,2], Flore Belmans[1], Ingrid Van Peufflik[1], Louis Deprez[4], Denis Danthine[4], Gregory Canivet[5], Philippe Lambin[2], Sean Walsh[1], Mariaelena Occhipinti[1], Paul Meunier[4], Wim Vos[1], Pierre Lovinfosse[6] and Ralph T.H. Leijenaar[1]

[1]Radiomics (Oncoradiomics SA), Liège, Belgium. [2]The D-Lab, Depts of Precision Medicine, and Radiology and Nuclear Medicine, GROW–School for Oncology, Maastricht University, Maastricht, The Netherlands. [3]Dept of Pneumology, University Hospital of Liège, Liège, Belgium. [4]Dept of Radiology, University Hospital of Liège, Liège, Belgium. [5]Dept of Computer Applications, University Hospital of Liège, Liège, Belgium. [6]Dept of Nuclear Medicine and Oncological Imaging, University Hospital of Liège, Liège, Belgium. [7]These authors have contributed equally to this work and share first authorship.

Corresponding author: Akshayaa Vaidyanathan (akshayaa.vaidyanathan@radiomics.bio)

Shareable abstract (@ERSpublications)
A fully automated artificial intelligence-based network is proposed to classify CT volumes of patients affected with COVID-19 or influenza/CAP, and in the uninfected https://bit.ly/3MJrVRi

Cite this article as: Vaidyanathan A, Guiot J, Zerka F, *et al*. An externally validated fully automated deep learning algorithm to classify COVID-19 and other pneumonias on chest computed tomography. *ERJ Open Res* 2022; 8: 00579-2021 [DOI: 10.1183/23120541.00579-2021].

Received: 8 Oct 2021
Accepted: 4 March 2022

Abstract

Purpose In this study, we propose an artificial intelligence (AI) framework based on three-dimensional convolutional neural networks to classify computed tomography (CT) scans of patients with coronavirus disease 2019 (COVID-19), influenza/community-acquired pneumonia (CAP), and no infection, after automatic segmentation of the lungs and lung abnormalities.

Methods The AI classification model is based on inflated three-dimensional Inception architecture and was trained and validated on retrospective data of CT images of 667 adult patients (no infection n=188, COVID-19 n=230, influenza/CAP n=249) and 210 adult patients (no infection n=70, COVID-19 n=70, influenza/CAP n=70), respectively. The model's performance was independently evaluated on an internal test set of 273 adult patients (no infection n=55, COVID-19 n= 94, influenza/CAP n=124) and an external validation set from a different centre (305 adult patients: COVID-19 n=169, no infection n=76, influenza/CAP n=60).

Results The model showed excellent performance in the external validation set with area under the curve of 0.90, 0.92 and 0.92 for COVID-19, influenza/CAP and no infection, respectively. The selection of the input slices based on automatic segmentation of the abnormalities in the lung reduces analysis time (56 s per scan) and computational burden of the model. The Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) score of the proposed model is 47% (15 out of 32 TRIPOD items).

Conclusion This AI solution provides rapid and accurate diagnosis in patients suspected of COVID-19 infection and influenza.

Introduction

Imaging with computed tomography (CT) plays a central role in the diagnosis of respiratory diseases [1, 2]. Since the outbreak of coronavirus disease 2019 (COVID-19) in 2020, more emphasis has been given to the different types of pneumonias and to the distinctive features of COVID-19 from all others [3, 4]. Viral pneumonias, either COVID-19 or others, can all present with reticulation, ground-glass opacities and consolidations at chest CT scan, creating a challenge for radiologists in their routine differential diagnosis. Previous studies on the performance of radiologists in discriminating between COVID-19 and other pneumonias on chest CT scans have shown high variability in both sensitivity (73–94%) and specificity

(24–100%), with, on average, high sensitivity and moderate specificity [5]. This variability of interpretation of CT findings of pneumonia still creates a routine challenge for clinicians in their differential diagnosis, which is key to properly treat the patients and prevent infection spread during pandemics and in the next future.

In this context, the development of innovative artificial intelligence (AI) imaging solutions to support radiologists in swift and precise differential diagnosis would be invaluable. Convolutional neural networks (CNN) have shown great potential in detection, segmentation and classification tasks in radiological images [6]. A recent study demonstrated the application of CNN for differentiation between influenza, COVID-19 and no infection on chest CT scans, with an overall accuracy of 86.7%. The proposed method incorporated training on image patches extracted from CT volumes where each image patch required manual labelling as "pneumonia" or "irrelevant information" [7]. Another study compared the performance of different AI models in classifying COVID-19 from other atypical and viral pneumonias, showing 99.5% accuracy in classifying COVID-19 [8]. However, these approaches involve all manual detection (*i.e.* drawing boxes around the lesions), labelling of the lesions in all the slices and training the models on the patches of detected lesions and manual labels. The time required to perform these manual operations is usually not considered when addressing the real-world application of these models and probably represent one of the major hurdles to widespread clinical adoption.

A fully automatic tool running on chest CT images for the differential diagnosis of pneumonias can represent an important step forward for decreasing the variability of interpretation among clinicians and speeding up the diagnostic process. This will unburden medical staff and in turn provide better and faster diagnosis for patients, reducing the use of hospital resources. Better allocation of both material and human resources can be essential in a time of crisis, as the COVID-19 pandemic demonstrated with dramatic clarity [9]. To attain this goal, we developed and externally validated a fully automated deep-learning framework with a three-dimensional (3D) CNN, able to classify chest CT scans of patients with COVID-19, influenza/CAP or no infection without manual intervention. Individual AI-based whole lung and lung abnormalities segmentation models were used to pre-process the CT images to train the 3D CNN model and are an integral part of the workflow to assure that only the patients presenting abnormalities in the lung volume are processed by the model, saving time and computational power.

Material and methods

The study was approved by the local ethics committee of CHU-Liège (Liège, Belgium; EC number 116/2020). The institutional review board waived the requirement to obtain written informed consent for this retrospective case series, since all analyses were performed on de-identified (*i.e.* anonymised) data and there was no potential risk to patients.

Subjects

Three cohorts of patients were included retrospectively in this study for model training, validation and testing. Cohorts came from two university hospitals (CHU Sart-Tilman and CHU Notre Dame des Bruyères) in Liège, Belgium. The first cohort (label: COVID-19) consisted of all patients with COVID-19 infection confirmed by reverse transcriptase (RT)-PCR who underwent chest CT imaging before 28 March 2020. The second cohort (label: influenza/community-acquired pneumonia (CAP)) consisted of patients with influenza, parainfluenza or CAP infection confirmed by RT-PCR or positive antigen testing who underwent chest CT imaging between March 2014 and March 2020. The third cohort (label: no infection) consisted of consecutive patients who underwent chest CT imaging during October 2019, with confirmed no infection in the lungs disregarding any other lung disease. The three cohorts were pooled together and randomly split between training, validation and testing sets (figure 1). Additionally, the open-source dataset COVID-CT-MD was used as an external validation set [10]. The final population consisted of 169 RT-PCR confirmed positive COVID-19 cases (from February 2020 to April 2020), 60 CAP cases (from April 2018 to November 2019) and 76 no infection cases (from January 2019 to May 2020): all the patients were treated at the Babak Imaging Center (Tehran, Iran) and labelled by three experienced radiologists.

Imaging scans

In this retrospective study, CT scans of the three cohorts of patients included were acquired from different scanners (Siemens and GE) with diverse reconstruction kernels (soft and sharp). In case of presence of more than one series per case, all the available series were used in training the model (as the reconstruction kernels corresponding to the series were considered as a form of image augmentation). Slice thickness of the scans ranged between 0.5 mm and 2 mm, while pixel spacing was between 1 mm and 2.5 mm. A complete summary of the imaging parameters of both training and external validation set is reported in supplementary table S1.

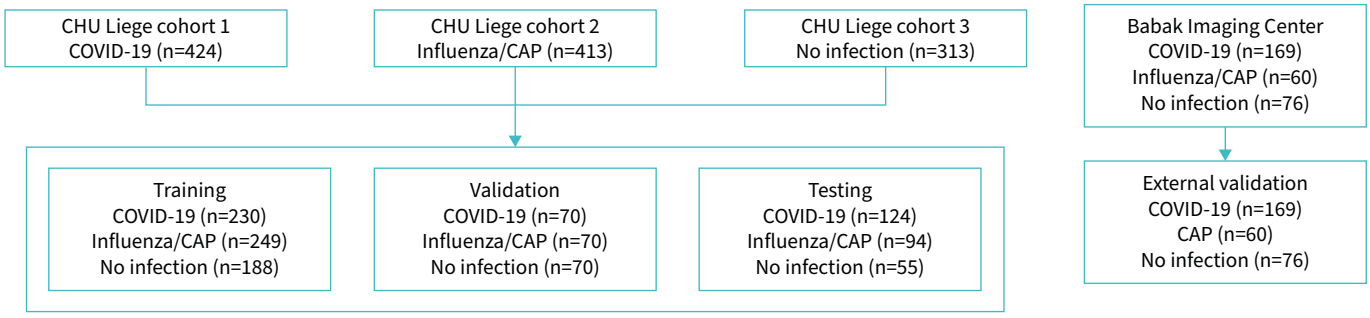

FIGURE 1 Flow chart of patient cohort division. COVID-19: coronavirus disease 2019; CAP: community-acquired pneumonia.

Image pre-processing

The prevalence of COVID-19 cases in the three datasets was adjusted to avoid class imbalance and bias in classification [11]. COVID-19 cases represented 35–45% of the whole cohort for each dataset. A fully automated lung segmentation model (supplementary material: lung segmentation) was used to filter out the slices not containing lungs from the CT scan series. The presence of abnormalities in each filtered slice was confirmed using a lung abnormalities segmentation model (supplementary material: abnormalities segmentation). If no abnormalities were present in the filtered slices, the scan was discarded from model processing. Different sets of 48 consecutive axial slices with an overlap of 10 slices between one set and the other (extracted from the whole volume with axial slices containing lungs) were obtained, while each set including at least one slice containing abnormalities in the lung was used to train the model. The workflow for the pre-processing protocol is depicted in figure 2. The entirety of datasets provided by clinicians were used in the model training and validations, without any prior scan quality selection.

Each data point containing the 48 consecutive axial slices was processed in three different ways to obtain a three-channel input for the model:

1. The first channel (channel 1) contained slices with intensities clipped at lung window level settings (width 1500 HU, length −600 HU) with lungs and the abnormalities cropped.

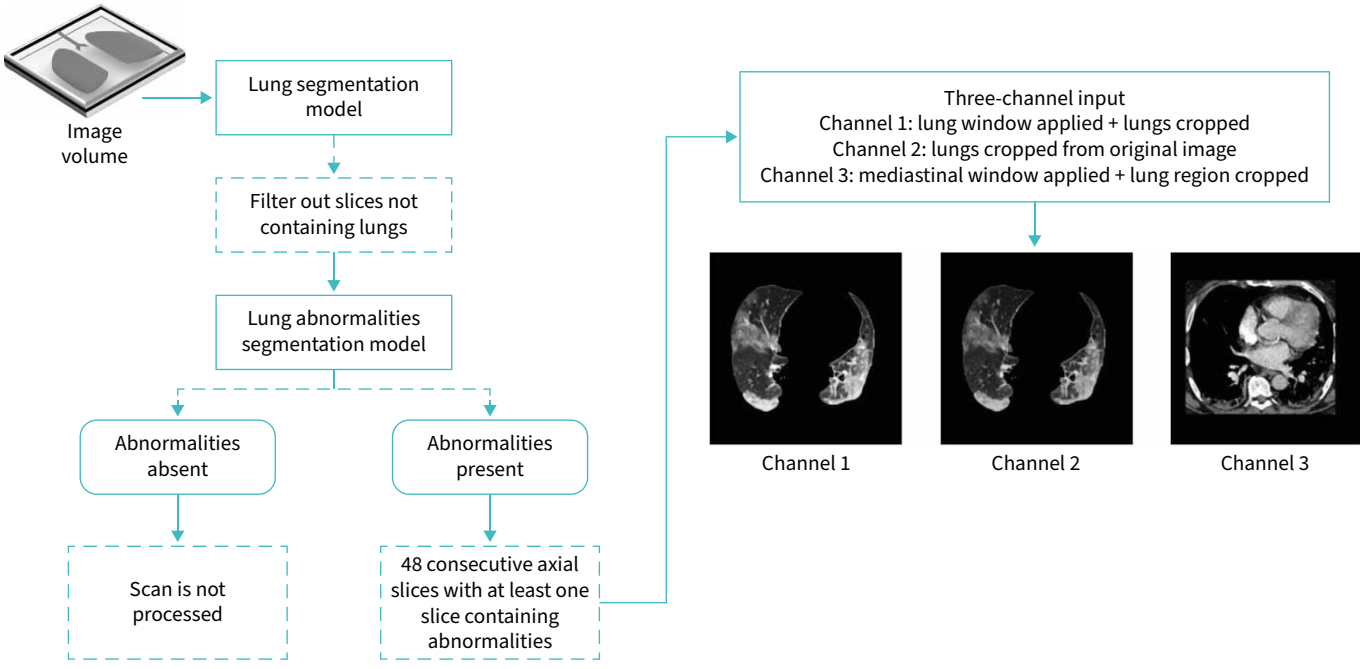

FIGURE 2 Scheme of the pre-processing workflow applied.

2. The second channel (channel 2) contained the slices with the original intensities of lungs along with the abnormalities cropped.
3. The third channel (channel 3) contained slices with intensities clipped at mediastinal window level settings (width 350 HU, length 50 HU) within the region containing the cropped lungs, for which the bounding rectangular crop within which lungs or lung abnormalities pixels are present was obtained. This operation was performed in order to better assess pleural effusion [12].

Finally, the slices were centre-cropped to a slice size of 448×448 pixels. An example of the resulting lung and abnormalities segmentation is reported in figure 3 for an influenza/CAP patient. Two more examples for COVID-19 (supplementary figure S1) and no infection (supplementary figure S2) patients are reported in the supplementary material.

3D CNN architecture

An inflated 3D Inception model [13], pre-trained on a Kinetics dataset [14], was trained on 48 consecutive axial slices as 3D input. Inflated 3D Inception, also known as "two-stream inflated 3D ConvNets", is based on the Inception v1 architecture [15] and consists of inflated filters and pooling kernels into 3D, leading to very deep, naturally spatiotemporal classifiers. The model is trained for five epochs and early stopping was performed after the fifth epoch, as the validation loss started to increase while the training loss decreased, using the categorical cross-entropy loss as an objective function at a batch size of two. A batch size of two was preferred to fit graphics processing unit (GPU) memory of 11 GB. The model was trained on 10 500 data points (which are different sets of 48 consecutive axial slices obtained from the image volume with an overlap to 10 slices between one set and the other) and validated on 6000 data points. The network weights were updated by using an Adam optimiser at a constant learning rate of 1×10^{-4} [16]. The model's architecture is depicted in supplementary figure S3.

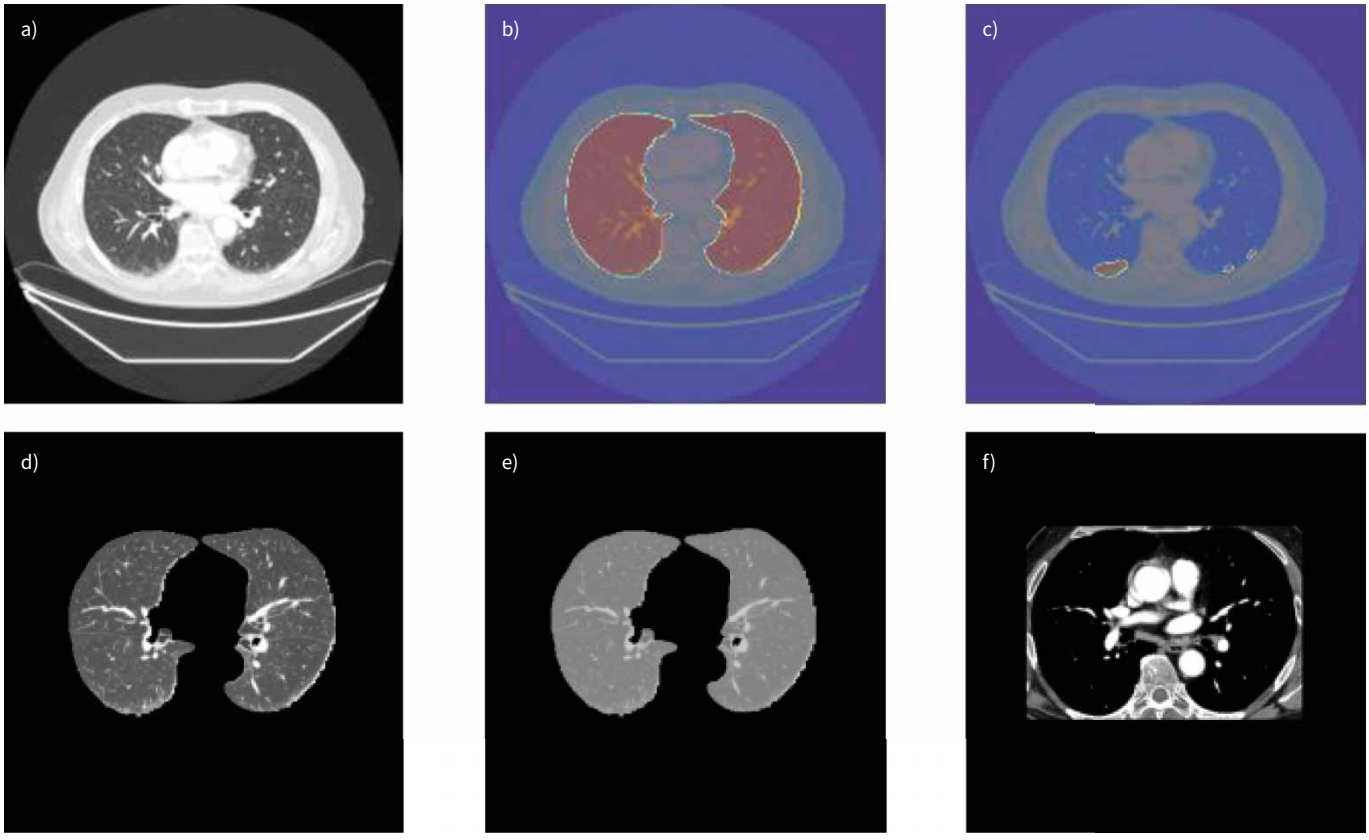

FIGURE 3 Lungs plus abnormalities segmentation on a slice from an influenza/community-acquired pneumonia (CAP) patient. a) Original axial slice from case with influenza/CAP label; b) lung segmentation obtained on the same slice; c) ground-glass opacities segmented by the lung abnormalities model. d–f) Three-channel input obtained from the same slice: d) channel 1; e) channel 2; f) channel 3.

Model prediction

The model's predictions on the probability of each class was obtained on all the 48 consecutive axial slices of the test datasets. The overall class and the overall class probability were computed: if >20% of the predictions correspond to the class COVID-19, then the patient is assigned to that class. If the probabilities for the class influenza/CAP are >20%, then the patient is assigned to the class influenza/CAP. Otherwise, the scan is classified in the no infection class.

Performance metrics

Classification performances of the deep-learning model in the internal testing set and external vaidation set are expressed in term of area under the curve (AUC), specificity and sensitivity. AUC, sensitivity and specificity are calculated for each class by considering the respective class as positive and the rest of the classes as negatives. For instance, AUC of the influenza class is calculated by considering the influenza class as positive and the no infection and COVID-19 classes as negatives. All data elaborations were performed in Python (version 3.6.5) with Keras API. The computation time was calculated on average per scan on the external test set for a RTX 2080 ti 11 GB GPU. The model was evaluated according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) [17] (supplementary material).

Clinical summary report

The clinician is presented with an automatically generated report containing the results of the classification algorithm. The report presents basic patients data (patient's identifier, scan number and scan date) along with the diagnosis (no infection, influenza/CAP, COVID-19) and the probability calculated by the model for each class. In addition, the report shows the 48 consecutive slices with the corresponding lung and lung abnormalities segmentations masks used by the model to make the classification.

Results

Study population

Table 1 lists the study population characteristics for the COVID-19, influenza/CAP and no infection cohorts for the training, validation, internal testing and external validation sets. In the training set, 69% of the COVID-19 patients needed oxygen therapy at admission, with 37% of patients being admitted to the intensive care unit. 17% of COVID-19 patients needed mechanical ventilation and 4% died.

Performance on the internal test set

Model performance is reported in figure 4. The receiver operating characteristic curves for each class (COVID-19, influenza/CAP and no infection) are depicted in figure 4a. The performance for COVID-19 classification in the internal test set has an AUC of 0.91 (95% CI 0.88–0.94) with a sensitivity of 87.90% (109 out of 124) and a specificity of 87.24% (130 out of 149). The influenza/CAP and no infection classes present an AUC of 0.89 (95% CI 0.84–0.93) with a sensitivity of 82.97% (78 out of 94) and a specificity of 89.38% (160 out of 179); and 0.98 (95% CI 0.96–0.99) with a sensitivity of 78.18% (43 out of 55) and specificity of 97.71% (213 out of 218), respectively. The confusion matrix (figure 4b) reports the classification performances (*i.e.* predicted *versus* real values) for each class.

Performance on the external validation set

The lung abnormalities segmentation model identified 19 cases with no abnormalities in the external validation set. These scans were not processed by the deep-learning architecture: performance metrics reported in figure 5a and b are based on the 57 cases from the no infection class which presented abnormalities in the lung. Classification for the COVID-19 class had an AUC of 0.90 (95% CI 0.86–0.94) (sensitivity 83.43%, 141 out of 169; specificity 91.15%, 103 out of 113), while influenza/CAP presented

TABLE 1 Study population characteristics

	Training and validation set	Internal test set	External validation set
Age (years)	63.8±14.44	64.4±15.8	50.67±5.87
Female	48	44	40
Pixel spacing (mm)	0.71±0.10	0.70±0.07	0.67±0.07
Slice thickness (mm)	1.19±0.61	1.19±0.59	2±0

Data are presented as mean±SD or %.

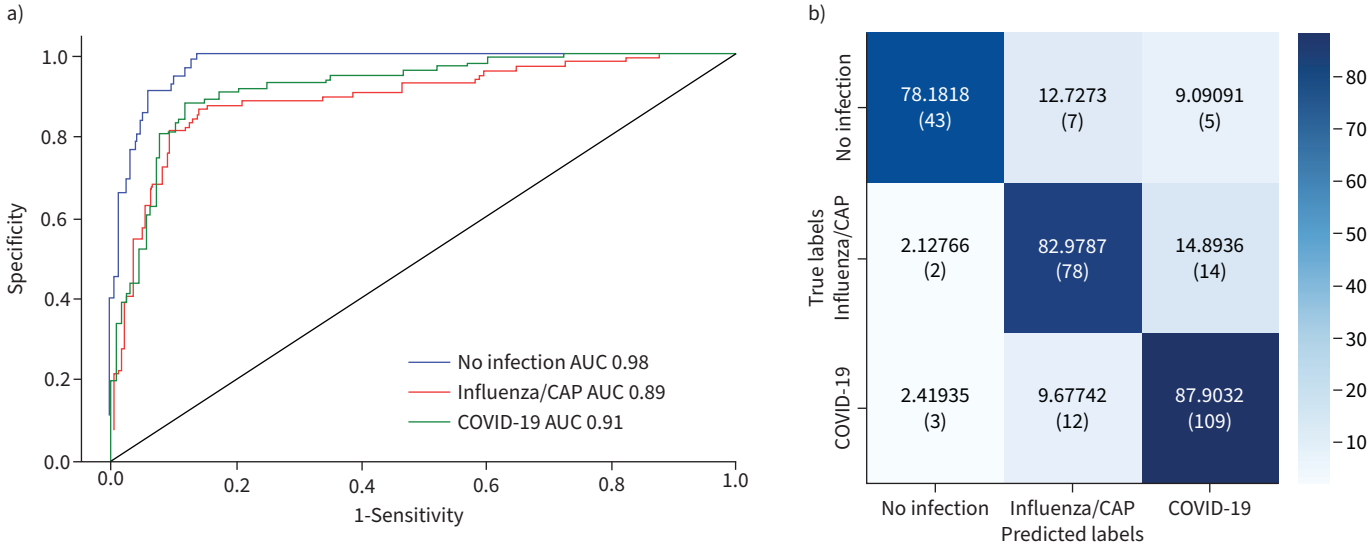

FIGURE 4 a) Receiver operating characteristic curve and b) confusion matrix for the internal test set. AUC: area under the curve; CAP: community-acquired pneumonia; COVID-19: coronavirus disease 2019.

an AUC of 0.92 (95% CI 0.87–0.96) (sensitivity 78.57%, 44 out of 56; specificity 94.54%, 208 out of 220) and no infection presented an AUC of 0.92 (95% CI 0.88–0.95) (sensitivity 84.21%, 48 out of 57; specificity 90.66%, 204 out of 225) (figure 5a). The confusion matrix for the external validation set is reported in figure 5b.

The performance in the external validation set is in good agreement with the internal testing set. A summary of the performance metrics for both internal test set and external validation set are presented in table 2. The TRIPOD score of the proposed model is 47% (15 out of 32 TRIPOD items). The output of the classification workflow is also reported in the clinical summary report. A sample report for influenza/CAP and COVID-19 patients is presented in supplementary figures S4 and S5.

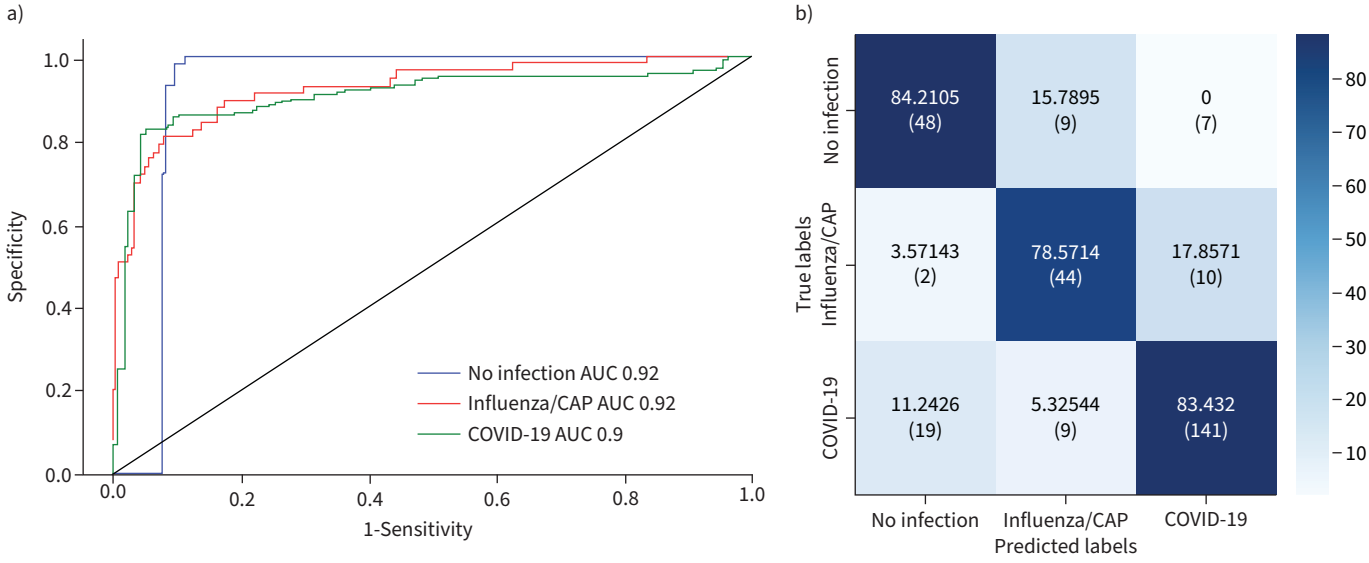

FIGURE 5 a) Receiver operating characteristic curve and b) confusion matrix for the external test set. AUC: area under the curve; CAP: community-acquired pneumonia; COVID-19: coronavirus disease 2019.

TABLE 2 Performance metrics results

	AUC		Sensitivity (%)		Specificity (%)	
	Internal test set	External validation set	Internal test set	External validation set	Internal test set	External validation set
No infection	0.98	0.92	78.18	84.21	97.72	92.59
Influenza/CAP	0.89	0.92	82.97	78.57	88.79	89.44
COVID-19	0.91	0.90	87.90	83.43	88.01	91.07

AUC: area under the curve; CAP: community-acquired pneumonia; COVID-19: coronavirus disease 2019.

Discussion

We developed and validated a deep-learning AI model for the classification of no infection, COVID-19 or influenza/CAP cases based upon CT imaging. The model showed a performance of AUC of 0.90, 0.92 and 0.92 for COVID-19, influenza/CAP and no infection, respectively, in the external validation. The proposed workflow automatically segments and detects both lungs and lung abnormalities, reducing the time and computational burden of the classification task. Moreover, the network produces an automatic clinical summary report that can be used by the clinician to verify the model decision.

The datasets used for this study come from different countries and different centres. The training cohort is from the University Hospital in Liege, while the external validation set is from the Babak Imaging Center in Tehran. The training dataset presents a certain homogeneity in imaging acquisition parameters, barring the use of different scanners at the different centres. However, the validation data present different characteristics, coming from a different country with different standards of care, and thus image acquisition protocols. This is an indication of the difference existing within the dataset and indirect proof of the generalisability of the performance of our model, which attained good performance in the external validation dataset.

Several deep-learning COVID-19 classification networks have been published thus far, both 2D [18] and 3D [19], also based on automatic segmentation of the lungs [20, 21]. Both machine learning [22], deep learning [23, 24] or a combination of both [25] have been explored for this classification task. The models' performances are high to very high for all the published approaches (AUC between 0.8 and 0.95) and several authors compared the AI workflow with clinicians' performances [26, 27], reporting comparable if not better performances from the AI models, and faster and more reproducible diagnosis. Our model has a performance of AUC ∼0.9 for all the classes, in line with those reported in the literature [21, 28].

The possibility to integrate a fully automatic tool for evaluation of pneumonia source in the clinical workflow can be instrumental to improve patient management and hospital resource allocation. Automatic identification of COVID-19, influenza/CAP and no infection patients can reduce the diagnostic errors related to human reader experience. The possibility of fast throughput of CT scans analysis will unburden medical staff and free resources to be allocated to more urgent needs. Dubious cases will have to be confirmed by clinicians upon examination, but the time and effort required to do so will be drastically reduced. A careful evaluation of the real cost/benefit of these tools is sorely needed to promote their application in the clinical practice.

However, these automatic tools still have important limitations of applicability in the clinical setting. Overfitting, lack of generalisability and of explainability are the most relevant ones for deep-learning models [29, 30]. In this study, several techniques were used to prevent overfitting. The model was trained on a multivendor (GE, Siemens) dataset with diverse acquisition protocols and on differently reconstructed series of the same case. In this way, the model learnt how to generalise in varying image acquisition parameters, which is well reflected by the high sensitivity when evaluated on a held-out internal test set with diverse acquisition protocols, and on the external validation set, coming from a different medical centre. The ability of the model architecture to generalise to images with diverse imaging parameters is a desired property for real-world clinical applications. Another important aspect of deep learning applied to medical image analysis is explainability, with the "black box" perception hampering the widespread adoption of these methods by clinicians. The production of parsimonious models (*i.e.* clinicians clearly comprehend and agree with how the model reached the result to support a clinical decision) is instrumental to build confidence and acceptance [31, 32]. In the field of AI there are two main explainability approaches: *post hoc* systems, which provide explanation for a single specific

decision and make it possible to obtain it on demand, and *ante hoc* systems (also known as "glass box"), in which the model is built to be intrinsically explainable, so it is possible to follow each step that the model takes to reach its classification decision [32–34]. Usually, (gradient) class activation maps are used to visualise the region of the scan on which the model based its classification decision [35]; thus, this explainability approach falls under the *post hoc* systems category. In the present study, the use of pre-selected and segmented slices containing lung abnormalities can be seen as an *ante hoc* explainability system, as the model is specifically looking at the abnormal areas of the lung, segmented by the lung abnormalities segmentation model. In this way the end user can verify on which slices and on which areas of the slice (*i.e.* the abnormalities) the model based its classification decision. This can be confirmed easily by the clinicians looking at the 48 consecutive slices along with lung and lung abnormalities segmentation masks, used by the model for the classification, and reported in the automatic clinical summary report (supplementary figures S1 and S2).

Indeed, our model selected only those slices containing abnormalities in the lungs, while most deep-learning models published in the literature [7, 36] are still based on manual segmentation of the CT scans and use as input all the slices containing lungs or the whole 3D lung volume when automatic segmentation is implemented. Moreover, in previous studies the identification of the regions of the slice used by the model to make its classification decision are the output of the model, helping with interpretability. In our model the identification of the abnormalities in the lungs, linked to the different kinds of pneumonia, is done *a priori*, removing irrelevant information (*e.g.* other pathological presentations in the lung). An additional advantage of our approach is the possibility to select upfront the scans for the model to process. If the selected slices do not present any abnormalities, the model will not process the image, saving time and computational power. This was verified in the external validation set. The no-infection patient cohort (n=76) of the COVID-CT-MD dataset includes healthy patients, among other no-infection patients: our segmentation model correctly identified all the slices without abnormalities and the corresponding scans were not processed by the model (19 out of 76 cases). Furthermore, the pre-selection of slices to be evaluated by the model allows a reduction of computational burden, also researched in this study by using Inception architecture. Indeed, the use of Inception architecture compared to other approaches based on ResNet or ResNext reduces the computational burden of the model, while maintaining equivalent performances [37]. This approach can allow shallow networks to achieve results comparable to their deeper and more complex counterparts with shorter training times, enabling good classification performances, even when using limited hardware [38]. The computation time (57 s per scan), which can be seen as an indication of the computational burden of the model, was faster than alternatives reported in the literature. Moreover, compared to other studies that used Inception architecture for similar classification tasks (table 3), our network showed comparable performances [39, 40] and was validated on an external testing set. This validation step is very important to verify the generalisability of the model to patients other than those used for model development (*i.e.* training and testing).

Limitations

Considering the limitations of this study, a relevant point related to the external validation test set is the presence of only CAP cases for the influenza/CAP class. This could lead to a misestimation of the model performance for this classification task. However, influenza cases were present in the internal validation and testing cohorts and the performances of the model were tested there. An additional external validation dataset with direct clinician assessment of source of pneumonia would strengthen the generalisability and add credibility to our approach.

TABLE 3 Performances of other classification models to distinguish coronavirus disease 2019 and other sources of pneumonia, based on Inception modules

	AUC		Sensitivity (%)		Specificity (%)		Sample size		Computation time[#] (s)
	Internal test set	External validation set	Internal test set	External validation set	Internal test set	External validation set	Internal test set	External validation set	
This work	0.91	0.90	87	83	88	91	273	305	57
Wang *et al.* [36]	0.93	0.81	88	83	87	67	455	290	n.r.
El Asnaoui *et al.* [40]	n.r.		92		96		n.r.		262
Gifani *et al.* [39]	0.85		77		n.r.		186		n.r.

AUC: area under the curve; n.r.: not reported. [#]: calculated as average time per scan on the external validation set.

The further distinction between bacterial and viral (non-COVID) pneumonia would represent an additional step forward, allowing the clear identification of the best therapeutic treatment for each patients. This can also result in a better therapeutic management, regarding for example the administration of antibiotics. The misuse and abuse of antibiotic is a cause of great concerns in the research and clinical communities. The insurgence of antimicrobial resistance is regarded as one of the top 10 global public health threats for the near future [41]. The timely identification of patients with pneumonias that does not require antibiotics can inform better therapy decisions and procedures, contributing to ease the burden of healthcare-associated infections from resistant strains of bacteria [42].

Looking at the dataset used for this study, the provenance of all scans from scanner from only two different vendors might somehow limit the generalisability of our approach, even though the images were acquired with two of the most commonly used scanner manufacturers on the market. Adding more data on different vendors, different acquisition and reconstruction settings might improve the model performances. Ideally, these kind of clinical decision-making support tools need to be continuously updated with new and heterogeneous data to attain accuracy, specificity and sensitivity comparable to the latest implementation of diagnostic and therapeutic state of the art, for example *via* distributed learning [43, 44].

To verify the real clinical utility of the proposed tool, a prospective clinical validation study should be carried out comparing performance and time to diagnosis of the AI tool to the current standard of care. Moreover, the clinical use of this tool might need to be updated and modified according to the development of the COVID-19 pandemic. We can expect that pneumonia from COVID-19 infection will become endemic and recurring in the future. Our approach could be adapted to spot undiagnosed cases or to provide a second independent verification of the occurrence of the disease, past the emergency status of this pandemic.

Conclusion

COVID-19-associated lung diseases can mimic other viral lung diseases such as (para-)influenza or CAP, which may result in misdiagnosis and delayed and unproper treatment. In this context, the development of new diagnostic tools based on AI could become critical for deployment in the daily practice in the near future. This approach could be exploited for other type of pulmonary diseases, fine-tuning the abnormalities segmentation model to recognise and select only the slices that contain the abnormalities relevant to the investigated disease. To reach this goal, a close collaboration between clinicians and data scientists is essential and will promote the future application of these decision support tools in the clinic.

Acknowledgements: The authors thank Fabio Bottari (Radiomics, Liège, Belgium) for providing medical writing support in accordance with Good Publication Practice (GPP3) guidelines.

Provenance: Submitted article, peer reviewed.

Author contributions: A. Vaidyanathan and J. Guiot were responsible for conceptualisation, methodology, software, validation, formal analysis, investigation, writing (original draft, and revision and editing), visualisation and data curation. F. Zerka, F. Belmans and I. Van Peufflik were responsible for data curation, validation and visualisation. L. Deprez, D. Danthine, G. Canivet, P. Meunier and M. Occhipinti were responsible for data curation, formal analysis, resources, validation and writing (revision and editing). P. Lambin, S. Walsh, W. Vos, P. Lovinfosse and R.T.H. Leijenaar were responsible for supervision, writing (revision and editing), funding acquisition and project administration.

Conflict of interest: A. Vaidyanathan, F. Zerka, F. Belmans, I. Van Peufflik and M. Occhipinti are salaried employees of Radiomics (Oncoradiomics SA). J. Guiot reports, within and outside the submitted work, research agreements from Radiomics (Oncoradiomics SA); he is in the permanent SAB of Radiomics (Oncoradiomics SA) for the SALMON trial without any specific consultancy fee for this work; he is co-inventor of one issued patent on radiomics licensed to Radiomics (Oncoradiomics SA); he confirms that none of these entities or funding was involved in the preparation of this work. P. Lambin reports, within and outside the submitted work, grants/sponsored research agreements from Radiomics (Oncoradiomics SA), ptTheragnostic/DNAmito, Health Innovation Ventures; he received an advisor/presenter fee and/or reimbursement of travel costs/consultancy fee and/or in-kind manpower contribution from Radiomics (Oncoradiomics SA), BHV, Merck, Varian, Elekta, ptTheragnostic, BMS and Convert Pharmaceuticals; he has minority shares in Radiomics (Oncoradiomics SA), Convert Pharmaceuticals, Comunicare Solutions and LivingMed Biotech; he is co-inventor of two issued patents with royalties on radiomics (PCT/NL2014/050248 and PCT/NL2014/050728) licenced to Radiomics (Oncoradiomics SA), one issued patent on mtDNA (PCT/EP2014/059089) licenced to ptTheragnostic/DNAmito, one unissued patent on LSRT (PCT/ P126537PC00) licenced

to Varian Medical, three nonpatented inventions (software) licenced to ptTheragnostic/DNAmito, Radiomics (Oncoradiomics SA) and Health Innovation Ventures, and three unissued, unlicenced patents on deep and handcrafted radiomics (US P125078US00, PCT/NL/2020/050794 and N2028271); he confirms that none of these entities or funding was involved in the preparation of this paper. R.T.H. Leijenaar has shares in the company Radiomics (Oncoradiomics SA) and is co-inventor of an issued patent with royalties on radiomics (PCT/NL2014/050728) licenced Radiomics (Oncoradiomics SA). S. Walsh and W. Vos have shares in the Radiomics (Oncoradiomics SA). The rest of the co-authors have no known competing financial interests or personal relationships to declare.

Support statement: The authors acknowledge financial support from European Marie Curie grant (PREDICT – ITN, number 766276) and the European Union's Horizon 2020 research and innovation programme under grant agreements DRAGON – 101005122 (call: H2020-JTI-IMI2-2020-21) and iCOVID – 101016131 (call: H2020-SC1-PHE-CORONAVIRUS-2020-2) for the execution of this work. Funding information for this article has been deposited with the Crossref Funder Registry.

References

1 Marchiori E, Zanetti G, Hochhegger B, *et al.* High-resolution computed tomography findings from adult patients with influenza A (H1N1) virus-associated pneumonia. *Eur J Radiol* 2010; 74: 93–98.
2 Gao L, Zhang J. Pulmonary high-resolution computed tomography (HRCT) findings of patients with early-stage coronavirus disease 2019 (COVID-19) in Hangzhou, China. *Med Sci Monit* 2020; 26: e923885.
3 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. *Lancet* 2020; 395: 507–513.
4 Guan W, Ni Z, Hu Y, *et al.* Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* 2020; 382: 1708–1720.
5 Bai HX, Hsieh B, Xiong Z, *et al.* Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT. *Radiology* 2020; 296: E46–E54.
6 Litjens G, Kooi T, Bejnordi BE, *et al.* A survey on deep learning in medical image analysis. *Med Image Anal* 2017; 42: 60–88.
7 Xu X, Jiang X, Ma C, *et al.* A deep learning system to screen novel coronavirus disease 2019 pneumonia. *Engineering* 2020; 6: 1122–1129.
8 Ardakani AA, Kanafi AR, Acharya UR, *et al.* Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: results of 10 convolutional neural networks. *Comput Biol Med* 2020; 121: 103795.
9 Emanuel EJ, Persad G, Upshur R, *et al.* Fair allocation of scarce medical resources in the time of COVID-19. *N Engl J Med* 2020; 382: 2049–2055.
10 Afshar P, Heidarian S, Enshaei N, *et al.* COVID-CT-MD, COVID-19 computed tomography scan dataset applicable in machine learning and deep learning. *Sci Data* 2021; 8: 121.
11 Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Netw* 2018; 106: 249–259.
12 Ilsen B, Vandenbroucke F, Beigelman-Aubry C, *et al.* Comparative interpretation of CT and standard radiography of the pleura. *J Belgian Soc Radiol* 2016; 100: 106.
13 Carreira J, Zisserman A. Quo vadis, action recognition? A new model and the kinetics dataset. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2017; pp. 4724–4733.
14 Kay W, Carreira J, Simonyan K, *et al.* The kinetics human action video dataset. *arXiv* 2017; preprint [https://arxiv.org/abs/1705.06950v1].
15 Szegedy C, Liu W, Jia Y, *et al.* Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2015; pp. 1–9.
16 Kingma DP, Ba JL. Adam: A Method for Stochastic Optimization. 3rd International Conference for Learning Representations, San Diego, 2015; pp. 1–15.
17 Collins GS, Reitsma JB, Altman DG, *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC Med* 2015; 13: 1.
18 Ibrahim DM, Elshennawy NM, Sarhan AM. Deep-chest: multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases. *Comput Biol Med* 2021; 132: 104348.
19 Li L, Qin L, Xu Z, *et al.* Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. *Radiology* 2020; 296: E65–E71.
20 Ghaderzadeh M, Asadi F. Deep learning in the detection and diagnosis of COVID-19 using radiology modalities: a systematic review. *J Healthc Eng* 2021; 2021: 6677314.
21 Mohammad-Rahimi H, Nadimi M, Ghalyanchi-Langeroudi A, *et al.* Application of machine learning in diagnosis of COVID-19 through X-ray and CT images: a scoping review. *Front Cardiovasc Med* 2021; 8: 638011.
22 Wu Z, Li L, Jin R, *et al.* Texture feature-based machine learning classifier could assist in the diagnosis of COVID-19. *Eur J Radiol* 2021; 137: 109602.

23 Yan T, Wong PK, Ren H, *et al.* Automatic distinction between COVID-19 and common pneumonia using multi-scale convolutional neural network on chest CT scans. *Chaos Solitons Fractals* 2020; 140: 110153.
24 Wang S, Zha Y, Li W, *et al.* A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *Eur Respir J* 2020; 56: 2000775.
25 Wang H, Wang L, Lee EH, *et al.* Decoding COVID-19 pneumonia: comparison of deep learning and radiomics CT image signatures. *Eur J Nucl Med Mol Imaging* 2021; 48: 1478–1486.
26 Liu H, Ren H, Wu Z, *et al.* CT radiomics facilitates more accurate diagnosis of COVID-19 pneumonia: compared with CO-RADS. *J Transl Med* 2021; 19: 29.
27 Zhang K, Liu X, Shen J, *et al.* Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell* 2020; 181: 1423–1433.
28 Ozsahin I, Sekeroglu B, Musa MS, *et al.* Review on diagnosis of COVID-19 from chest CT images using artificial intelligence. *Comput Math Methods Med* 2020; 2020: 9756518.
29 Ying X. An overview of overfitting and its solutions. *J Phys Conf Ser* 2019; 1168: 022022.
30 Caruana R, Lawrence S, Giles L. Overfitting in neural nets: backpropagation, conjugate gradient, and early stopping. *Adv Neural Inf Process Syst* 2001; 13: 402–408.
31 Singh A, Sengupta S, Lakshminarayanan V. Explainable deep learning models in medical image analysis. *J Imaging* 2020; 6: 52.
32 Holzinger A, Langs G, Denk H, *et al.* Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip Rev Data Min Knowl Discov* 2019; 9: e1312.
33 Holzinger A, Biemann C, Pattichis CS, *et al.* What do we need to build explainable AI systems for the medical domain? *ArXiv* 2017; preprint [http://arxiv.org/abs/1712.09923].
34 Holzinger A. Explainable AI and multi-modal causability in medicine. *I-Com* 2021; 19: 171–179.
35 Jin C, Chen W, Cao Y, *et al.* Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. *Nat Commun* 2020; 11: 5088.
36 Wang S, Kang B, Ma J, *et al.* A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19). *Eur Radiol* 2021; 31: 6096–6104.
37 Bianco S, Cadene R, Celona L, *et al.* Benchmark analysis of representative deep neural network architectures. *IEEE Access* 2018; 6: 64270–64277.
38 Bressem KK, Adams LC, Erxleben C, *et al.* Comparing different deep learning architectures for classification of chest radiographs. *Sci Rep* 2020; 10: 13590.
39 Gifani P, Shalbaf A, Vafaeezadeh M. Automated detection of COVID-19 using ensemble of transfer learning with deep convolutional neural network based on CT scans. *Int J Comput Assist Radiol Surg* 2021; 16: 115–123.
40 El Asnaoui K, Chawki Y. Using X-ray images and deep learning for automated detection of coronavirus disease. *J Biomol Struct Dyn* 2020; 39: 3615–3626.
41 Global Action Plan on Antimicrobial Resistance. *Microbe Mag* 2015; 10: 354–355.
42 Haque M, McKimm J, Sartelli M, *et al.* Strategies to prevent healthcare-associated infections: a narrative overview. *Risk Manag Healthc Policy* 2020; 13: 1765–1780.
43 Zerka F, Urovi V, Vaidyanathan A, *et al.* Blockchain for privacy preserving and trustworthy distributed machine learning in multicentric medical imaging (C-DistriM). *IEEE Access* 2020; 8: 183939–183951.
44 Zerka F, Barakat S, Walsh S, *et al.* Systematic review of privacy-preserving distributed machine learning from federated databases in health care. *JCO Clin Cancer Inform* 2020; 4: 184–200.