Computers in Biology and Medicine 141 (2022) 105143

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Deep diagnostic agent forest (DDAF): A deep learning pathogen recognition system for pneumonia based on CT

Weixiang Chen [a,1], Xiaoyu Han [b,c,1], Jian Wang [d,e,1], Yukun Cao [b,c], Xi Jia [b,c], Yuting Zheng [b,c], Jie Zhou [a], Wenjuan Zeng [d,e,**], Lin Wang [d,e,***], Heshui Shi [b,c,****], Jianjiang Feng [a,*]

[a] Department of Automation, Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing, China
[b] Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[c] Department of Laboratory Medicine, Liyuan Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[d] Department of Clinical Laboratory, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[e] Research Center for Tissue Engineering and Regenerative Medicine, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

ARTICLE INFO

Keywords:
Pathogens of pneumonia
Deep learning
Imbalanced data

ABSTRACT

Background: Even though antibiotics agents are widely used, pneumonia is still one of the most common causes of death around the world. Some severe, fast-spreading pneumonia can even cause huge influence on global economy and life security. In order to give optimal medication regimens and prevent infectious pneumonia's spreading, recognition of pathogens is important.

Method: In this single-institution retrospective study, 2,353 patients with their CT volumes are included, each of whom was infected by one of 12 known kinds of pathogens. We propose Deep Diagnostic Agent Forest (DDAF) to recognize the pathogen of a patient based on ones' CT volume, which is a challenging multiclass classification problem, with large intraclass variations and small interclass variations and very imbalanced data.

Results: The model achieves 0.899 ± 0.004 multi-way area under curves of receiver (AUC) for level-I pathogen recognition, which are five rough groups of pathogens, and 0.851 ± 0.003 AUC for level-II recognition, which are 12 fine-level pathogens. The model also outperforms the average result of seven human readers in level-I recognition and outperforms all readers in level-II recognition, who can only reach an average result of 7.71 ± 4.10% accuracy.

Conclusion: Deep learning model can help in recognition pathogens using CTs only, which might help accelerate the process of etiological diagnosis.

1. Introduction

Despite the availability of antibiotics agents, pneumonia still constitutes the third most common cause of death and is the first cause of death from infections [1–3]. According to 2019 Global Burden of Diseases (GBD) [4], pneumonia and bronchiolitis affected 489 million people globally in 2019, and an another statistics showed that over 2.56 million people died from pneumonia in 2017 [5]. Since 2020, a viral pneumonia caused by a novel coronavirus (COVID-19) has killed over 5.1 million people worldwide so far [5]. It is crucial to identify the causative pathogens of pneumonia, because without a specific diagnosis, patients cannot get adequate antimicrobial therapy and some

Abbreviations: AUC, area under curves of receiver; CAP, community-acquired pneumonia; CT, computed tomography; CXR, Chest X-Ray; DDAF, deep diagnostic agent forest; GBD, Global Burden of Diseases; HAP, hospital-acquired pneumonia.

* Corresponding authors.
** Corresponding author. Department of Clinical Laboratory, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.
*** Corresponding authors. Department of Clinical Laboratory, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.
**** Corresponding author. Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.
E-mail addresses: 1124177514@qq.com (W. Zeng), lin_wang@hust.edu.cn (L. Wang), heshuishi@hust.edu.cn (H. Shi), jfeng@tsinghua.edu.cn (J. Feng).
[1] These authors are co-first authors: Weixiang Chen, Xiaoyu Han, Jian Wang.

https://doi.org/10.1016/j.compbiomed.2021.105143
Received 19 October 2021; Received in revised form 5 December 2021; Accepted 12 December 2021
Available online 16 December 2021

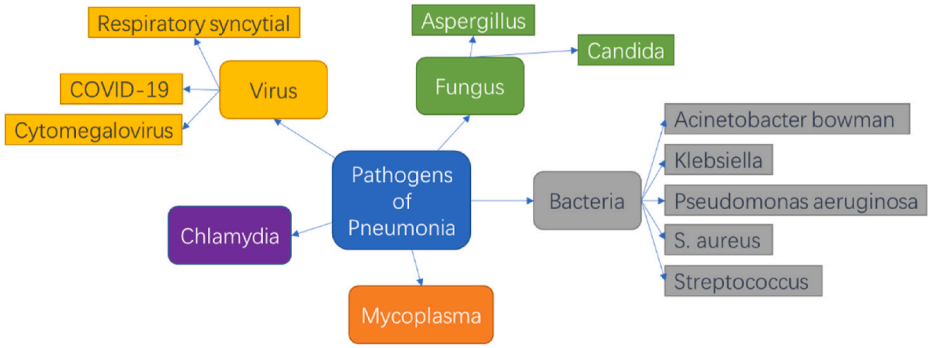


Fig. 1. The definition of level-I and level-II labels. The rounded rectangles mean level-I labels while the outer rectangles mean the level-II labels. Chlamydia and Mycoplasma have no fine-level categories in our database and are treated as both level-I and level-II labels.

Table 1
Information about the cohorts among all categories.

Pathogen categories		Gender		Age				Train /Val cohort	Test cohort	Total	Total for level I
level-I	level-II	M	F	<20	21–40	41–60	>60				
Virus	Cytomegalo-virus	63	56	24	38	40	17	58	61	119	1009
	Respiratory syncytial	24	19	3	4	17	19	20	23	43	
	COVID-19	401	446	3	182	357	306	428	419	847	
Fungus	Aspergillus	122	74	5	14	85	92	97	99	196	375
	Candida	111	68	0	12	47	120	88	91	179	
Bacteria	Acinetobacter bowman	144	55	2	22	79	96	98	101	199	873
	Klebsiella	118	29	0	20	42	86	72	75	147	
	Pseudomonas aeruginosa	96	54	5	22	50	73	74	76	150	
	S. aureus	139	69	6	24	95	83	103	105	208	
	Streptococcus	113	56	7	30	64	68	83	86	169	
Chlamydia		23	24	5	9	4	29	22	25	47	47
Mycoplasma		27	22	29	7	5	8	23	26	49	49
Total		1381	972	89	384	885	997	1166	1187	2353	2353
Mean of 2-cls p-value		0.302451598		0.17499104				–			

pneumonia could progress rapidly to severe illness or even spread widely [6]. Besides, though empiric wide spectrum therapy can almost reach similar efficacy without obtaining specific diagnosis, it can also cause risks of antibiotic resistance, which is especially harmful for children [7]. If specific fine-grained pathogen is known, accurate medication and infection prevention can be taken without delay.

On the other hand, it is challenging to reach an etiological diagnosis of pathogens. There are more than 100 kinds of bacteria, viruses and fungi that can cause pneumonia [8], and the symptoms overlap significantly among various lower respiratory infections [9]. To obtain a precise recognition of pathogens, a complex pipeline of biochemical tests is usually performed, which can be high cost, time-consuming or invasive with poor diagnostic efficiency. As a result, etiological diagnosis of pathogens is only achieved in half of patients [10]. Instead, pneumonia is usually broadly divided into community-acquired pneumonia (CAP) or hospital-acquired pneumonia (HAP), and the treatments of them are often based on the different experiential prior possibility of CAP and HAP. Chest X-Ray (CXR) is useful in establishing parenchymal lung involvement [11], but due to lower rendering quality and limited information, it often leads to misdiagnoses which cause delay of antimicrobial therapy [12]. Chest computed tomography (CT) is believed to be more informative than CXR, but it still faces difficulties that different types of pneumonia may share some common radiologic features [13]. Some researchers have focused on manual pathogen recognition of pneumonia, but they considered at most four different types of pathogens and took visual radiology features as clues for manual recognition [14–16], such as bronchial wall thickening, consolidation and so on. It is a consensus [6,7] that recognizing pathogens using CTs is extremely hard for radiologists. However, if people can get some reasonable prediction with the help of computers, the situation may change. The quick recognition using images can also be used together with biochemical tests, which can help doctors design the pipeline of test and determine the exact pathogen more efficiently.

In recent years computer-aided diagnosis (CAD) system has developed rapidly [17–19] thanks to the rapid development of deep neural network [20,21]. With the help of computers, compact and distinctive features can be extracted and the quantitative calculation of computers helps avoid subjective bias of doctors. Dimensionality of medical images is too high for human to extract important features, while CAD can help in this situation. When using deep neural network technology, the convolutional layers can extract distinctive features from huge quantities of data. Until now, deep learning based CAD systems have achieved high accuracies in many medical image analysis tasks, such as lung nodule detection [22–24], organ segmentation [25–27] and multiple category diagnosis [28–32]. Recent studies have shown that deep learning methods yielded great performance in distinguishing COVID-19 from other lung diseases, such as lung nodules, non-virus pneumonia, and healthy cases [33–40], which is a relatively easier classification task than pathogen recognition, due to the small number of categories and drastically different patterns among them. As far as our survey, there is no publication on automatic pathogen recognition of pneumonia using CTs.

Our contributions in this study are three-fold:

1. A multi-pathogen CT database with 12 different types of pathogens was collected, in which each subject has CT volume and determined pathogen ground truth.
2. A network called **D**eep **D**iagnostic **A**gent **F**orest network (DDAF) was proposed to deal with this fine-level, imbalanced pathogen recognition problem.
3. DDAF reached a 0.899 ± 0.004 multi-way AUC, better than the average performance of seven radiologists in a reader study.

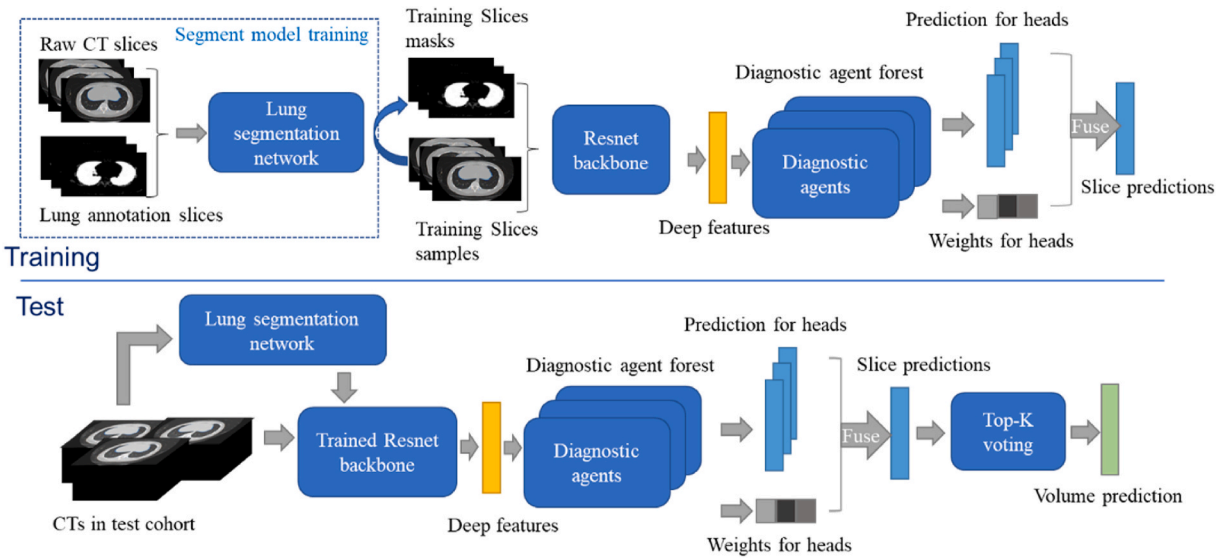


Fig. 2. The overall pipelines of DDAF for training and test process. The training and test (inference) process have slight differences. The training is done all at slice level, while the test process includes a slice fusion block, Top-K voting, to output the volume level result.

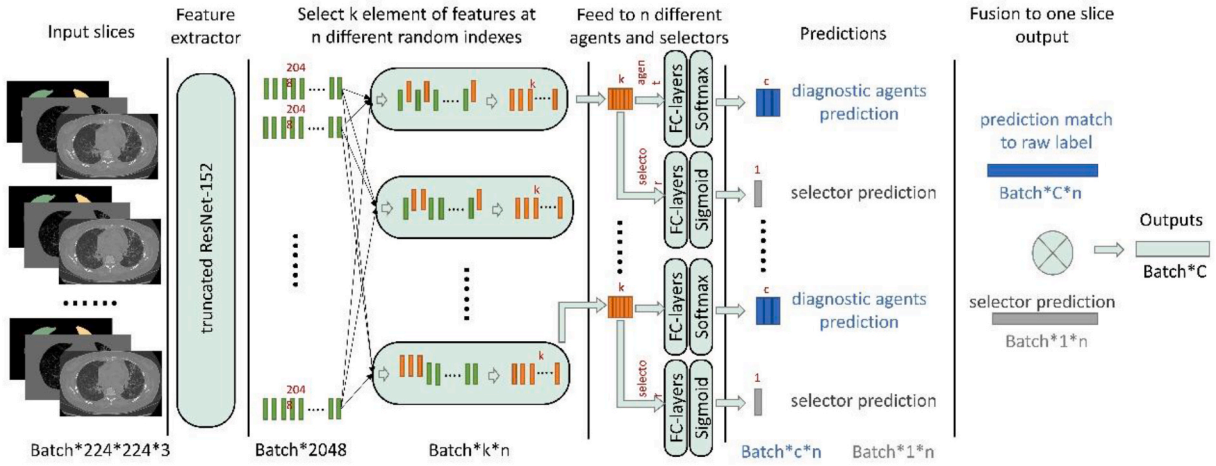


Fig. 3. The main structure of neural network in DDAF, in which the feature extractor is a truncated Resnet [21], *n* selectors and agents are paired and used to give classification predictions. Predictions from agents and selectors are finally fused. Inputs of our network are concatenated images, lung masks and masked images. The process that fuses slice-level results into volume-level, and the segmentation network are not shown in this figure.

2. The database

This study obtained ethical approval from the Ethics Commission of Wuhan Union Hospital. All participants remained anonymous, and the requirement for informed patient consent was waived by the ethics committee for this retrospective study. This study was registered with the Chinese Clinical Trial Registry, ChiCTR2000038609.

Initially, 4090 patients were included according to the inclusion criteria: (1) lower respiratory infections confirmed by PCR positive, or sputum culture; (2) clinically common pathogen of pneumonia infection; (3) available chest CT scans during acute infection phase; (4) pneumonia confirmed by CT; and (5) age 5–90 years old. In total, 1737 patients were excluded because of the following exclusion criteria: (1) with multiple or secondary infections (n = 1102); (2) heavy CT image artefacts (n = 126); (3) the interval between CT scan time point and onset time was >1 month (n = 388); (4) combined with pulmonary edema (n = 76), pulmonary malignancy (n = 22) or esophagus surgery (n = 9); (5) unclear diagnostic results or possible misdiagnosis according to double check (n = 14). Ultimately, 2353 patients with 2353 CT scans were included.

We defined the rough level of pathogens, virus, fungus and so on, as level-I labels, and the precise level of pathogens as level-II labels. As a result, 2,353 samples contain 5 different level-I labels and 12 different level-II labels, which is shown in Fig. 1 and Table 1.

All slices in volumes are resized to 224 × 224 pixels to meet the need of pretrained CNN backbone of DDAF. The Hounsfield value is clipped within −1200 to 500, and subsequently normalized to 0–1, to meet the need of pretrained backbone.

The whole database was divided into training/validation cohort and test cohort without any overlap. We divided our database randomly with the same proportion (almost 1:1) for every level-II category as shown in Table 1. The deep model is trained on the training/validation cohort and the performances are estimated on the test cohort. Among all cases of different pathogens, distributions of gender or ages are not significantly different (p = 0.302 for genders, and p = 0.174 for ages).

3. Methods

The main challenge of pathogen recognition is that the data are very imbalanced and limited in size. Some pathogens are relatively rare than

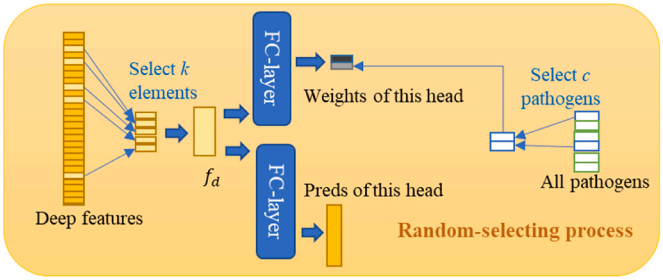


Fig. 4. Detailed process that how random-selecting is done on selectors and agents. For every agent, k elements of deep features are selected, whose position index are different among agents. The selected features are used to predict c selected pathogens which are also different among cases.

others, while it is still important to recognize them. Besides, the three-dimensional volume also brings some trouble to engineering, because it is hard to process the whole volume using deep networks within GPU memory. Our network uses a pipeline of two-dimensional training and volume-level inference which reduces the memory consumption of three-dimensional data. Besides, in order to overcome imbalanced data and long-tail distribution, our proposed network groups categories into hyper-groups and uses multiple different agents to handle different groups. In each agent, a classifier is designed to do inner group recognition which is trained on data within the group, while an agent selector is designed to recognize whether the input is suitable for its group. All these designs help DDAF yield better result on pathogens recognition.

3.1. DDAF system and pipelines

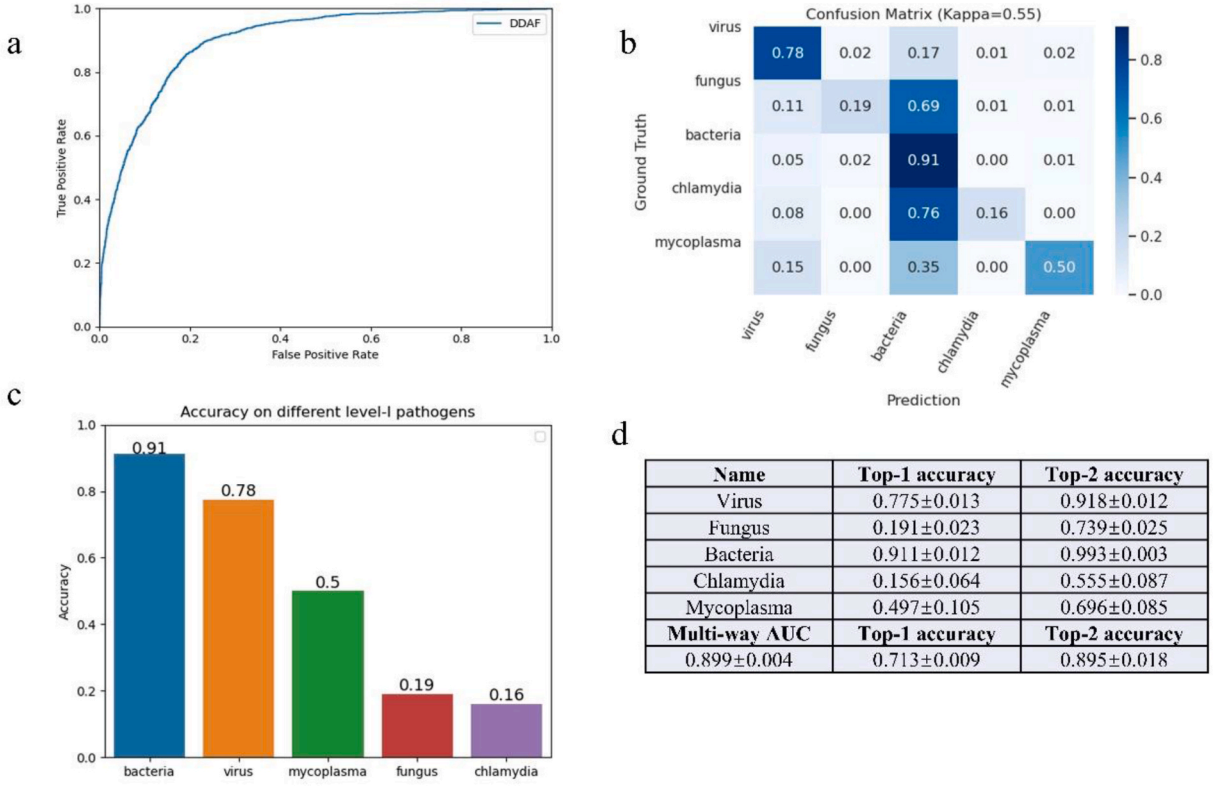
Our DDAF is composed of a segmentation network, a feature extraction backbone, a diagnostic agent forest, and a top-K voting module. To deal with the three-dimensional CT data, DDAF is trained on slice level, while during test, a slice fusion block is used to compute a single result for the input volume (shown in Fig. 2). This design helps the network easier to convergence because two-dimensional convolutional kernels have less parameters and pretrained parameters are available. On the other hand, when replacing volumes into slices of them, the number of training cases are hugely enlarged which is also helpful for training.

When training, all training volumes are cut into slices and the slices are processed by a lung segmentation CNN model firstly. Image slices are then concatenated with their lung segmentations, and are fed into DDAF model together. A manual selection is done among all training slices in order to remove those without lesion of pneumonia on images. This process is only done in training/validation cohort in order to make sure all slices in training are positive slices which do have lesion in them since other slices are irrelevant to the pathogen recognition task. Though the selection is tedious in some degree, it is still faster than totally manual segmentation which is often required for training lesion segmentation algorithms. The training and validation are done at slice level and then the parameters are fixed for inference. The detailed network structure will be described in latter section and is shown in Fig. 3.

While in the test stage, all slices of a volume are inputted into the system, and no manual selection is required. All slices for a volume are processed in inference, and classification results of all slices are fused by Top-K fusion block into a single volume-level result. Therefore, our method can handle variable size or span of pneumonia when inference. The feature extraction part is similar to Ref. [41], which is used as one of our baselines in experiments.

3.2. Lung segmentation network

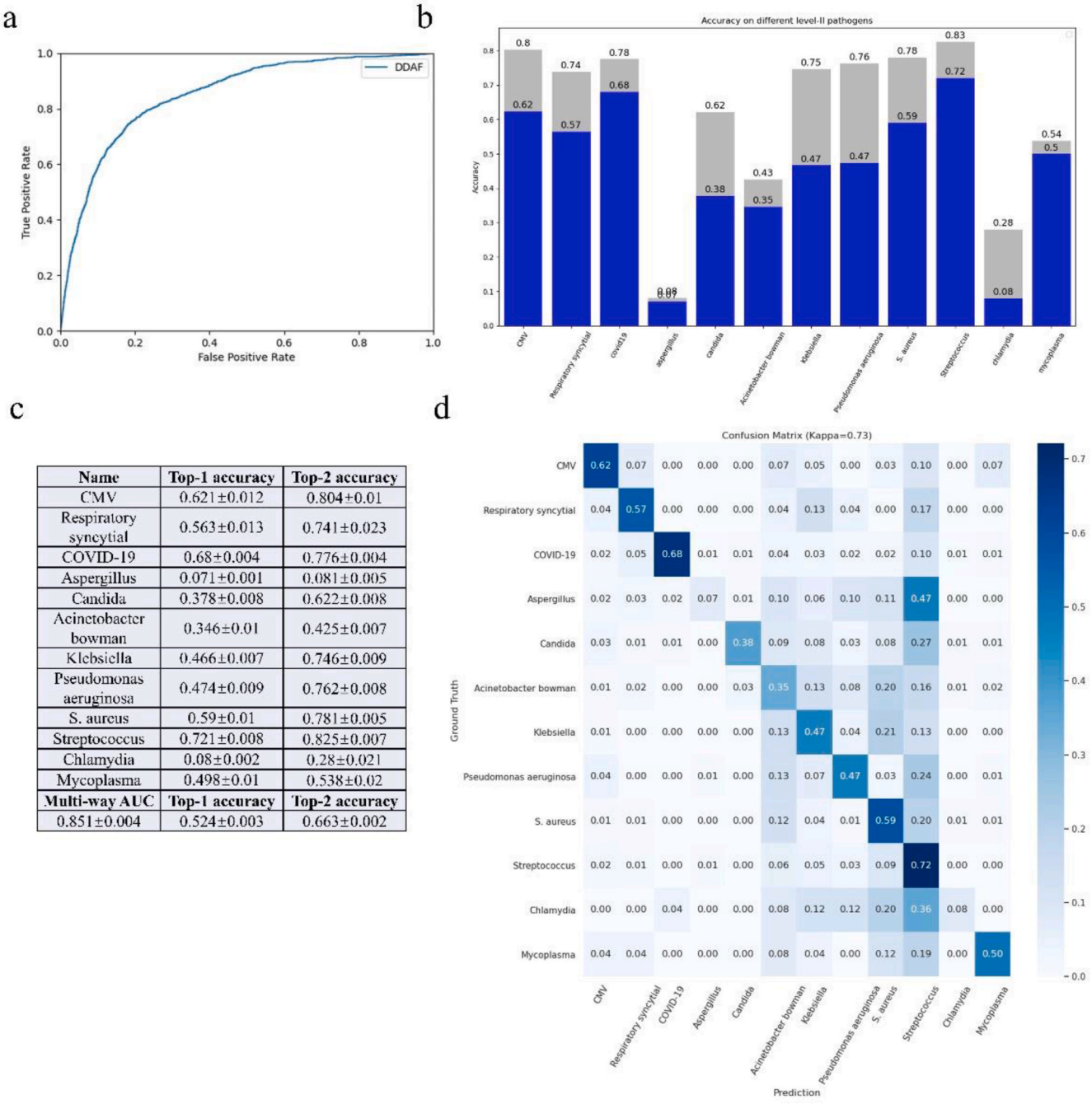
Both in training and test process, the input of DDAF are a concatenated images or volumes, which are consisted by the raw images and the lung segmentation masks. The segmentation masks force network to



d

Name	Top-1 accuracy	Top-2 accuracy
Virus	0.775±0.013	0.918±0.012
Fungus	0.191±0.023	0.739±0.025
Bacteria	0.911±0.012	0.993±0.003
Chlamydia	0.156±0.064	0.555±0.087
Mycoplasma	0.497±0.105	0.696±0.085
Multi-way AUC	**Top-1 accuracy**	**Top-2 accuracy**
0.899±0.004	0.713±0.009	0.895±0.018

Fig. 5. Results for level-I recognition. a) The overall ROC of our deep network. b) The confusion matrix among level-I categories. c) The bin-plot for accuracy of level-I categories. d) The metrics of our deep network.



Name	Top-1 accuracy	Top-2 accuracy
CMV	0.621±0.012	0.804±0.01
Respiratory syncytial	0.563±0.013	0.741±0.023
COVID-19	0.68±0.004	0.776±0.004
Aspergillus	0.071±0.001	0.081±0.005
Candida	0.378±0.008	0.622±0.008
Acinetobacter bowman	0.346±0.01	0.425±0.007
Klebsiella	0.466±0.007	0.746±0.009
Pseudomonas aeruginosa	0.474±0.009	0.762±0.008
S. aureus	0.59±0.01	0.781±0.005
Streptococcus	0.721±0.008	0.825±0.007
Chlamydia	0.08±0.002	0.28±0.021
Mycoplasma	0.498±0.01	0.538±0.02
Multi-way AUC	**Top-1 accuracy**	**Top-2 accuracy**
0.851±0.004	0.524±0.003	0.663±0.002

Fig. 6. Results for level-II recognition. a) The overall ROC of our deep network for level-II. b) The bin-plot for accuracy of level-II categories. The blue bins denote the top-1 accuracies while the gray bins denote top-2 accuracies. c) The metrics of our deep network for level-II. d) The confusion matrix among level-II categories.

focus on area of interests which in our cases are lung areas. We used U-net [26,27] to process all data and get their lung masks.

3.3. Feature extraction backbone

The feature extraction part extracts deep features from CT images using a truncated ResNet152 [21] backbone whose fully connected layers have been removed (also shown in Fig. 3). Resnet is combined by multiple Res-blocks, in which features of former layers skip and concatenate to the latter feature. As a result, a 2048-dimension (which means 2048 element-length features) deep features is extracted from every slice.

3.4. Deep diagnostic agent forest

The classification part of our network uses multiple diagnostic agents which are denoted as deep diagnostic agent forest (Fig. 3). Inspired by the real triage procedure in the hospital, we suppose different pathogens should be identified by different experts. When a diagnostic agent is trained specially using a pre-divided sub-group of pathogens, it can work as an expert in those pathogens which focuses more on the fine-level differences within the sub-group. Since all deep diagnostic agents are trained to find classification margins, the group of pathogens can also reduce the difficulties. What's more, when data is grouped, the data imbalance for each agent is reduced. Therefore, we proposed DDAF to bring in multi-expert principle and to deal with data imbalance.

Table 2
Performances of the proposed method and some other network structures. Best performances are shown in bold.

Backbone	Slice/ Volume	Classifier	AUC level-I	Acc level-I	AUC level-II	Acc level-II
DRE-Net			0.884 ± 0.006	0.613 ± 0.002	0.798 ± 0.043	0.375 ± 0.012
Deep-chest			0.822 ± 0.005	0.491 ± 0.011	0.721 ± 0.007	0.251 ± 0.010
Resnet152	Slice + LSTM	DDAF	0.529 ± 0.008	0.146 ± 0.007	0.439 ± 0.006	0.063 ± 0.005
Resnet50-3D	Volume	DDAF	0.659 ± 0.009	0.259 ± 0.010	0.450 ± 0.009	0.170 ± 0.010
Xception	Slice + TopK	DDAF	0.871 ± 0.003	0.633 ± 0.012	0.755 ± 0.007	0.372 ± 0.012
Densenet121	Slice + TopK	DDAF	0.867 ± 0.005	0.568 ± 0.012	0.761 ± 0.005	0.349 ± 0.011
Resnet152	Slice + TopK	Pure dense	0.718 ± 0.009	0.547 ± 0.013	0.618 ± 0.008	0.224 ± 0.010
Resnet152	Slice + TopK	Group Softmax	0.776 ± 0.005	0.697 ± 0.010	0.768 ± 0.005	0.481 ± 0.012
Resnet152	Slice + TopK	Designed agents	0.892 ± 0.004	0.702 ± 0.009	0.805 ± 0.005	0.434 ± 0.012
Resnet152	Slice + TopK	DDAF	**0.899 ± 0.004**	**0.713 ± 0.009**	**0.851 ± 0.004**	**0.524 ± 0.011**

To realize the principle and train the model end-to-end, we designed structure as shown in Fig. 3, of which some details are shown in Fig. 4. For a single agent, it chooses k element from deep features randomly by randomly determining k indexes for the multi-expert principle. On the classification target side, a choice is also made to determine c pathogens. The k indexes and c pathogens are fixed for a specific agent. This kind of agents have n copies in network forming agent forest. Multiple agents can help predict all pathogens when n is larger enough. In the experiments, we set up $\frac{n}{2}$ agents for level-I and $\frac{n}{2}$ agents for level-II. The selected features, denoted as f_d, are processed through two linear layers for classification. Cross-entropy loss is used as objective functions for classification, denoted as L_{cls}.

A selector is also designed for every agent to predict whether this agent can cover the pathogen of this input CT or CT slice. The selectors are trained together with classifiers which is called an end-to-end training, and when calculating losses, the loss of selector (also cross-entropy) is also added into the global loss, denoted as L_{sel}. When a case should not be predicted by one agent, the ground-truth of switch of that agent is zero, so the objective function of the training is:

$$L_{all} = \lambda_1 L_{cls} + \lambda_2 L_{sel} + \lambda_3 R,$$

in which R is regularization items to prevent overfitting, and λ_x are weighting parameters of different objective functions. Besides, L_{cls} and L_{sel} are defined as followed,

$$L_{cls} = \sum_{i=0}^{n} \sum_{j=0}^{b} -\widehat{p}_j \log\left(p_{ij}\right) \times \mathrm{I}\left(\widehat{p}_j \in c_i\right),$$

$$L_{sel} = \sum_{i=0}^{n} \sum_{j=0}^{b} -\widehat{w}_j \log\left(w_{ij}\right), \quad \text{where} \quad \widehat{w}_j := \mathrm{I}\left(\widehat{p}_j \in c_i\right),$$

When inference, the results of selectors are used as weights on majority voting of classifiers to fuse agents' results of a slice. The voting of agents can be further described as multiplication of two independent probabilities, which are selectors' prediction probability and agents' prediction probability respectively:

$$P(y\,|X) = \frac{\sum_i Selector_i(X) \times Agent_i(X)}{\sum_i Selector_i(X)}$$

Though we cannot get the ground truth agent when inference, we can get all predictions of agents $Agent_i(X)$ together with all selectors' prediction $Selector_i(X)$, and calculate result using the function above. Result of a slice is available after voting of agents and then the results of all slices in the same volume are processed by a top-K fusion step to form the result of the whole volume.

3.5. Training and evaluation details

The training/validation cohort is used to select hyper-parameters and inner network parameters of network. Since the cases of our database are not abundant, training and validation are done in a four-fold cross validation manner. After validation, the hyper-parameters (n, c, m) were selected. The training converged after 20 epochs for every fold, which costs 32 h on one NVIDIA GeForce 3090, and for level-I agents, $(n, m, c) = (50, 1024, 2)$ for level-II agents, $(n, m, c) = (50, 1024, 3)$.

In order to evaluate the performances, we computed the accuracy and areas under receiver curves (AUC) on both level-I and level-II categories. Since results on different databases can be different, we reproduced some common methods, and train and test them on the same situation and database. Reader studies were also done to compare performances of DDAF and human readers.

3.6. Sub-groups analysis

In order to analyze performances better, the bias of infectious degree should be removed [9,42,43]. Since not all patients came and underwent CTs in the same degree or period of infectious, and it is hard to revert the situations they went to hospital, we simplify the experiments to keep the infected area within the same levels.

We divided sub-groups according to the area of infected area obtained by another deep segmentation network. The segmentation network was trained on another database of COVID-19 cases with manual annotations. We asked two experienced radiologists to annotate 10 cases of COVID-19 lesion areas, the training cases were randomly selected from COVID-19 cases from Union Hospital of Wuhan, and none of them is included in the training or test cohort of this study. After training, the segmentation network processed on the whole test cohort, and we clustered the ratio of infected area to the whole lung areas into two clusters, and the higher one was denoted as large lesion sub-group, the lower one was small lesion sub-group.

3.7. Reader studies

Reader studies were carried out to obtain a relative baseline to compare our method. The readers were asked to perform two tasks, and readers could only check CTs and make decisions independently. Seven invited radiologists with different degrees of experience (HSS, JG, XQL, YKC, XYH, JX, and YTZ with 32, 12, 12, 6, 5, 2, and 1 years of experiences in thoracic radiology) independently performed all qualitative image analysis on Digital Imaging and Communications in Medicine (DICOM) images from the CT studies without access to clinical and laboratory findings. They were not specially trained for pneumonia pathogen recognition but all of them are experienced in CT reading and analysis, which is common for most hospitals.

In the first task, radiologists need to decide which pathogens had infected the patients. They have five choices and need to list the five possible level-II pathogens by the order of probabilities they supposed in the first test. At the same time, the radiologists were also asked to list the most possible level-I pathogen for each case. Readers could give

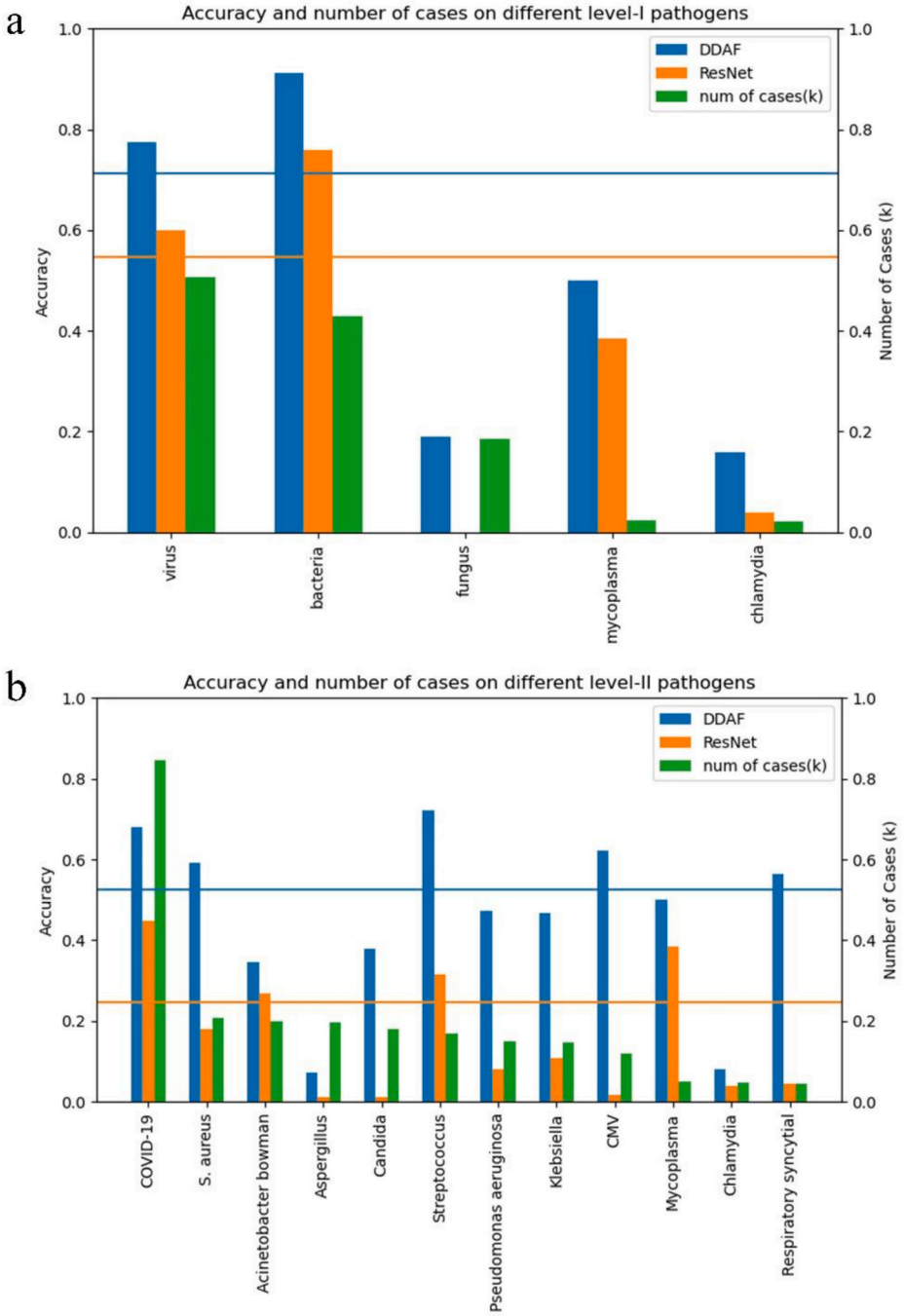


Fig. 7. Accuracy comparison between DDAF and ResNet on both level-I and level-II pathogens. Number of cases of every level-I and level-II labels are also shown. a) Bin-plot of accuracies for level-I. b) Bin-plot of accuracies for level-II. The index of pathogens has been sorted by the number of cases, and the numbers are plotted in thousands. The horizontal lines mean averaged accuracies of two methods.

predictions for level-I and level-II independently, which means they can choose any level-II pathogens without caring of the level-I results they closed. For the second task, they need to decide whether the case is a viral pneumonia or not, which is a common task for radiology department in hospitals. The decisions were made without limitation of time, and radiologists were well informed the rules and knew the list of all pathogens possible in the cohorts.

4. Experiments and results

4.1. Level-I performance

We evaluated the recognition accuracies of our network at both level-I and level-II. In level-I recognition, the deep network outputs predictions for input CT volumes among viral, fungal, bacterial, mycoplasmal or chlamydial pneumonia (Fig. 5), and it gets an averaged accuracy of 0.713 ± 0.009 and multi-way AUC of 0.899 ± 0.004. Bacteria pneumonia has the highest accuracy in our network, while fungus pneumonia and chlamydia pneumonia tend to be confused as bacteria pneumonia. We have tried to simplify the multi-class classification

Table 3
Number of cases in each sub-cohort of our test cohort.

level-I	Small lesion	Large lesion	level-II	Small lesion	Large lesion	Total
Virus	473	66	Cytomegalovirus	36	15	539
			Respiratory syncytial	18	5	
			Covid-19	373	46	
Fungus	158	32	Aspergillus	88	11	190
			Candida	70	21	
Bacteria	322	121	Acinetobacter bowman	64	37	443
			Klebsiella	47	28	
			Pseudomonas aeruginosa	53	23	
			S. aureus	81	24	
			Streptococcus	77	9	
Chlamydia				23	2	25
Mycoplasma				24	2	26
Total				964	223	1187

Table 4
Performances of the proposed method on different subgroups with lesion.

	AUC level-I	Acc level-I	AUC level-II	Acc level-II
Large lesion group	0.922 ± 0.009	0.766 ± 0.025	0.870 ± 0.008	0.537 ± 0.025
Small lesion group	0.893 ± 0.004	0.702 ± 0.012	0.841 ± 0.005	0.520 ± 0.012

problem to binary classification between fungus and bacteria, or chlamydia and bacteria pneumonia, and the results remained which means the misdiagnosis is mainly due to indistinctive imaging features. This finding is consistent with some former publications that lobar consolidation, cavitation, and pleural effusions usually suggest a bacterial etiology [44], but they are not unique and may overlap with fungal pneumonia [45].

4.2. Level-II performance

For the level-II recognition results (in Fig. 6), top-2 accuracies are also calculated since top-1 recognition is too hard. In clinical practice, physicians might also give multiple possibilities according to CTs, and we believe our level-II prediction can offer useful suggestions for physicians. According to the results, our network gets 0.524 ± 0.011 average accuracy, 0.851 ± 0.004 multi-way AUC, and the average top-2 accuracy is 0.663 ± 0.002. Chlamydia and Aspergillus are the two hardest pathogens with accuracies below 10%, and even taking top-2 choices into account, Aspergillus is still hard to recognize. According to confusion matrix, many of categories were easy to be confused to Streptococcus, which is one of the commonest pathogens. We suppose that the lower accuracy of certain pathogen is due to too few training samples and inherent difficulty, because when we removed all other cases to simplify the problem, the result was the same.

4.3. Performance comparison with other methods

All comparisons are carried out based on three parts of network pipeline which are backbone, slice or volume process, and classification structures (Table 2). According to the experiments, our proposed DDAF outperforms all of them on the pathogen recognition task. We took **Resnet152** [21] as our backbone to extract features. The result for Resnet152 means using pure Resnet [21] to directly predict the 12 categories, which in other words, uses pure dense layer as classifier instead of **DDAF**. This method can also be regarded as a reimplementation of the system in Jin et al. [41]. Different from Group SoftMax, our agents cover categories overlapping with each other and take different parts of features from the backbone, because we suppose different features should be used for fine-level recognition, while Group SoftMax groups only long tailed and non-long tailed categories. We also tried DDAF with manually designed agents, which groups categories based on prior knowledge that all level-II pathogens belong to their level-I. Some other diagnostic methods from other publications were tested, including DRE-Net [46] and Deep-chest [40]. For Deep-chest, we test the performance of its VGG [47]+CNN architecture which is the best model in original paper. We also test several commonly used backbone structures which uses different backbone from DDAF, such as Xception [48], Densenet [49]. Some different strategies on fusion among slices were also tested, because we found many publications used similar backbones and focused on data preprocessing and fusion among slices. Results showed that our deep diagnostic agent worked well comparing to other designs, and both Resnet50-3D [50] and Resnet152+LSTM [51] cannot converge using such few training data, and DDAF outperforms other classifier structures.

4.4. Improvement on deal with data imbalance

As shown in Fig. 7, all accuracies increased compared to the baseline Resnet pure classification method. For pathogens with small number of cases, the accuracies increase significantly, but there are some pathogens whose accuracies are still low. Fungus is one of these situations and we suppose that fungus should have huger intra-category differences and smaller inter-category differences to bacteria, which means it is hard even when training samples are plenty. Since our network helps not only in data imbalance, but also in finding identification margin between fine-level pathogens, the accuracy of fungus also increases compared to baseline. Similar situation happens on Aspergillus and Candida, while for those pathogens with fewer samples, the improvement is significant, such as CMV and Respiratory syncytial. The accuracy of Chlamydia increases only slightly. We suppose that the appearance of CTs may be too similar between Chlamydia and bacteria [44].

4.5. Performance in different infection level sub-groups

To investigate whether the severity of the pneumonia affects the diagnosis accuracy, it is necessary to distinguish pneumonia from each other at the same severity degree. Besides, we suppose that it is valuable to identify pathogens of pneumonia as early as possible. Therefore, we divided our test cohort into large lesion and small lesion sub-groups and evaluated our system on them (show in Table 3). The division was according to areas of infections, which were obtained by a trained infectious area segmentation network. The results in Table 4 and Fig. 8 show that in small-lesion group the accuracy decreases to 0.702 ± 0.012 because the task is more difficult, while in the large-lesion group, the accuracy is 0.766 ± 0.025. The accuracies are different but still on the same level, indicating that our method can work stably on different situations of infected areas.

4.6. Performance comparison with radiologists

To understand the performances of the proposed network, we conducted a reader study to evaluate the accuracy of human radiologists. We invited seven radiologists, who have an average of 10 years of experience to participate the reader study. These radiologists are not specially trained for pneumonia, which is a common situation for most general hospitals.

According to the result, the deep network outperformed most doctors (Fig. 9). In the first reader study, the deep network outperformed most readers with 68.42% accuracy and 36.84% level-II accuracy. Recognition of level-II pathogens is especially challenging even for the most experienced reader. In the second reader study, our network also outperformed most of human readers with 77.08% accuracy.

It is common that radiologists in most hospitals are usually not

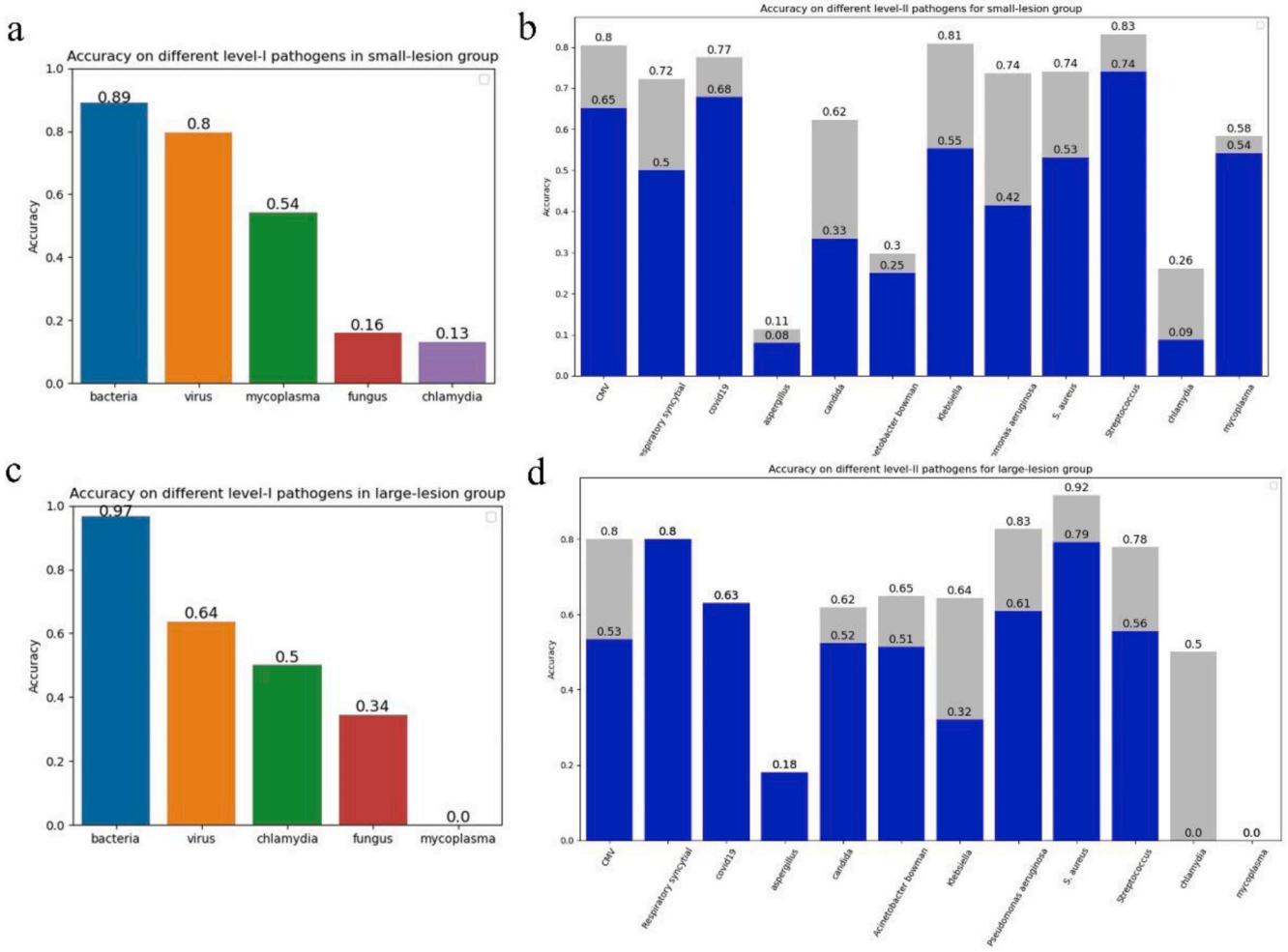


Fig. 8. Accuracies for different subgroups with lesion. a-b) level-I and level-II accuracies for small lesion group. c-d) level-I and level-II accuracies for large lesion group.

trained to identify pathogens unless in some specialist hospitals for infectious diseases. So, the experiments showed how much the algorithm can help in common hospitals in recognizing pathogens. On the other hand, the task of the second reader study is to identify viral pathogens, which is a common task for radiologists. On this task, our method also outperforms most radiologists.

4.7. Feature interpretation

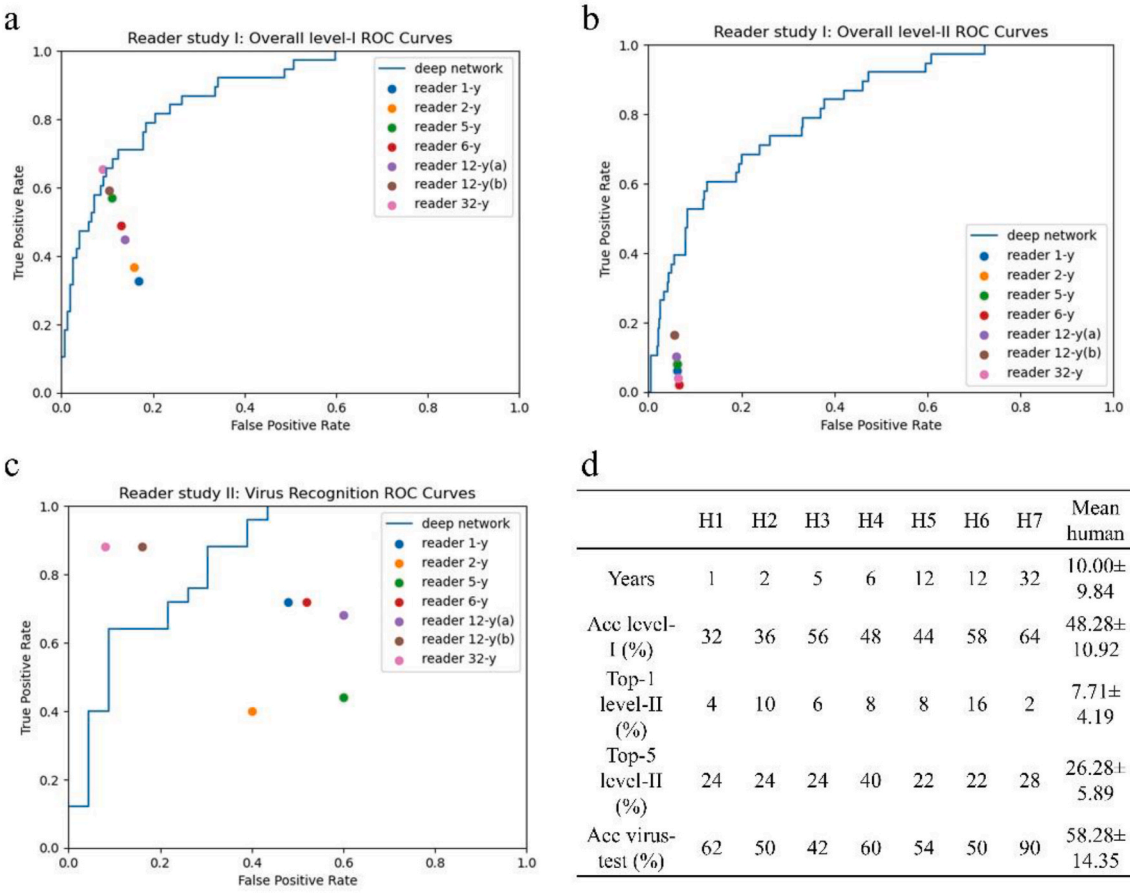
Since deep network is a black box, whose features are difficult to interpret, t-SNE map is a good tool to show the distribution of those deep features. This interpretation helps check whether the different categories are well divided by deep network. In Fig. 10, we show samples in the test cohort in a two-dimension map. The cases from the same level-I categories share the same color and the cases of the same level-II categories use the same shape of points. We can see that points of different colors are almost divided into different clusters in the map except that the purple ones (Chlamydia) and some red ones (Fungus) are spread among the yellows (Bacteria). The figure also shows the imbalanced number of cases of different categories.

5. Conclusion and discussion

In this study, we developed a deep network to extract clues of pathogen recognition from chest CT. The proposed network is specially designed for imbalanced data using multiple diagnostic agents. Each agent is trained by different groups of data cases, which helps identify fine level differences among similar cases and finally help improve accuracy. Despite of a very complicated task, our system achieved 71.3 $\pm$ 0.9% level-I accuracy and 52.4 $\pm$ 0.3% level-II accuracy, which are higher than baselines. When we allow the system to make two most probable predictions, the accuracy increases to 89.53 $\pm$ 1.8% for level-I and 66.3 $\pm$ 0.2% for level-II.

A diagnosis system with such accuracy can be used to optimize the use of biochemical tests. In many situations, patients need to take many different tests to determine pathogens, resulting in high expenditure and long waiting time. Informed by the prediction results of our system, doctors can make optimized biochemical tests, with lower cost and faster process. For example, if one is suspected to suffer from fungal or bacterial pneumonia by our system, culture and smear detection in sputum should be firstly tested. While if viral pneumonia has higher possibility, the IgM detection and nucleic acid detection are preferred. If we can recognize specific fine-grained pathogen as quickly as possible, accurate medication and infection prevention can be taken without delay, which can decrease antibiotic abuse and risk of hospital infection. On the other hand, if possible, pathogens are determined before laboratory test, the time and financial cost for patients can be reduced.

In our reader study, the accuracies of radiologists were relatively low. This indicates that recognizing pathogen based on CT is indeed a hard task for radiologists, which is consistent with previous findings [42]. Experienced pneumonia specialists may have higher accuracy since our readers are from radiology departments of general hospitals



d

	H1	H2	H3	H4	H5	H6	H7	Mean human	DDAF
Years	1	2	5	6	12	12	32	10.00± 9.84	--
Acc level-I (%)	32	36	56	48	44	58	64	48.28± 10.92	68.42
Top-1 level-II (%)	4	10	6	8	8	16	2	7.71± 4.19	36.84
Top-5 level-II (%)	24	24	24	40	22	22	28	26.28± 5.89	57.89
Acc virus-test (%)	62	50	42	60	54	50	90	58.28± 14.35	77.08

Fig. 9. ROC curves of our system and performances of human readers in reader study. a) ROC for level-I diagnosis of the first reader test. b) ROC for level-II diagnosis of the first reader test. c) ROC for virus-not-virus diagnosis of the second reader test. d) Accuracies of human readers and deep network in reader studies. H1, H2 and so on mean different human readers.

and are not specialized on pneumonia. Most patients are turn for helps to general hospitals in practice so that our reader study design is suitable. In addition, radiologists may have higher accuracy in clinical applications since they can refer to other prior information, such as some pathogens are not that frequently seen in certain season or certain populations or in immunocompetent subjects.

Since the beginning of 2020, artificial intelligence methods on COVID-19 diagnosis or detection attract much attention. It is important to note that out task is much more difficult than COVID-19 detection. COVID-19 is pneumonia caused by virus, and it has obvious difference from bacterial pneumonias in CTs. Our task is recognizing 12 kinds of pathogens which means many types of pneumonias which share similar appearance should also be identified from each other. Within the range of viral pneumonia, some imaging features of COVID-19 are also present in other viral pneumonia, and thus it is a much harder classification problem. We also tested the performances of our proposed DDAF on the same test cohort of our previous study [41], and found DDAF performed slightly better than the former one with 0.981 multi-way AUC. This indicates that the classification problem in this study is indeed much more challenging than COVID-19 detection problem addressed in previous study. This work is different from the former in not only topic, but also network structures. In order to promote the performance, we replace the last two layers of the network in Ref. [41] with agent forests which can help reduce the influence of imbalanced data and help identify fine-level differences among pathogens. The key to this design is to use multiple intra-group classifiers together with grouping classifiers, which can reduce the degree of data imbalance when training each classifier and may be beneficial to fine-level classification among different intra-group pathogens. The result also shows that our DDAF improves accuracies a lot both for the averaged or single pathogen.

We could learn some details from the accuracy and confusion matrix of our experiments. For differential diagnosis of level-I, it is not surprising that fungal and bacterial pneumonia were easier to be confused with each other according to the confusion matrix. It is known that lobar consolidation, cavitation, and pleural effusions usually suggest a bacterial etiology, such as Streptococcus and Klebsiella pneumonia [44]. However, the above imaging features are not unique and may overlap with those of other bacterial infections, such as Acinetobacter bowman or S. aureus [44], or fungal pneumonia [45]. Hence, the low diagnostic performance of fungal pneumonia in current study may be mainly due to the poor specificity of the imaging finding. Similarly, for differential diagnosis of level-II, Streptococcus had the highest sensitivity but lowest specificity, namely, many other infections such as Aspergillus, Chlamydia were misclassified as Streptococcus. We supposed this situation might mainly come from the varying imaging appearance of Streptococcus, which overlaps with some other pneumonia. Therefore, more attention should be paid to the identification of these pathogens in future research.

Though there are already many publications on COVID-19 classification, our proposed method has a novel network structure which is designed specifically for the imbalanced category situation in pathogen recognition. A key to our network is a random forest-like classifier, in which multiple simple classifiers are combined to predict at the same time. In order to get some randomness and variety as the random-forest principle [52], our classifiers take different elements of deep features as input using random sampling (described in Method). Partial-fully-connection and multiple classifier bagging are two keys of this design. The randomly sampled features result in

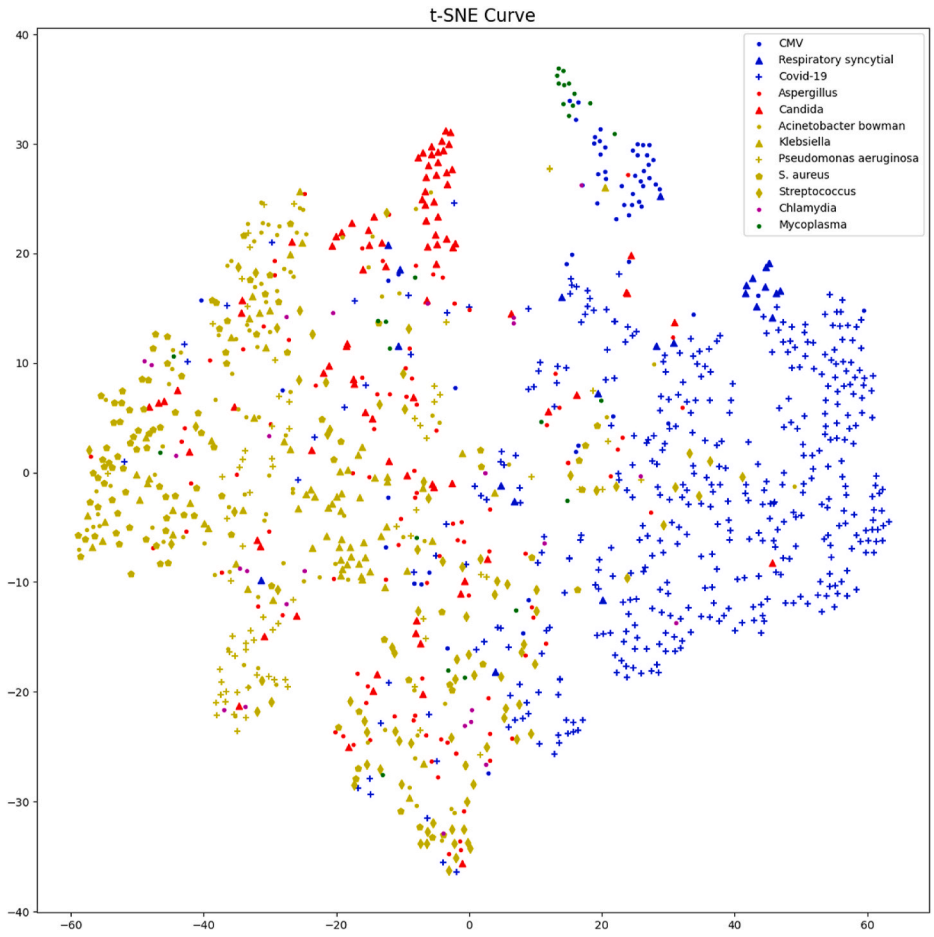


Fig. 10. T-SNE map of deep features of test samples extracted by our network, which are projected into two-dimension. According to the map, we can understand the similarity of different categories vividly.

partial-fully-connection, which decreases the connection in fully-connection layers, helps decrease the complexity, and can prevent overfitting just like the well-known dropout process [53]. Some other researchers also found that more connections between nodes of network cannot always increase the performances [54]. For classifier bagging, the ensemble of them can help increase the stability and accuracy [55, 56]. Besides, different classifiers can be regarded as different fine-level classifiers working well in different groups of similar categories. From the view of clinical practice, it is quite common that some radiologists do especially well in recognition of some pathogens because they are experts on them. Our network is a simulation of this situation that an agent represents radiologists who are good at recognizing some of the pathogens. The final prediction if a committee of those radiologists so that the results can foreseeably better than Resnet which means a single radiologist arranging all. We believe our method is suitable for situations where the number of subtypes is large, and the number of training samples is limited.

Our research has several limitations. Firstly, the accuracies are still relatively low, especially for some pathogens with limited number of samples. Larger database should be collected and better classification methods should be explored in the future. Secondly, clinical information, epidemiological knowledge, immunocompetent situation and so on, are important clues for diagnosis of pathogens, but have not been exploited in our method. We plan to collect more complete data and develop classifier utilizing both CT images and this additional information. Thirdly, this study was retrospectively conducted on databases from two medical centers. To improve the generalization ability and optimization of the model, multicenter and prospective research are needed. Fourthly, although we have strictly excluded cases of multiple or secondary infections, some cases of multiple infections might not be clinically recognized due to the limitation of laboratory testing approach. Finally, although our system can recognize 12 different pathogens, which are among the most common ones, there are still many other pathogens for infections that were not included due to limited numbers of samples.

To our knowledge, this is the first deep learning system for pneumonia pathogen recognition using CTs, and the results shows that our network works better than average human radiologists. Although there is still plenty of room for improvement, the current system can provide useful information for optimizing biochemical tests.

Data and code availability

The databases we used in this research are available after reasonable request and with permission of the Union Hospital of Wuhan, China. The code of this work is publicly available at https://github.com/ChenWWWeixiang/TriageNet_pneumonia.

Ethics

This study obtained ethical approval from the Ethics Commission of Wuhan Union Hospital. All participants remained anonymous, and the requirement for informed patient consent was waived by the ethics committee for this retrospective study. This study was registered with the Chinese Clinical Trial Registry, ChiCTR2000038609.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contributions

W.C., X.H., H.S. and J.F. contributed to the conception of the study; W.C., J.Z. and J.F. designed the algorithms; X.H., Y.C., X.J and Y.Z. contributed to acquisition and annotation of the CT data; J.W., W.Z. and L.W. contributed to acquisition of the laboratory data; H.S., L.W. and J. Z. contributed to analysis and interpretation of the data; W.C., H.X. and J.F. contributed to drafting and revising the manuscript.

Acknowledgements

We would like to acknowledge the radiologists participating the reader study. This study was supported by the National Natural Science Foundation of China (grant 82071921).

References

[1] T.L. Wiemken, P. Peyrani, J.A. Ramirez, Global changes in the epidemiology of community-acquired pneumonia, Semin. Respir. Crit. Care Med. 33 (2012) 213–219, https://doi.org/10.1055/s-0032-1315633.
[2] J.P. Lynch, Hospital-acquired pneumonia: risk factors, microbiology, and treatment, Chest 119 (2001) 373S–384S, https://doi.org/10.1378/chest.119.2_suppl.373S.
[3] GBD 2017 Causes of Death Collaborators, Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017, Lancet 392 (2018) 1736–1788, 10159, https://doi.org/10.1016/S0140-6736(18)32203-7.
[4] T. Vos, S.S. Lim, C. Abbafati, K.M. Abbas, M. Abbasi, M. Abbasifard, M. Abbasi-Kangevari, H. Abbastabar, F. Abd-Allah, A. Abdelalim, Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019, Lancet 396 (2020) 1204–1222. https://doi.org/10.1016/S0140-6736(20)30925-9.
[5] B. Dadonaite, M. Roser, Pneumonia, Our World in Data, 2018. https://ourworldindata.org/pneumonia.
[6] A. Torres, C. Cilloniz, M.S. Niederman, R. Menéndez, J.D. Chalmers, R. G. Wunderink, T. van der Poll, Pneumonia, Nat. Rev. Dis. Prim. 7 (2021) 25, https://doi.org/10.1038/s41572-021-00259-0.
[7] G. Mackenzie, The definition and classification of pneumonia, Pneumonia 8 (14) (2016), https://doi.org/10.1186/s41479-016-0012-z s41479-016-0012-z.
[8] S. Carbonara, L. Monno, B. Longo, G. Angarano, Community-acquired pneumonia, Curr. Opin. Pulm. Med. 15 (2009) 261–273.
[9] A. Nambu, Imaging of community-acquired pneumonia: roles of imaging examinations, imaging diagnosis of specific pathogens and discrimination from noninfectious diseases, WJR 6 (2014) 779, https://doi.org/10.4329/wjr.v6.i10.779.
[10] A. Torres, N. Lee, C. Cilloniz, J. Vila, M. Van der Eerden, Laboratory diagnosis of pneumonia in the molecular age, Eur. Respir. J. 48 (2016) 1764–1778, https://doi.org/10.1183/13993003.01144-2016.
[11] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M.P. Lungren, A.Y. Ng, CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning, ArXiv:1711.05225 [Cs, Stat], 2017. http://arxiv.org/abs/1711.05225. (Accessed 3 May 2021).
[12] Y.-E. Claessens, M.-P. Debray, F. Tubach, A.-L. Brun, B. Rammaert, P. Hausfater, J.-M. Naccache, P. Ray, C. Choquet, M.-F. Carette, C. Mayaud, C. Leport, X. Duval, Early chest computed tomography scan to assist diagnosis and guide treatment decision for suspected community-acquired pneumonia, Am. J. Respir. Crit. Care Med. 192 (2015) 974–982, https://doi.org/10.1164/rccm.201501-0017OC.
[13] P. Goodman, H. Prosch, D. Kienzl, C. Herold, Imaging of Pulmonary Infections. Diseases of the Heart and Chest, Including Breast 2011–2014, Springer, Milano, 2011, pp. 60–65.
[14] Y. Kunihiro, N. Tanaka, R. Kawano, T. Yujiri, M. Kubo, K. Ueda, T. Gondo, T. Kobayashi, T. Matsumoto, Differential diagnosis of pulmonary infections in immunocompromised patients using high-resolution computed tomography, Eur. Radiol. 29 (2019) 6089–6099, https://doi.org/10.1007/s00330-019-06235-3.
[15] J. Qin, J. Xu, Y. Dong, W. Tang, B. Wu, Y. An, H. Shan, High-resolution CT findings of pulmonary infections after orthotopic liver transplantation in 453 patients, BJR (Br. J. Radiol.) 85 (2012) e959–e965, https://doi.org/10.1259/bjr/26230943.
[16] T. Saraya, K. Ohkuma, Y. Tsukahara, T. Watanabe, D. Kurai, H. Ishii, H. Kimura, H. Goto, H. Takizawa, Correlation between clinical features, high-resolution computed tomography findings, and a visual scoring system in patients with pneumonia due to Mycoplasma pneumoniae, Respiratory Investigation 56 (2018) 320–325, https://doi.org/10.1016/j.resinv.2018.03.001.
[17] J. Yanase, E. Triantaphyllou, A systematic survey of computer-aided diagnosis in medicine: past and present developments, Expert Syst. Appl. 138 (2019) 112821.
[18] G. Haskins, U. Kruger, P. Yan, Deep learning in medical image registration: a survey, Mach. Vis. Appl. 31 (2020) 1–18. https://doi.org/10.1007/s00138-020-01060-x.
[19] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88. https://doi.org/10.1016/j.media.2017.07.005.
[20] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, ImageNet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (2015) 211–252, https://doi.org/10.1007/s11263-015-0816-y.
[21] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 770–778, 7, https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html.
[22] J.H. Lee, H.Y. Sun, S. Park, H. Kim, E.J. Hwang, J.M. Goo, C.M. Park, Performance of a deep learning algorithm compared with radiologic interpretation for lung cancer detection on chest radiographs in a health screening population, Radiology 297 (2020) 687–696, https://doi.org/10.1148/radiol.2020201240.
[23] J.G. Nam, S. Park, E.J. Hwang, J.H. Lee, K.-N. Jin, K.Y. Lim, T.H. Vu, J.H. Sohn, S. Hwang, J.M. Goo, C.M. Park, Development and validation of deep learning–based automatic detection algorithm for malignant pulmonary nodules on chest radiographs, Radiology 290 (2019) 218–228, https://doi.org/10.1148/radiol.2018180237.
[24] P.M. Shakeel, A. Tolba, Z. Al-Makhadmeh, M.M. Jaber, Automatic detection of lung cancer from biomedical data set using discrete AdaBoost optimized ensemble learning generalized neural networks, Neural Comput. Appl. 32 (2020) 777–790, https://doi.org/10.1007/s00521-018-03972-2.
[25] J. Dolz, K. Gopinath, J. Yuan, H. Lombaert, C. Desrosiers, I. Ben Ayed, HyperDense-net: a hyper-densely connected CNN for multi-modal image segmentation, IEEE Trans. Med. Imag. 38 (2019) 1116–1126, https://doi.org/10.1109/TMI.2018.2878669.
[26] O. Cicek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-net: learning dense volumetric segmentation from sparse annotation. International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 424–432. https://link.springer.com/chapter/10.1007%2F978-3-319-46723-8_49.
[27] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241. https://linkspringer.fenshishang.com/chapter/10.1007/978-3-319-24574-4_28.
[28] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (2017) 115–118. https://doi.org/10.1038/nature21056.
[29] D. Ardila, End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography, Nat. Med. 25 (2019) 24. https://doi.org/10.1038/s41591-019-0447-x.
[30] S.M. McKinney, M. Sieniek, V. Godbole, J. Godwin, N. Antropova, H. Ashrafian, T. Back, M. Chesus, G.S. Corrado, A. Darzi, International evaluation of an AI system for breast cancer screening, Nature 577 (2020) 89–94. https://doi.org/10.1038/s41586-019-1799-6.
[31] A. Ahmadi, M. Kashefi, H. Shahrokhi, M.A. Nazari, Computer aided diagnosis system using deep convolutional neural networks for ADHD subtypes, Biomed. Signal Process Control 63 (2021) 102227, https://doi.org/10.1016/j.bspc.2020.102227.
[32] A. Ahmadi, S. Davoudi, M.R. Daliri, Computer Aided Diagnosis System for multiple sclerosis disease based on phase to amplitude coupling in covert visual attention, Comput. Methods Progr. Biomed. 169 (2019) 9–18. https://doi.org/10.1016/j.cmpb.2018.11.006.
[33] D. Dong, Z. Tang, S. Wang, H. Hui, L. Gong, Y. Lu, Z. Xue, H. Liao, F. Chen, F. Yang, R. Jin, K. Wang, Z. Liu, J. Wei, W. Mu, H. Zhang, J. Jiang, J. Tian, H. Li, The role of imaging in the detection and management of COVID-19: a review, IEEE Reviews in Biomedical Engineering 14 (2021) 16–29, https://doi.org/10.1109/RBME.2020.2990959.
[34] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, D. Shen, Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19, IEEE Reviews in Biomedical Engineering 14 (2021) 4–15, https://doi.org/10.1109/RBME.2020.2987975.
[35] L. Saba, M. Agarwal, A. Patrick, A. Puvvula, S.K. Gupta, A. Carriero, J.R. Laird, G. D. Kitas, A.M. Johri, A. Balestrieri, Z. Falaschi, A. Paschè, V. Viswanathan, A. El-Baz, I. Alam, A. Jain, S. Naidu, R. Oberleitner, N.N. Khanna, A. Bit, M. Fatemi, A. Alizad, J.S. Suri, Six artificial intelligence paradigms for tissue characterisation and classification of non-COVID-19 pneumonia against COVID-19 pneumonia in computed tomography lungs, Int J CARS 16 (2021) 423–434, https://doi.org/10.1007/s11548-021-02317-0.
[36] H. Panwar, P.K. Gupta, M.K. Siddiqui, R. Morales-Menendez, V. Singh, Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet, Chaos, Solitons & Fractals 138 (2020) 109944. https://doi.org/10.1016/j.chaos.2020.109944.
[37] M. La Salvia, G. Secco, E. Torti, G. Florimbi, L. Guido, P. Lago, F. Salinaro, S. Perlini, F. Leporati, Deep learning and lung ultrasound for Covid-19 pneumonia detection and severity classification, Comput. Biol. Med. 136 (2021) 104742. https://doi.org/10.1016/j.compbiomed.2021.104742.

[38] S.H. Khan, A. Sohail, A. Khan, M. Hassan, Y.S. Lee, J. Alam, A. Basit, S. Zubair, COVID-19 detection in chest X-ray images using deep boosted hybrid learning, Comput. Biol. Med. 137 (2021) 104816. https://doi.org/10.1016/j.compbiomed.2021.104816.

[39] M. Toğaçar, B. Ergen, Z. Cömert, COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches, Comput. Biol. Med. 121 (2020) 103805. https://doi.org/10.1016/j.compbiomed.2020.103805.

[40] D.M. Ibrahim, N.M. Elshennawy, A.M. Sarhan, Deep-chest: multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases, Comput. Biol. Med. 132 (2021) 104348. https://doi.org/10.1016/j.compbiomed.2021.104348.

[41] C. Jin, W. Chen, Y. Cao, Z. Xu, Z. Tan, X. Zhang, L. Deng, C. Zheng, J. Zhou, H. Shi, J. Feng, Development and evaluation of an artificial intelligence system for COVID-19 diagnosis, Nat. Commun. 11 (2020) 5088, https://doi.org/10.1038/s41467-020-18685-1.

[42] H.J. Koo, S. Lim, J. Choe, S.-H. Choi, H. Sung, K.-H. Do, Radiographic and CT features of viral pneumonia, Radiographics 38 (2018) 719–739. https://doi.org/10.1148/rg.2018170048.

[43] T. Franquet, Imaging of community-acquired pneumonia, J. Thorac. Imag. 33 (2018) 282–294. https://doi.org/10.1097/RTI.0000000000000347.

[44] C.M. Walker, G.F. Abbott, R.E. Greene, J.-A.O. Shepard, D. Vummidi, S. R. Digumarthy, Imaging pulmonary infection: classic signs and patterns, Am. J. Roentgenol. 202 (2014) 479–492. https://www.ajronline.org/doi/full/10.2214/AJR.13.11463.

[45] N.P. Dueck, S. Epstein, T. Franquet, C.C. Moore, J. Bueno, Atypical pneumonia: definition, causes, and imaging features, Radiographics 41 (2021) 720–741. https://doi.org/10.1148/rg.2021200131.

[46] Y. Song, S. Zheng, L. Li, X. Zhang, X. Zhang, Z. Huang, J. Chen, R. Wang, H. Zhao, Y. Zha, J. Shen, Y. Chong, Y. Yang, Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images, IEEE/ACM Trans. Comput. Biol. Bioinf 18 (6) (2021) 2775–2780. https://doi.org/10.1109/TCBB.2021.3065361.

[47] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Y. Bengio, Y. Lecun (Eds.), International Conference on Learning Representations, Conference Track Proceedings, 2015. http://arxiv.org/abs/1409.1556.

[48] F. Chollet, Xception: deep learning with depthwise separable convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, {IEEE} Computer Society, 2017, pp. 1251–1258. https://doi.org/10.1109/CVPR.2017.195.

[49] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: IEEE Conference on Computer Vision and Pattern Recognition, {IEEE} Computer Society, 2017, pp. 4700–4708. https://openaccess.thecvf.com/content_cvpr_2017/html/Huang_Densely_Connected_Convolutional_CVPR_2017_paper.html.

[50] K. Hara, H. Kataoka, Y. Satoh, Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet?, in: IEEE Conference on Computer Vision and Pattern Recognition {IEEE} Computer Society, 2018, pp. 6546–6555. https://openaccess.thecvf.com/content_cvpr_2018/html/Hara_Can_Spatiotemporal_3D_CVPR_2018_paper.html.

[51] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

[52] V. Svetnik, A. Liaw, C. Tong, J.C. Culberson, R.P. Sheridan, B.P. Feuston, Random forest: a classification and regression tool for compound classification and QSAR modeling, J. Chem. Inf. Comput. Sci. 43 (2003) 1947–1958. https://doi.org/10.1021/ci034160g.

[53] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (2014) 1929–1958. https://dl.acm.org/doi/10.5555/2627435.2670313.

[54] J. You, J. Leskovec, K. He, S. Xie, Graph structure of neural networks, in: Proceedings of Machine Learning Research, PMLR, 2020, pp. 10881–10891, in: https://proceedings.mlr.press/v119/you20b.html.

[55] T.G. Dietterich, Ensemble methods in machine learning, in: International Workshop on Multiple Classifier Systems, Springer, 2000, pp. 1–15. https://linkspringer.fenshishang.com/chapter/10.1007/3-540-45014-9_1.

[56] L. Rokach, Ensemble-based classifiers, Artif. Intell. Rev. 33 (2010) 1–39. https://doi.org/10.1007/s10462-009-9124-7.