Medical & Biological Engineering & Computing (2023) 61:1395–1408
https://doi.org/10.1007/s11517-022-02746-2

ORIGINAL ARTICLE

PneuNet: deep learning for COVID-19 pneumonia diagnosis on chest X-ray image analysis using Vision Transformer

Tianmu Wang[1,2,3] · Zhenguo Nie[1,2,3] · Ruijing Wang[4] · Qingfeng Xu[1,5] · Hongshi Huang[6] · Handing Xu[1,2,3] · Fugui Xie[1,2,3] · Xin-Jun Liu[1,2,3]

Received: 23 June 2022 / Accepted: 22 December 2022 / Published online: 31 January 2023
© International Federation for Medical and Biological Engineering 2023

Abstract

A long-standing challenge in pneumonia diagnosis is recognizing the pathological lung texture, especially the ground-glass appearance pathological texture. One main difficulty lies in precisely extracting and recognizing the pathological features. The patients, especially those with mild symptoms, show very little difference in lung texture, neither conventional computer vision methods nor convolutional neural networks perform well on pneumonia diagnosis based on chest X-ray (CXR) images. In the meanwhile, the Coronavirus Disease 2019 (COVID-19) pandemic continues wreaking havoc around the world, where quick and accurate diagnosis backed by CXR images is in high demand. Rather than simply recognizing the patterns, extracting feature maps from the original CXR image is what we need in the classification process. Thus, we propose a Vision Transformer (VIT)–based model called PneuNet to make an accurate diagnosis backed by channel-based attention through X-ray images of the lung, where multi-head attention is applied on channel patches rather than feature patches. The techniques presented in this paper are oriented toward the medical application of deep neural networks and VIT. Extensive experiment results show that our method can reach 94.96% accuracy in the three-categories classification problem on the test set, which outperforms previous deep learning models.

Keywords Deep learning · Pneumonia diagnosis · COVID-19 · Vision Transformer · Multi-head attention

1 Introduction

According to the World Health Organization (WHO), pneumonia accounts for 14% of all deaths of children under 5 years old and is blamed as one main murderer of children [1]. Pneumonia can be divided into two categories which are bacterial pneumonia and viral pneumonia. In the past decades, the viral pneumonia, SARS, for example, has not only murdered children but also claimed the lives of people of all ages, especially those who are suffering from chronic disease [2]. In 2019, a novel coronavirus called COVID-19 was first reported in Wuhan, China, in December 2019. The plague has continued to have a devastating effect on global health and has caused over 6 million death cases all over the world, out of over 611 million infected people, according to Johns Hopkins University [3]. Patients infected by COVID-19 share similar symptoms as usual pneumonia. The ignorance of such similar symptoms leads to the rapid spreading of this lethal virus and has become one of the leading causes of this global pandemic.

During the global pandemic caused by COVID-19, isolation has been proven to be the most effective method to control the spreading once an accurate diagnosis is conducted. RT-PCR test and antibody test have become two wildly used solutions to make a quick diagnosis. However, the sensitivity of RT-PCR is still under debate. It is reported that there is a 3% false-negative rate exists in the RT-PCR test [4]. In the meanwhile, the result of the antibody test cannot be convincing if the suspected patient gets infected in the first 7 days. An alternative choice to make the diagnosis is based on the evaluation of radiographic images of the lung, such as chest X-ray (CXR) images and computerized tomography (CT) images. Previous research has concluded that pulmonary manifestation of COVID-19 infection is predominantly characterized by ground-glass opacification with occasional consolidation [5, 6].

✉ Zhenguo Nie
zhenguonie@tsinghua.edu.cn

Extended author information available on the last page of the article.

Researchers are confident that preliminary screening can be applied to suspected cases based on the evaluation of CXR images [7].

Preliminary screening based on CXR images has the advantage over other screening methods not only from the perspective of precision but also from its availability and efficiency. CXR imaging has been counted as part of the primary health care system and is readily available in community clinics. A large batch of CXR images can be evaluated at the same time with the help of computer vision methods and artificial neural networks [8].

An accurate classification can be conducted by convolutional neural networks (CNN) with the help of spatial feature extraction and pattern recognition. However, conventional CNN and CV methods do not care about the number of spatial features responsible for classification. In this paper, we propose a multi-head attention-based network called PneuNet, inspired by Vision Transformer (VIT), to conduct a diagnosis of COVID-19. We treat each channel after convolution calculation as one patch, and then the transformer module is applied to evaluate the contribution of each patched channel in classification. Unlike paying attention to the primary feature block of images, we pay more attention to the global extracted spatial features, which could be more efficient in the classification process. Meanwhile, attention applied to convoluted channels has practical implications, where each channel theoretically represents a particular feature from the raw image on the scale of a higher dimension. An extensive experiment shows the proposed model can reach 94.96% test accuracy in three category classifications, where none pneumonia, normal pneumonia, and COVID-19 are classified. Our model also reaches 99.30% accuracy when applied to binary classification, which is used to detect if pneumonia is caused by a coronavirus.

The full implementation and trained network are published on https://github.com/TianmuWang/PneuNet. Our main contributions are as follows:

1. We propose a novel deep learning method, which jointly employs ResNet18 and multi-head attention to make the diagnosis of COVID-19 based on CXR images.
2. We treat the extracted features as patches, and apply channel-based attention to implement the transformer encoder after the application of ResNet18.
3. We evaluate the model not only from the perspective of prediction accuracy but also backed by statistical criteria. In this paper, the prediction based on up to four-category classification is evaluated, corresponding to diagnosing COVID-19 from COVID-19, none pneumonia, bacterial pneumonia, and viral pneumonia.

2 Related work

Inspired by the rapid development of image recognition and classification backed by artificial intelligence, many intelligent pneumonia diagnosis methods have been proposed [9], which are based on making classification among radiography images backed by deep learning methods. The prediction results of the above models are proven to be convincing from the aspect of prediction accuracy. Our review highlights popular deep learning models on image classification and their applications in pneumonia diagnosis from radiography images.

The deep learning method is of a broader family of machine learning methods based on artificial neural networks with representation learning, inspired by the human brain's structure and function [10, 11]. Convolutional neural network (CNN) is one of the most effective deep learning methods dealing with image classification resulting from extracting spatial features through convolution calculation [12]. With the help of deeper network layers and its more complicated structure, CNN outperforms conventional computer vision methods such as Generalized Search Tree (GIST) [13] and Histogram of Oriented Gradient (HOG) [14]. The art of deep learning methods is the generalization capability brought from deeper hidden layers [15] and the gradient descent methods through backpropagation. He et al. [16] propose ResNet, where residual blocks are used to prevent potential gradient vanishing and gradient exploding during the training of a deep CNN model. Due to the outstanding capacity of extracting spatial features, the ResNet family is commonly used in the field of pattern recognition.

Motivated by CNN and ResNet models, much research on pneumonia diagnosis focuses on radiography image classification based on CNN models [17–22]. A lightweight CNN-based model proposed by Bhosale et al. reaches high prediction accuracy under RaspberryPi [23]. Zhang et al. [24] apply ResNet18 to COVID-19 diagnosis and reach a 95.18% accuracy on binary classification. Hemdan et al. [25] make a fine tune on ResNet50 and rename it COVIDX-Net. Narin et al. [26] compare ResNet18, InceptionV3 [27], and Inception-ResNetV2 szegedy2017inception based on a small-scale pneumonia datasets. Wang et al. [28] propose a deep CNN-based model called COVID-Net, where CNN layers from different depths are tailored and obtain 83.5% accuracy in classifying COVID-19, normal, pneumonia-bacterial, and pneumonia-viral classes. Apostolopoulos and Mpesiana [29] apply transfer learning based on pre-trained VGG19 models and obtain the best accuracy of 98.75% and 93.48% for two and three classes, respectively. Ozturk et al. [30] propose DarkCovidNet based on CNN layers with LeakyReLU as an activation function and obtain 98.08%

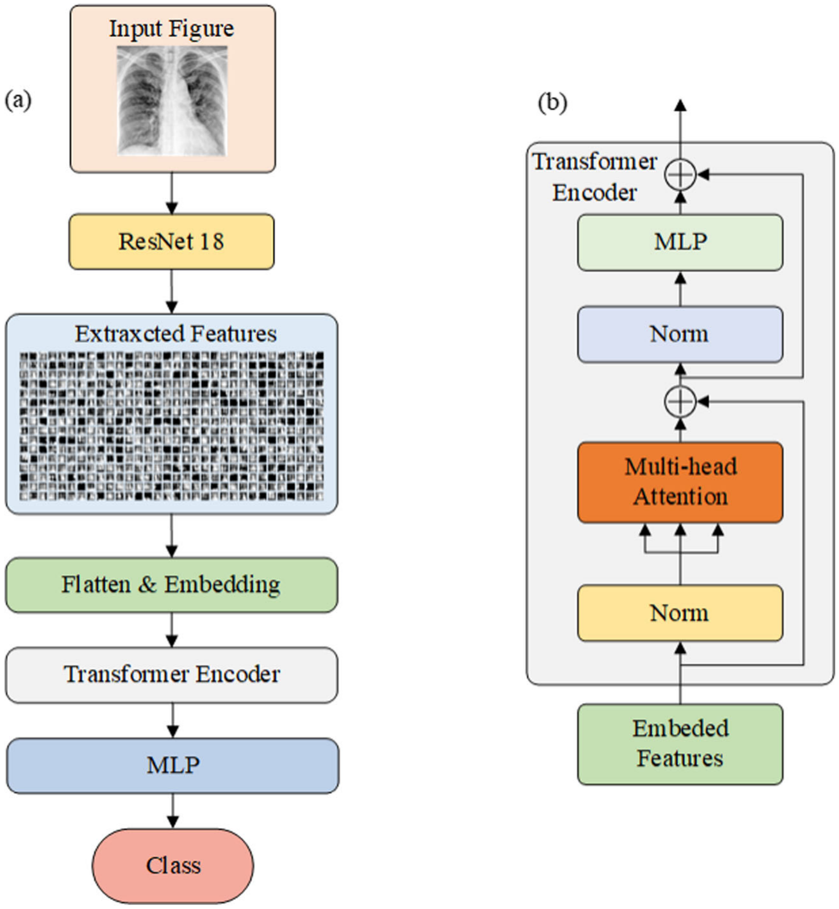


Fig. 1 Architecture of PneuNet (a) and the details of Transformer Encoder (b)

and 87.02% accuracy for binary classification and three-category classification, respectively. CoroNet proposed by Khan et al. [31] evolves the model from the Xception structure and reaches a 92% accuracy on multi-category classification for pneumonia diagnosis. Shazia et al. [32] analyze VGG, ResNet, DenseNet101, and Inception model and compare their performance in diagnosing COVID-19, where DenseNet101 and ResNet51 reveal the latent transfer ability to make an accurate prediction of COVID-19 diagnosis. Considering that we want to find a lightweight model available in suburban and undeveloped areas, we need to compress the scale of the model. Therefore, we consider employing ResNet18, a lightweight standard ResNet model, to first extract useful spatial features from raw images.

Recurrent neural networks (RNN) is another wildly used deep learning method [33]. Unlike CNN models, RNN models consider the context within the message and are initially developed to solve natural language processing (NLP) questions such as voice recognition and translation. However, RNN has been developed and is ready to work on pattern recognition and object detection in computer vision. Long short-term memory (LSTM) is one typical application of RNN networks to detect the object in an image and shows great potential in medical diagnosing [34–36]. Mousavi et al. [37] combine CNN and LSTM and make an analysis based on seven binary categorical classifications among COVID-19, viral pneumonia, bacterial pneumonia, and healthy people.

However, both CNN and LSTM are limited in recognizing a pattern on a dynamic and large scale. The size of the kernel blocks the receptive field of CNN, and LSTM can only realize the context in a narrow range. Apart from CNN and LSTM, transformer [38] has become a hot topic in computer vision, where attention modular is employed to analyze the hot spot region of each image and then make a classification based on it. Compared with LSTM, Transformer is much more accessible in parallel computation and is powerful in global context reconstruction. The transformer is firstly applied in an NLP problem where several powerful general-purpose models are introduced, such as BERT [39] and GPT-2 [40], where multi-head attention is proposed that allows the model to jointly attend to information from different representation subspaces at different positions. Whereas the transformer is initially proposed to solve NLP problems, and it has been transferred to solve computer vision problems. Although the transformer structure is suitable for image recognition and classification of objects, the CNN module still plays an important role in classification [41]. Dosovitskiy et al. [42] bring up the conception of Vision in Transformer (VIT), where multi-head attention can be applied to small patches, which are divided from original images. Inspired by the combination of CNN models and transformer structure, lots of research have been carried out. Sitaula et al. [43] propose an attention-based VGG-16 model and get 79.58% accuracy on multi-class pneumonia diagnosis. Zhang et al. combine the swin transformer block with U-Net together and got a maximum F-1 score of 0.935 based on training on 1560 CT scans. However, precise prediction accuracy is not mentioned. Park et al. [44] introduce a probabilistic-CAM (PCAM) pooling backbone network before applying the transformer and obtain an AUC score of 0.941 for three-category classification based on CXR images. The transformer can be jointly used with basic CNN models, outperforms simple CNN-based models, and follows an interesting internal logic similar to human cognition during the training process. Consequently, we combine transformer and ResNet in our proposed PneuNet model to help diagnose COVID-19 based on CXR images.

3 Technical approach

Our proposed PneuNet is backed by ResNet18 and VIT models. As the previous study does, ResNet18 works as the backbone of the whole model, extracting spatial features with the help of deep convolutional layers. However, unlike other VIT models, we keep the extracted features from splitting them into patches but take the whole channel, downsized from deep convolution calculation and max pooling, as one individual patch. After being encoded and embedded, patches will pass multi-head attention layers before the final classification process with the help of three fully connected layers, which can also be called multi-layer perceptron (MLP). Figure 1 illustrates the overall structure of our proposed model. The detailed architecture of our proposed model is presented in Table 1.

Table 1 Detailed architecture of PneuNet

Layer (submodel)	Output shape	Number of parameter
Input Layer	22 × 224 × 1	0
ResNet18	7 × 7 × 512	11184640
Batch Normalization	7 × 7 × 512	2048
Embedding&Encoder	512 × 80	44960
Transformer_1	512 × 80	439680
Transformer_2	512 × 80	439680
Transformer_3	512 × 80	439680
Transformer_4	512 × 80	439680
Transformer_5	512 × 80	439680
Transformer_6	512 × 80	439680
Layer normalization	512 × 80	160
Flatten	40960	0
Dropout	40960	0
Dense	1024	41944064
Dropout	1024	0
Dense	64	65600
Dropout	64	0
Dense	16	1040
Dropout	16	0
LogitDense	3	51

3.1 Application of ResNet18

ResNet18 is first proposed by He et al. [16] and has been proven to have excellent performance on spatial feature extraction, which is a series of deep neural networks that are derived from the base repeated building blocks. ResNet18 contains four kinds of Residual Blocks, and each Residual block is repeated twice, as shown in Fig. 2. The entire architecture is shown in Table 2. Compared with other deep convolutional networks such as VGG16, ResNet18 can prevent the gradient vanishing and exploding during the backpropagation process. In the meanwhile, ResNet18 requires much less computational capacity compared with ResNet51 and ResNet101 and can be easily trained on a local PC. ResNet18 can extract 512 channels of different spatial features which will be thought of as patches. At the end of ResNet18, an additional max-pooling layer is used to downsample the extracted spatial features better. Considering the input of neural networks is a batch of gray-scale images, of which the shape can be defined as

$$Input \in R^{h\times w\times 1} \tag{1}$$

where h denotes the height of the input image and w denotes the width. The output of this submodule can be described as

$$Output_{ResNet18} \in R^{28\times 28\times 512} \tag{2}$$

3.2 Application of transformer

In conventional VIT models, patches are generated from raw images, where images are divided into pieces in each channel, as shown in Fig. 3. However, in our proposed PneuNet, we think of each channel of extracted spatial features entirely as one patch, shown in Fig. 4. Convolution is an efficient way to extract spatial features from multiple dimensions but lacks the capacity to tell how important the feature stands for during the classification process.

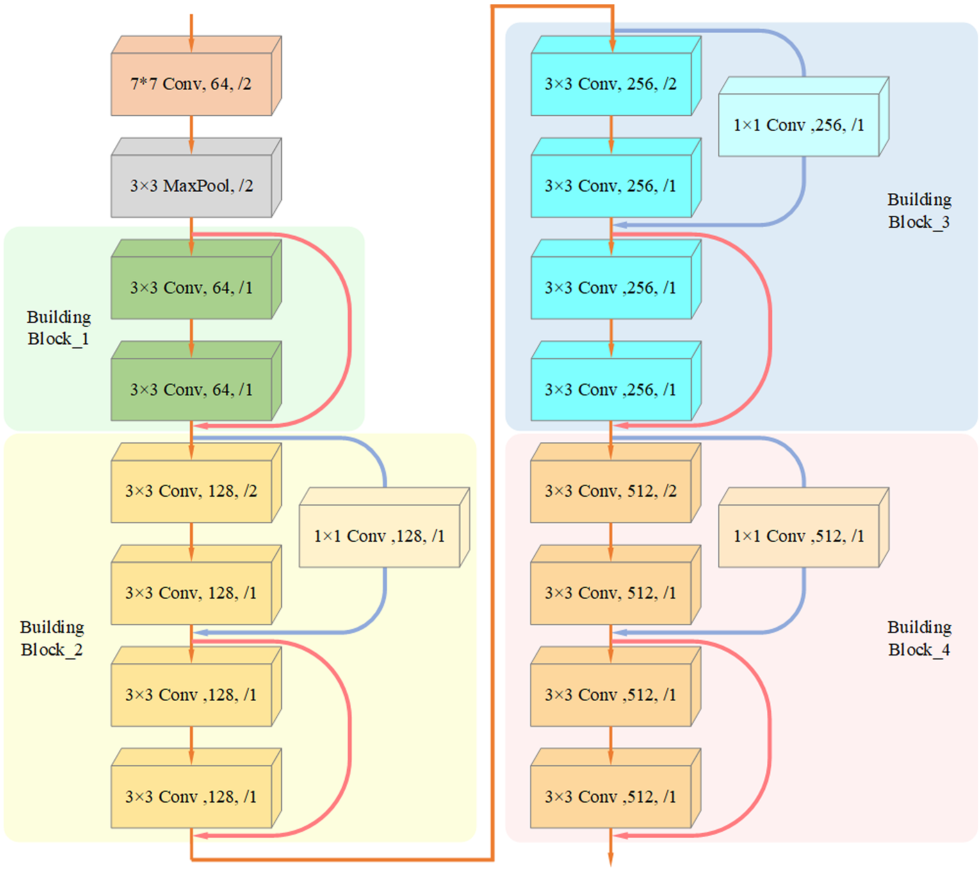


Fig. 2 Architecture of ResNet18, without flatten layer nor logit layer

However, the transformer can evaluate how much one feature patch contributes to classifying with the help of multi-head attention. The number of heads represents the number of subspace that allows the model to focus jointly on information from different positions. In our proposed PneuNet, we employ four transformer layers with four-head attention. Each two-dimensional patch is embedded into a one-dimensional vector with a length of 80.

The intermediate output of transformer modular can be described as

$$Output_{inter} = T_i(head, Output_{ResNet18}) \in R^{512\times p} \tag{3}$$

Table 2 Detailed architecture of ResNet18

Layer (submodel)	Output shape
Conv2D	112 × 112 × 64
Max Pooling	56 × 56 × 64
Building Block_1	56 × 56 × 64
Building Block_2	28 × 28 × 64
Building Block_3	14 × 14 × 64
Building Block_4	7 × 7 × 64
Total parameters	11184640

where $head$ denotes the number of heads applied in attention modular, p denotes the number of projection dimensions, and $Output_{inter}$ denotes the intermediate output of transformer modular before an extra flatten layer to map higher-dimensional features into a one-dimensional vector, counted as $Output_{transformer}$, which can be indicated as

$$Output_{Transformer} = T_{tr}(Output_{inter}) \in R^N \tag{4}$$

$$N = 512 \times p \tag{5}$$

where N denotes the dimension of the flattened vector. In PneuNet, multiple Transformers are employed in model architecture. The least repeating unit of the Transformer submodel is shown in Table 3.

3.3 Application of multi-layer perceptron

Multi-layer perceptron consists of several fully connected layers, which are also called Dense layers. Particularly in our PneuNet. MLP section consists of three Dense layers, where ReLU is used as the activation function and a 20% dropout is applied after each Dense layer.

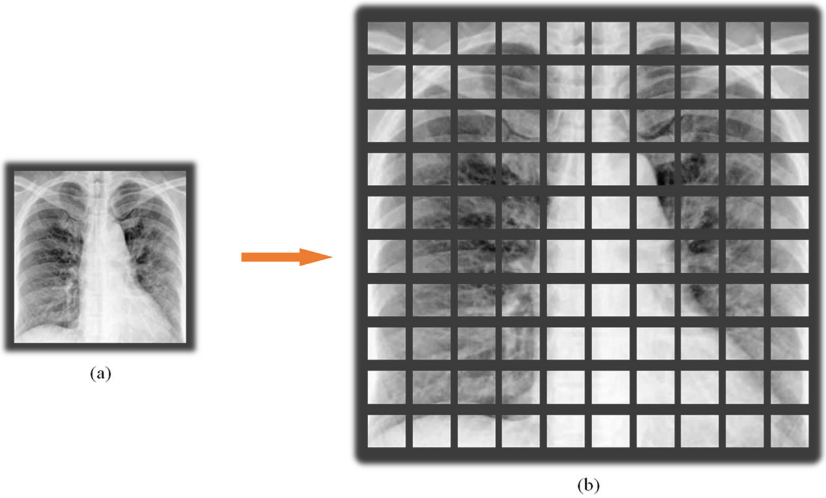


Fig. 3 Image partition in conventional VIT process: raw image (a) is divided into several patches (b) and each patch will be encoded and embedded with its position

3.4 Application of logit classification layer

A particular fully connected layer can be treated as a logit layer where Softmax acts as an activation function. Softmax is a wildly used logit function that maps the multinomial distribution of the probability score to a vector. The length of the vector equals the number of categories to be classified. The output of this distribution can be described as

$$P(y) = \frac{e^{W_y}}{\sum_{y=1}^{C} e^{W_y}} \tag{6}$$

where y denotes y^{th} category, C denotes the number of categories, and W_y denotes the intermediate weight for y^{th} category calculated from the Dense layer.

4 Experiments setup

4.1 Dataset

In this study, CXR images from seven online public repositories are assembled as our datasets. COVID-19 CXR images collected from [45–51], normal pneumonia CXR images collected from [52–54], and CXR images of healthy people from [52–54] have been divided into three sub-datasets out of total 33920 CXR images. We divided our datasets into three categories which are used for training and validation during training, and for testing the generalization performance of the well-trained model. The dataset is divided in a ratio of 64:16:20. Details of our dataset are shown in Table 4. However, most online datasets except [48] do not distinguish bacterial and viral pneumonia out of normal pneumonia. When we generate another dataset

Fig. 4 Illustration of patches employed in Transformer modular which obtained from ResNet18

Table 3 Architecture of Transformer submodel

Layer (submodel)	Output shape	Number of parameter
Layer Normalization	512 × 80	160
Multi-head Attention	512 × 80	413520
Add	512 × 80	0
Layer Normalization	512 × 80	160
Dense	512 × 160	12960
Dropout	512 × 160	0
Dense	512 × 80	12880
Dropout	512 × 80	0
Add	512 × 80	0

with four categories, we employ a down-sampling method to collect COVID-19 CXR images and healthy CXR images from the latter datasets to balance this four-categorical dataset. Details of this dataset are shown in Table 5. Some typical CXR images from four categories are shown in Fig. 5. In this study, to realize the intelligent screening of COVID-19 patients, we focus on the classification of three categories, which are none pneumonia, COVID-19, and normal pneumonia. Discussion of the four-category classification from none pneumonia, COVID-19, bacterial pneumonia, and viral pneumonia is also mentioned in the following sections.

4.2 Evaluation metrics

Cross-entropy (CE) acts as a loss function and plays an essential role in training a deep neural network during the backpropagation process. It can be described as

$$CE = \frac{1}{N} \sum_{i} \sum_{c=1}^{M} y_{ic} \log(P_{ic}) \tag{7}$$

where N denotes dimension, M denotes the number of categories, and y_{ic} is a symbolic function where y_{ic} equals 1 if the sample i and the sample c belong to the same category based on ground truth, otherwise equals 0. And P_{ic} denotes predicted probability where sample i belongs to category c.

For a better description of the effectiveness of the proposed PneuNet model, five statistical evaluation criteria are used to evaluate the performance of the PneuNet model, which are Prediction Accuracy, Recall rate, Precision, and F-1 Score. These criteria are defined as follows:

Table 4 Details of three-categorical dataset

Image category	Training	Validation	Test	Amount
COVID-19	7658	1903	2395	11956
Normal Pneumonia	7208	1802	2253	11263
None Pneumonia	6849	1712	2140	10701

Table 5 Details of four-categorical dataset

Image category	Training	Validation	Test	Amount
COVID-19	1245	120	120	1485
Bacterial Pneumonia	1245	120	120	1485
Viral Pneumonia	1225	120	120	1465
None Pneumonia	1245	120	120	1485

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$F - 1Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{11}$$

where TP is short for True Positive, TN for Ture Negative, FP for False Positive, and FN for False Negative. TP, TN, FP, and FN can be obtained from the confusion matrix, which is a square matrix containing the predictionary label and the ground truth. Additionally, receiver operating characteristic (ROC) curve is another widely used metric to describe the performance of a classifier, where the false-positive rate (FPR) and true-positive rate (TPR) is used as the horizontal and vertical axis, respectively. In statics, true-positive rate (TPR) is defined as the probability of a positive response when the correct answer is positive, and false-positive rate (FPR) is defined as the probability of a positive response when the correct answer is negative [55]:

$$FPR = \frac{FP}{FP + TN} \tag{12}$$

$$TPR = \frac{TP}{TP + FN} \tag{13}$$

4.3 Implementation

In all the tested models, AdamW [56] is used for optimization, which performs better in preventing gradient vanishing during training compared with jointly using Adam and L2 regularization. A mini-batch with a size of 16 is used in both the training and test process, where the maximum training epoch is set as 100. Considering that the distilled spatial features in each channel act as patches in the process of transformer encoder, we trained ResNet18 locally to get optimal parameters of kernels. Data augmentation is applied before training. The detailed parameters, which include basic settings for training and data augmentation,

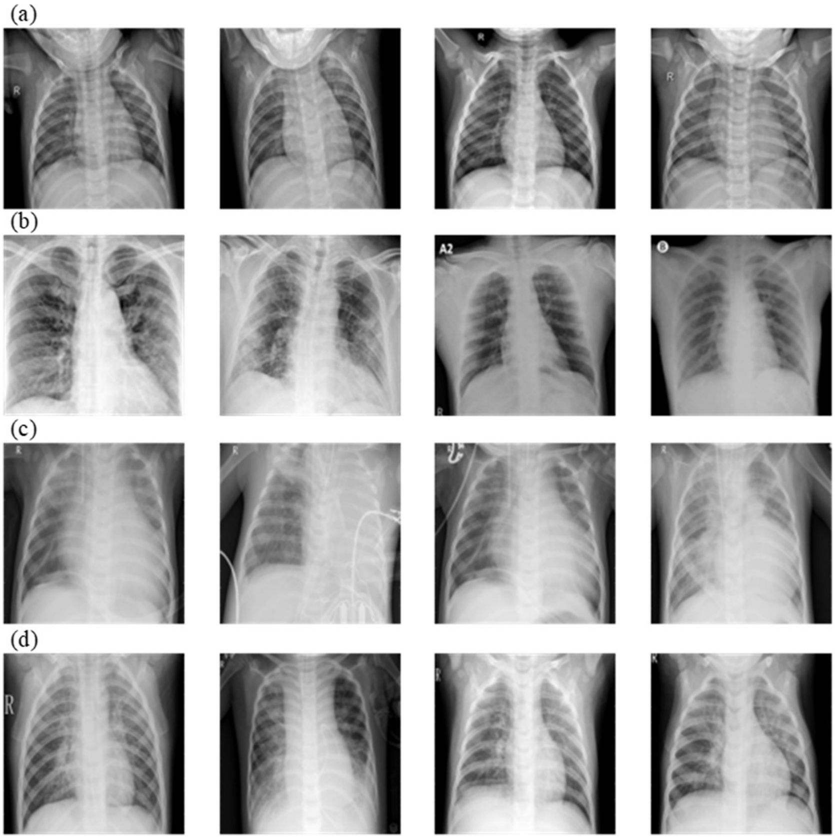


Fig. 5 Typical chest X-ray images from the combined dataset: CXR image of None Pneumonia (a), CXR images of COVID-19 (b), CXR images of Bacterial Pneumonia (c), and CXR images of Viral Pneumonia (d)

are listed in Table 6. We initialize the hyperparameters with recommended values from previous work [16, 38, 42, 56].

5 Results and discussion

All the experiments in this paper were implemented in Python using Keras with TensorFlow running on an NVIDIA GeForce RTX 3090 GPU. The training history of our model is shown in Fig. 6, where cross-entropy and categorical accuracy are used as a criterion to help make early stopping in case of overfitting. The model is trained with 300 epochs. Both validation loss and validation accuracy seem to begin to converge during the training process. There is a significant increase in accuracy values at the beginning of the training within 50 epochs. Whereas training loss presents a significant trend in decreasing, the validation loss does not decrease obviously after 200 epochs. We should early stop the training process before 200 epochs in case of overfitting.

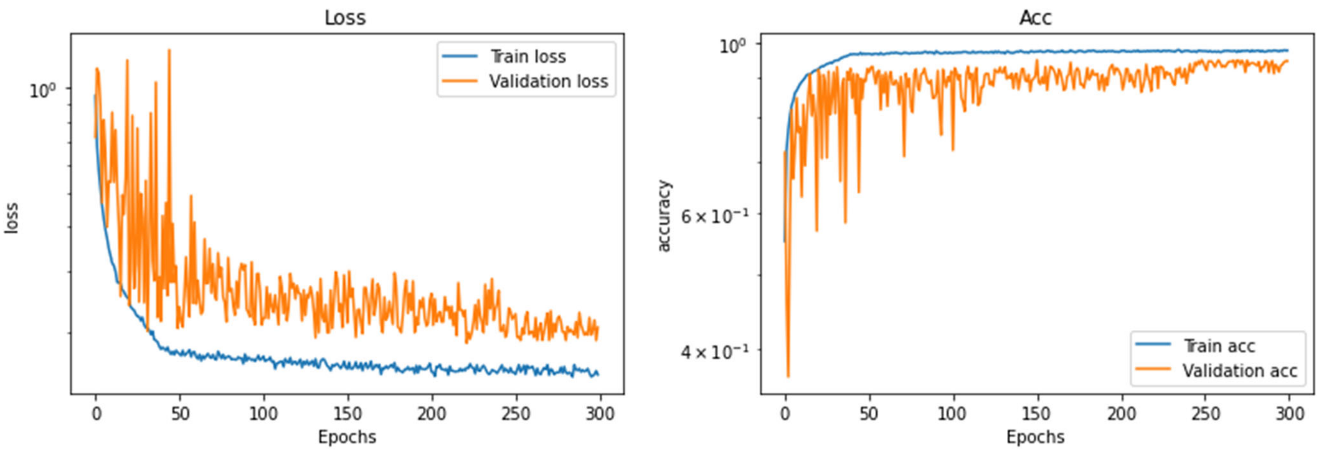
Table 6 Parameters during training and data augmentation

Name of parameter	Value
Image size	224 × 224 × 1
Batch size	16
Optimizer	AdamW
Learning rate	0.0001
Weight decay	0.00001
Dropout rate	0.2
Loss	Categorical cross-entropy
Zoom range	0.1
Rotation range	0.1
Width-shift range	0.1
Height-shift range	0.1

5.1 Model evaluation

The aforementioned statistical metrics are the top metrics used to measure the performance of classification algorithms. Our proposed PneuNet obtained a prediction accuracy of 95.16%, whereas the precision, recall, and F1-Score for class COVID-19 are 96.95%, 98.45%, and 97.69% respectively. The class-wise performance of PneuNet is presented in Table 7, which is generated through the confusion matrix shown in Fig. 7.

We compare the proposed PneuNet with some other deep learning methods from previous work [28, 30, 31, 43] based on the three-category classification using the same dataset. Table 8 shows the comparison between our proposed PneuNet and other deep learning, where the



(a) (b)

Fig. 6 History during training: history of cross-entropy loss (a) and history of categorical accuracy (b)

aforementioned statistical metrics are used. Our proposed PneuNet outperforms other models in all aspects and has a significant increase in overall precision, from 93.55% (Wang et al. [28], COVID-Net) to 97.11%.

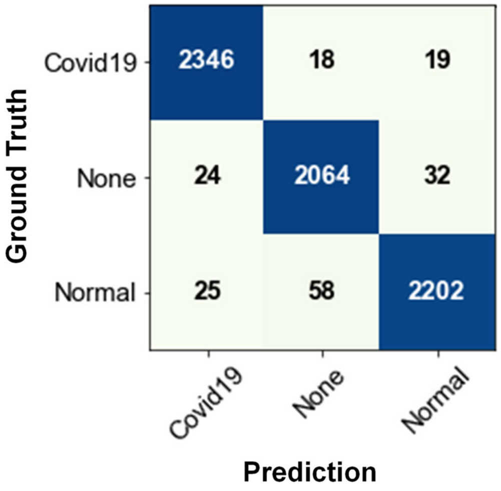
CoroNet [31] and COVID-Net [31] are conventional deep neural networks based on convolution layers and residual connections. However, the model from Siltuala et al. [43] is a VGG-16 model concatenating with traditional VIT modular, and DarkCovidNet [30] takes a heat map of the original CXR images into account inspired by self-attention. Apart from our proposed model, all CNN-based models perform better than transform models on the same dataset. It could be derived that simply cropping the raw image is weak in extracting the latent spatial features, especially when the texture of the lung does not present an apparent difference between that from CXR images of COVID-19 and of other pneumonia. However, latent spatial features in high dimensions could be extracted from CNN methods, allowing the transformer encoder to perform a better prediction.

Table 7 Statistical performance of PneuNet on three-category classification

Class	Precision (%)	Recall (%)	F-1 Score (%)
COVID-19	96.95	98.45	97.69
None Pneumonia	96.64	97.35	96.99
Pneumonia	97.74	96.37	97.10
Average	97.11	97.39	97.26
Prediction Accuracy	95.16%		

5.2 Model performance under other circumstances

We also test the performance of the proposed PneuNet on binary classification and four-category classification to detect its robustness. The binary dataset is generated from the original dataset by deleting CXR images of normal pneumonia from both the training set and validation set, whereas the four-category dataset is generated from the original dataset by relabeling bacterial pneumonia against viral pneumonia from normal pneumonia. The confusion matrices are shown in Fig. 8 and the aforementioned statistical metrics are shown in Table 9.



	Covid19	None	Normal
Covid19	2346	18	19
None	24	2064	32
Normal	25	58	2202

Fig. 7 The confusion matrix for three-category classification: None Pneumonia, Normal Pneumonia, and COVID-19

Table 8 Comparison among proposed PneuNet and other deep learning methods

Model	Precision (%)	Recall (%)	F-1 Score (%)	Accuracy (%)
Siltuala et al. [43] Attention-based VGG-16	85.61	80.10	82.77	81.36
Ozturk et al. [30] DarkCovidNet	89.96	85.35	87.59	87.02
Khan et al. [31] CoroNet	91.85	94.63	93.22	91.30
Wang et al. [28] COVID-Net	93.55	93.33	93.44	93.33
Proposed PneuNet	97.11	97.39	97.26	95.16

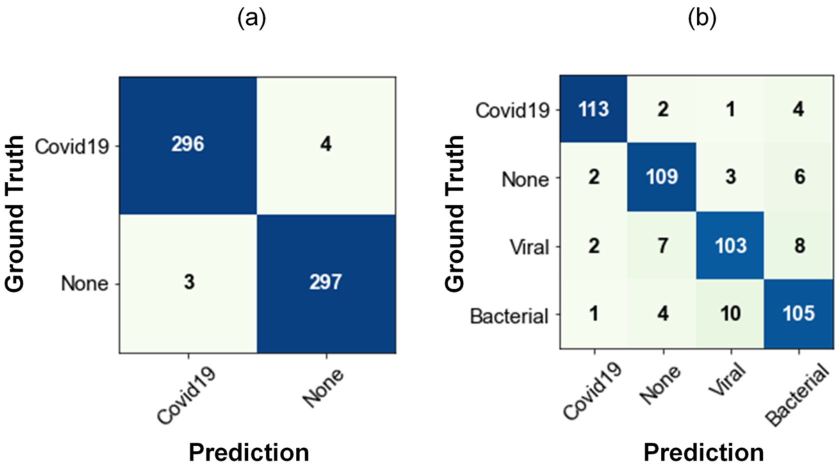


Fig. 8 Confusion matrix generated from binary classification model (a) and four-category classification model (b)

Table 9 Statistical performance of PneuNet

Class	Precision (%)	Recall (%)	F-1 Score (%)
Binary classification			
COVID-19	98.87	99.00	98.93
None Pneumonia	99.00	98.67	98.83
Average	98.94	98.84	98.88
Prediction accuracy	99.32%		
Four-category classification			
COVID-19	94.17	95.76	94.96
None Pneumonia	90.83	89.34	90.07
Bacterial Pneumonia	85.83	88.03	86.92
Viral Pneumonia	87.50	85.37	86.42
Average	89.58	89.62	89.59
Prediction accuracy	90.03%		

Table 10 Comparison of statistical performance among PneuNet and other deep learning models

Model	Precision (%)	Recall (%)	F-1 Score (%)	Accuracy (%)
Binary classification				
Rarin et al. [26] InceptionV3	82.4	100	90.3	97.7
Ozturk et al. [30] DarkCovidNet	98.04	95.13	96.56	98.08
Khan et al. [31] CoroNet	98.33	99.31	98.81	99.01
Proposed PneuNet	98.94	98.84	98.88	99.32
Four-category classification				
Siltuala et al. [43] Attention-based VGG-16	87.48	96.31	86.89	86.15
Ozturk et al. [30] DarkCovidNet	89.96	85.35	87.59	87.02
Khan et al. [31] CoroNet	87.61	87.82	87.71	87.36
Proposed PneuNet	89.58	89.62	89.59	90.03

We compare the prediction performance with other related works [26, 30, 31, 43], as shown in Table 10 for binary classification and four-category classification. Our proposed PneuNet performed better than the above models with a prediction accuracy of 99.32%, where the precision, recall, and F1-Score for class COVID-19 are 98.94%, 98.84%, and 98.88% respectively when detecting COVID-19 out of none pneumonia.

5.3 Future work

PneuNet exhibits good performance compared to other deep learning methods, but there are still several limitations, especially when dealing with multi-category classification problems compared with some CNN-based models such as LDC-NET [57]. This can be caused by a couple of reasons. Firstly, compared with computerized tomography (CT) images, CXR images only contain planar texture from one fixed angle of view, having lost plenty of lung texture such as that on the coronal plane. In the meanwhile, our proposed PneuNet used ResNet18 for feature extraction but is ready to use other deeper CNN encoders, such as ResNet51, to extract latent patterns much deeper in higher dimensions. Future directions thus include augmenting the dataset and applying a deeper CNN encoder dealing with complex input such as CT images, as well as extending the application of our proposed PneuNet, such as predicting how severe the patient is and predicting dates before patients are cured, which could be useful for proper and efficient allocation of medical resources. Finally, like most other deep learning models, our proposed PneuNet is a black box model. The feature extraction process, especially the channel-wise transformer encoder, is nonrepresentational and complicated. So far, we still cannot find a proper method to make better the model interpretable. This is another important direction for our follow-up research.

6 Conclusion

Quick and efficient diagnosis in the early stage is critical during such a tough time caused by the raging plague. Inspired by high diagnostic demand but limited medical resources, we propose a deep learning method named PneuNet, which is based on ResNet18 and applied to extract spatial features, to detect COVID-19 cases from the chest X-ray images. The proposed PneuNet is evaluated on a combined CXR dataset and reached a 95.13% prediction accuracy as well as a 95.16% precision, which is state of the art over other deep learning methods. The model performed well in binary classification and obtained a 99.29% training accuracy as well as a 98.79 precision when distinguishing COVID-19 against none pneumonia. PneuNet also obtained a promising prediction accuracy (86.94%) on four-category classification among COVID-19, none pneumonia, bacterial pneumonia, and viral pneumonia, revealing its latent capacity in the diagnosis of more kinds of pneumonia based on CXR images. The performance of our proposed PneuNet could get improved with the extension of the dataset in the future. From the comparison with other deep learning methods, it is convincing that channel-based attention has great potential in the field of feature recognition and image classification.

Despite the convincing prediction result obtained from PneuNet, the model still needs clinical study and testing but reveals great potential in quick remote diagnosis of suspected COVID-19 patients on a large scale.

Funding This work is supported by the Tsinghua Precision Medicine Foundation under Grant 2022TS009.

Declarations

Conflict of Interest The authors declare no competing interests.

References

1. Rudan I, Boschi-Pinto C, Biloglav Z, Mulholland K, Campbell H (2008) Epidemiology and etiology of childhood pneumonia. Bull World Health Organ 86:408–416B
2. Loo WK, Hasikin K, Suhaimi A, Yee PL, Teo K, Xia K, Qian P, Jiang Y, Zhang Y, Dhanalakshmi S et al (2022) Systematic review on COVID-19 readmission and risk factors: future of machine learning in COVID-19 readmission studies. Front Public Health, 1311
3. Dong E, Du H, Gardner L (2020) An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 20(5):533–534
4. Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, Ji W (2020) Sensitivity of chest CT for COVID-19: comparison to RT-PCR. Radiology 296(2):E115–E117
5. Ng M-Y, Lee EYP, Yang J, Yang F, Li X, Wang H, Lui MM-S, Lo CS-Y, Leung B, Khong P-L et al (2020) Imaging profile of the COVID-19 infection: radiologic findings and literature review. Radiol Cardiothorac Imaging 2(1):e200034
6. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395(10223):497–506
7. Xie X, Zhong Z, Zhao W, Zheng C, Wang F, Liu J (2020) Chest CT for typical coronavirus disease 2019 (COVID-19) pneumonia: relationship to negative RT-PCR testing. Radiology 296(2):E41–E45
8. Salehinejad H, Colak E, Dowdell T, Barfett J, Valaee S (2018) Synthesizing chest X-ray pathology for training deep convolutional neural networks. IEEE Trans Med Imaging 38(5):1197–1206
9. Vineth Ligi S, Kundu SS, Kumar R, Narayanamoorthi R, Lai KW, Dhanalakshmi S (2022) Radiological analysis of COVID-19 using computational intelligence: a broad gauge study. J Healthc Eng, 2022
10. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
11. Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 35(8):1798–1828
12. LeCun Y, Bengio Y et al (1995) Convolutional networks for images, speech, and time series. Handb Brain Theory Neural Netw 3361(10):1995
13. Hellerstein JM, Naughton JF, Pfeffer A (1995) Generalized search trees for database systems
14. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol 1. IEEE, pp 886–893
15. Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? Adv Neural Inf Process Syst, 27
16. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
17. Nayak SR, Nayak DR, Sinha U, Arora V, Pachori RB (2021) Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: a comprehensive study. Biomed Sig Process Control 64(102365):1–12
18. Shwab C, Drn D, Dsg E, Xin ZF, Ydzb G (2021) COVID-19 classification by CCSHNet with deep fusion using transfer learning and discriminant correlation analysis. Inf Fusion 68:131–148
19. Serena Low WC, Chuah JH, Tee CATH, Anis S, Shoaib MA, Faisal A, Khalil A, Lai KW (2021) An overview of deep learning techniques on chest X-ray and CT scan identification of COVID-19. Comput Math Methods Med, 2021
20. Sheykhivand S, Mousavi Z, Mojtahedi S, Rezaii TY, Farzamnia A, Meshgini S, Saad I (2021) Developing an efficient deep neural network for automatic detection of COVID-19 using chest X-ray images. Alex Eng J 60(3):2885–2903
21. Woan Ching SL, Lai KW, Chuah JH, Hasikin K, Khalil A, Qian P, Xia K, Jiang Y, Zhang Y, Dhanalakshmi S (2022) Multiclass convolution neural network for classification of COVID-19 CT images. Comput Intell Neurosci, 2022
22. Bhosale YH, Patnaik KS (2022) Application of deep learning techniques in diagnosis of COVID-19 (coronavirus) A systematic review. Neural Process Lett, 1–53
23. Bhosale YH, Zanwar S, Ahmed Z, Nakrani M, Bhuyar D, Shinde U (2022) Deep convolutional neural network based COVID-19 classification from radiology X-ray images for IoT enabled devices. In: 2022 8th international conference on advanced computing and communication systems (ICACCS), vol 1. IEEE, pp 1398–1402
24. Zhang J, Xie Y, Li Y, Shen C, Xia Y (2020) COVID-19 screening on chest x-ray images using deep learning based anomaly detection. arXiv:2003.12338, 27
25. Hemdan EEl-D, Shouman MA, Karar ME (2020) COVIDx-net: a framework of deep learning classifiers to diagnose COVID-19 in x-ray images. arXiv:2003.11055
26. Narin A, Kaya C, Pamuk Z (2021) Automatic detection of coronavirus disease (COVID-19) using x-ray images and deep convolutional neural networks. Pattern Anal Appl, 1–14
27. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2818–2826
28. Wang L, Lin ZQ, Wong A (2020) COVID-net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest x-ray images. Sci Rep 10(1):1–12
29. Apostolopoulos ID, Mpesiana TA (2020) COVID-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. Phys Eng Sci Med 43(2):635–640
30. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Acharya UR (2020) Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med 121:103792
31. Khan AI, Shah JL, Bhat MM (2020) CoroNet: a deep neural network for detection and diagnosis of COVID-19 from chest X-ray images. Comput Methods Programs Biomed 196:105581
32. Shazia A, Xuan TZ, Chuah JH, Usman J, Qian P, Lai KW (2021) A comparative study of multiple neural network for detection of COVID-19 on chest X-ray. EURASIP J Adv Sig Process 2021(1):1–16

33. Medsker LR, Jain LC (2001) Recurrent neural networks. Des Appl 5:64–67
34. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
35. Jia X, Gavves E, Fernando B, Tuytelaars T (2015) Guiding the long-short term memory model for image caption generation. In: Proceedings of the IEEE international conference on computer vision, pp 2407–2415
36. Udritoiu AL, Cazacu IM, Gruionu LG, Gruionu G, Iacob AV, Burtea DE, Ungureanu BS, Costache MI, Constantin A, Popescu CF (2021) Real-time computer-aided diagnosis of focal pancreatic masses from endoscopic ultrasound imaging based on a hybrid convolutional and long short-term memory neural network model. PLoS ONE, 6
37. Mousavi Z, Shahini N, Sheykhivand S, Mojtahedi S, Arshadi A (2022) COVID-19 detection using chest X-ray images based on a developed deep neural network. SLAS Technol 27(1):63–75
38. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In: Advances in neural information processing systems, pp 5998–6008
39. Devlin J, Chang M-W, Lee K, Toutanova K (2018) BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805
40. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I et al (2019) Language models are unsupervised multitask learners. OpenAI Blog 1(8):9
41. Jaderberg M, Simonyan K, Zisserman A et al (2015) Spatial transformer networks. Adv Neural Inf Process Syst 28:2017–2025
42. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S et al (2020) An image is worth 16x16 words: transformers for image recognition at scale. arXiv:2010.11929
43. Sitaula C, Hossain MB (2021) Attention-based VGG-16 model for COVID-19 chest X-ray image classification. Appl Intell 51(5):2850–2863
44. Park S, Kim G, Oh Y, Seo JB, Lee SM, Kim JH, Moon S, Lim JK, Ye JC (2021) Vision transformer for COVID-19 CXR diagnosis using chest x-ray feature corpus. arXiv:2103.07055
45. Qata-cov19 database. https://www.kaggle.com/aysendegerli/qatacov19-dataset
46. Covid-19-image-repository. https://github.com/ml-workgroup/COVID-19-image-repository/tree/master/png
47. Eurorad. https://www.eurorad.org/
48. COVID-chestxray-dataset. https://github.com/ieee8023/COVID-chestxray-dataset
49. COVID-19 database. https://www.sirm.org/category/senza-categoria/COVID-19/
50. Kaggle (2020) COVID-19 radiography database. https://www.kaggle.com/tawsifurrahman/COVID19-radiography-database
51. Github (2020) COVID-cxnet. https://github.com/armiro/COVID-CXNet
52. RSNA pneumonia detection challenge. https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data
53. Chest x-ray images (pneumonia). https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia
54. Medical imaging databank of the valencia region. padchest: a large chest X-ray image dataset with multi-label annotated reports. https://bimcv.cipf.es/bimcv-projects/padchest/
55. Weller SC (2005) Cultural consensus model. In: Kempf-Leonard K (ed) Encyclopedia of social measurement. Elsevier, New York, pp 579–585
56. Loshchilov I, Hutter F (2017) Decoupled weight decay regularization. arXiv:1711.05101
57. Bhosale YH, Sridhar Patnaik K (2022) IoT deployable lightweight deep learning application for COVID-19 detection with lung diseases using RaspberryPI. In: 2022 international conference on IoT and blockchain technology (ICIBT). IEEE, pp 1–6

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Tianmu Wang received the B.E. degree in mechanical engineering from Nanjing University of Science and Technology in 2019, and received the M.S degree in mechanical engineering from Columbia University in 2021. He is now a Ph.D. candidate in the Department of Mechanical Engineering at Tsinghua University and tracks the field of intelligent medical robots in the Advanced Mechanism and Roboticized Equipment Lab.

Zhenguo Nie received the B.S. degree in materials science and engineering from Shandong University in 2006, and received the M.S. and Ph.D. degree in mechanical engineering from Tsinghua University in 2012 and 2016.

He is currently an Assistant Professor at the Department of Mechanical Engineering, Tsinghua University, Beijing, China. From 2016 to 2018, he was a Postdoctoral Researcher at Georgia Institute of Technology. He was a Postdoctoral Researcher and a lecturer at Carnegie Mellon University from 2018 to 2020. His research interests include AI-based research on mechanical engineering, intelligent medical robots, topology optimizations, and AI-based CAD/CAM.

Ruijing Wang received the B.E. degree in electrical engineering from Hebei University of Technology in 2018, and received the M.S degree in computer engineering from New York University in 2021. She is now a Ph.D. student in the School of Systems & Enterprises at Stevens Institute of Technology. She tracks the field of human-computer interaction in Human-AI Interaction design lab.

Qingfeng Xu received the B.E. degree in Computer Science from Ocean University of China in 2018, and received the M.S degree in Information Technology from the University of Melbourne in 2021. He is working as a Research Assistant in the Department of Mechanical Engineering at Tsinghua University in the Advanced Mechanism and Roboticized Equipment Lab.

Hongshi Huang is now an associate chief physician at Peking University Institute of Sports Medicine. His research interest covers intelligent medical data analysis, diagnosing, sports medicine/rehabilitation, elite athletes sports training and performance enhancement methods, East-meets-West in rehabilitation techniques in sports injuries and prevention programs, biomechanics of sports and lower extremity injury, isokinetic measurement and training, brace and shoe functions.

Handing Xu received the B.E. degree in mechatronic engineering from Beijing Institute of Technology in 2021. He is now a Ph.D. candidate in the Department of Mechanical Engineering at Tsinghua University and tracks the field of intelligent medical robots in the Advanced Mechanism and Roboticized Equipment Lab.

Fugui Xie received the B.S. in mechanical engineering from Tongji University in 2005 and the Ph.D. in manufacturing engineering from Tsinghua University in 2012. He is an Associated Professor in the Department of Mechanical Engineering at Tsinghua University, China. From 2012 to 2014, he was a Postdoctoral Researcher at Tsinghua University. He was the Alexander von Humboldt Research Fellow with Fraunhofer IWU, Germany, from 2015 to 2016. His research interests include theory and design on the issues of mechanisms, parallel kinematics machines, parallel robots, and advanced manufacturing equipments.

Xin-Jun Liu received the B.S. and M.S. degrees in machine design and manufacture and mechanics from Northeast Heavy Machinery Institute, Qinhuangdao, China, in 1994 and 1996, respectively, and the Ph.D. degree in mechanical design and theory from Yanshan University, Qinhuangdao, in 1999.

He is currently a Professor at the Department of Mechanical Engineering, Tsinghua University, Beijing, China. From 2000 to 2001, he was a Postdoctoral Researcher at Tsinghua University. He was a Visiting Researcher with Seoul National University, Seoul, South Korea, from 2002 to 2003. He was the Alexander von Humboldt Research Fellow with the University of Stuttgart, Stuttgart, Germany, from 2004 to 2005. He has authored/co-authored more than 110 papers in refereed journals and refereed conference proceedings. His research interests include parallel mechanisms, parallel robotics, parallel kinematic machines, and motion simulators.

Prof. Liu is currently the Chair of the IFToMM China-Beijing and the Associate Editor for Mechanism and Machine Theory.

Affiliations

Tianmu Wang[1,2,3] · Zhenguo Nie[1,2,3] · Ruijing Wang[4] · Qingfeng Xu[1,5] · Hongshi Huang[6] · Handing Xu[1,2,3] · Fugui Xie[1,2,3] · Xin-Jun Liu[1,2,3]

Tianmu Wang
wtm21@mails.tsinghua.edu.cn

Ruijing Wang
rwang79@stevens.edu

Qingfeng Xu
qingfeng.xu.academic@gmail.com

Hongshi Huang
qhuanghs@bjmu.edu.cn

Handing Xu
xhd21@mails.tsinghua.edu.cn

Fugui Xie
xiefg@tsinghua.edu.cn

Xin-Jun Liu
xinjunliu@tsinghua.edu.cn

[1] Department of Mechanical Engineering, Tsinghua University, Beijing, 100084, China

[2] State Key Laboratory of Tribology in Advanced Equipment, Tsinghua University, Beijing, 100084, China

[3] Beijing Key Lab of Precision/Ultra-precision Manufacturing Equipments and Control, Tsinghua University, Beijing, 100084, China

[4] School of System & Enterprises, Stevens Institute of Technology, Hoboken, NJ, 07030, USA

[5] National Cancer Center, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100060, China

[6] Institute of Sports Medicine, Peking University Third Hospital, Beijing, 100091, China