



A pyramid GNN model for CXR-based COVID-19 classification

Chang Jie¹ · Chen Jiming² · Shao Ying² · Tong Yanchun³ · Ren Haodong⁴

Accepted: 27 August 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

The urgent need for efficient COVID-19 diagnosis has spurred advancements in chest X-ray (CXR) radiography, particularly with the aid of deep learning technologies like convolutional neural networks (CNNs) and graph neural networks (GNNs). Yet, the scarcity of labeled CXR images due to privacy constraints and the complexity of COVID-19 phenotypes often hamper model performance. In this study, we present an innovative pyramid GNN model that effectively tackles these challenges. By segmenting a CXR image into patches, our model leverages a CNN to capture shallow features, then employs a pyramid graph structure within GNN layers to gain the inter-relationship of infected region in distant patches and to amalgamate high-level features. These are subsequently processed by a multi-layer perceptron classifier for final diagnosis. Our approach offers multiple benefits, including noise elimination without the need for pre-treatment, efficient examination of remote infection regions, and the ability to accommodate the intricate structure of the lungs. Evaluations conducted on three distinct public CXR image datasets suggest that our pyramid GNN model offers a promising pathway for enhancing the accuracy and efficiency of COVID-19 diagnosis.

Keywords COVID-19 · Graph neural network · Convolutional neural network · Chest X-ray images

1 Introduction

COVID-19, caused by the SARS-CoV-2 virus, has been a global health concern since 2019. As of June 2022, it has affected over 531 million individuals globally, resulting in over 6.3 million deaths [1]. Existing studies [2] have found that COVID-19 could affect nearly every major organ system. The common symptoms include cough, fever, breathlessness, etc. Reverse transcription-polymerase chain reaction (RT-PCR) testing is the gold standard for identifying patients with COVID-19. The

Extended author information available on the last page of the article

RT-PCR test, however, is time-consuming with high specificity yet low sensibility. In other words, a patient with COVID-19 positive get a negative RT-PCR result.

In clinical scenarios, COVID-19 manifests a wide range of phenotypes across different individuals. Consequently, the medical imaging analysis through chest radiography modalities such as chest X-ray (CXR) and Chest Computerized Tomography (CT) has proven indispensable in the clinical diagnosis, efficacy prediction, and accurate lung function assessment in COVID-19 patients. Advancements in computer vision and machine learning technologies have inevitably drawn researchers' attention toward AI-empowered diagnostic techniques based on chest radiography images. These automatic methods not only enable radiologists to detect infection from chest radiography images more efficiently and accurately, but also significantly alleviate their workload.

While the CT modality offers high-resolution imaging, it comes at a considerable cost. In contrast, the CXR modality, with its affordability, accessibility, and lower radiation dose, stands out as a superior alternative. Notably, portable CXR equipment minimizes the spread of COVID-19 by reducing patient movement to and from radiology facilities [3]. However, the low image resolution of CXR poses challenges for radiologists distinguishing between COVID-19 and other types of pneumonia, given their similar imaging manifestations under the CXR modality. Therefore, the development of effective CXR-based automatic methods holds significant value for the accurate diagnosis of COVID-19 from CXR images.

The involvement of deep learning in the field of computer vision enables the high quality of extracted high-level features to replace those based on conventional manual feature extraction methods in the domain of 'feature engineering'. Based on the classical CNN models, VGG [4], ResNet [5], Inception [6], CNN-based classifiers [7–12] are raised gradually to aid the COVID-19 diagnosis based on CXR images. The CXR-based COVID-19 classification task faces several challenge.

1. First, the limited availability of labeled medical image samples is not sufficient to inspire a good generalization performance of the deep learning models, which have a large amount of parameters to be trained.
2. Second, recently researchers tend to use a model pretrained on the ImageNet database to solve the problem of medical images by fine-tuning parameters of the model. The different image-forming principles between natural visual images and medical images induce a so-called negative transfer, during the transfer learning from source domain (natural visual images) to target domain (medical images), which reversely undermines the ability of the fine-tuned model in the target domain.
3. Third, the characteristics of COVID-19 images are generally in diverse manifestations and in different locations of pulmonary region. Even an experienced radiologist makes a conclusion diagnosis of COVID-19 based on combining global and local representation of CXR images.

To tackle these challenges, Chola et al. [13] proposed a hybrid deep learning approach for COVID-19 diagnosis using CT and X-ray images. While their model

showed promising results, it relies on the combination of different imaging modalities, which may not always be available in all clinical settings. Two ensemble models [14, 15] show effectivity in handling the diversity of data for pneumonia identification from chest X-ray images. But ensembling several CNN-based components adds to the model's complexity and computational demand.

In light of these limitations, our research is motivated by the need for a model that can efficiently and accurately diagnose COVID-19 using chest X-ray images alone, reducing reliance on other imaging modalities or complex computational methods. Moreover, we aim to design a model that maintains its performance across diverse datasets, increasing its applicability in real-world clinical settings. Our proposed approach tackles these issues by employing a Graph Neural Network (GNN)-based model that simplifies the feature extraction process and enhances the diagnosis accuracy. The contributions of our work are as follows:

1. A CXR image is split into a number of patches, which can remove background noises without the pre-treatment of lung segmentation.
2. The pulmonary region can be viewed as a composition of patches, and graph structure can obtain the inter-relationship of infected region in distant patches;
3. A graph structure is more flexible than grid or sequence to model the complex object as an object in the image is usually not quadrate whose shape is irregular.

The rest of this paper is organized as follows: Section 2 introduces some recent works which applied deep learning technologies to the domain of medical imaging. Section 3 elaborates the proposed model in details. Section 4 analyzes the results based on comparison to the other four classic models.

2 Related Work

The emergence of deep learning-based methods, beginning with LeNet [16], and their ability to extract high-level image features [4, 5, 16–20] has revolutionized various visual tasks such as image classification [17], object detection [20], and semantic segmentation [21]. Predominantly, Convolutional Neural Networks (CNNs)—encompassing ResNet [5], MobileNet [18], and EfficientNet [19]—have significantly evolved in the past decade. In the context of medical image processing, most deep learning techniques leverage large datasets with pre-existing labels (e.g., ImageNet) from diverse domains for model pretraining, or employ various data augmentation strategies to circumvent overfitting due to the scarcity of labeled medical image samples. These models excel in high-level feature extraction, a crucial aspect of image classification, object detection, and semantic segmentation tasks. However, they often require large labeled datasets or significant data augmentation to circumvent overfitting, particularly in medical image processing where labeled samples are typically scarce.

Various models have been proposed to categorize patient chest X-ray (CXR) images. Hemdan et al. [22] developed COVIDX-Net, an ensemble model

utilizing several pretrained models such as VGG [4] and MobileNet [18]. Nigam et al. [23] experimented with several pretrained CNN models for a three-class classification task, achieving different levels of accuracy with each model. Ozturk et al. [24] introduced a classifier called DarkNet within the You Only Look Once (YOLO) [25] object detection system. In contrast, Chouhan et al. [26] applied traditional data enhancement methods to enlarge the dataset.

Building upon the Auxiliary Classifier Generative Adversarial Network (ACGAN) [27], Waheed et al. [28] developed the CovidGAN to generate additional synthetic CXR images, improving CNN-based classifier performance. Similarly, Haritha et al. [29] and Apostolopoulos et al. [30] applied pretrained models to classify CXR images with high accuracy. More innovative models were proposed by Narin et al. [7], Tripti et al. [8], Fitriasaki et al. [10], Karaci et al. [11], Jyoti et al. [31], Chamseddine et al. [32], and Barshooi et al. [33].

The promotion of lightweight COVID-19 classification models has been facilitated by research like Wang et al. [34], Afshar et al. [35], and Montalbo et al. [36]. These models offer more representation capacity and less computational complexity. Federated learning, a burgeoning research field, offers data privacy protection suitable for medical image processing. Models proposed by Feki et al. [37] and Bhattacharya et al. [38] highlight the viability of federated learning for diagnosing COVID-19 from CXR images without sharing private databases. Despite these advances, a common limitation is the inability to fully exploit local and global image details concurrently, which our model aims to address.

The introduction of the transformer architecture to visual tasks [39–42] has led to numerous transformer-based classifiers [43–46] being proposed for distinguishing COVID-19 patients from a large volume of medical images. Nonetheless, these classifiers often face challenges related to high computational load and overfitting due to limited training samples. Our proposed Pyramid GNN model addresses this gap by leveraging an innovative pooling strategy to reduce computational complexity, and a novel patch-wise feature extraction technique to prevent overfitting.

Graph Neural Networks (GNNs), structural learning tools that capture relationship information of each node's neighborhood, have garnered increasing attention [47–50]. They have been widely used for COVID-19 related applications such as drug discovery [51, 52] and forecasting [53–55]. While Song et al. [56] have proposed a notable COVID-19 diagnosis model utilizing multi-center Graph Convolutional Networks (GCN) on Chest CT scans, and Lu et al. [57] have introduced a neighboring aware GNN (NAGNN) focusing on the neighboring relationships between image-level representations (ILR) based on a deep convolutional neural network, these models primarily depend on Chest CT scans. Additionally, their fine granularity scale often limits their ability to capture the varied manifestations of COVID-19 images. In contrast, our Pyramid GNN model has been strategically designed to leverage these associations, significantly enhancing its ability to detect intricate patterns in CXR images.

3 Method

The subtle variances in imaging features between COVID-19 and other types of pneumonia can be challenging to identify, especially when small lesion areas are dispersed throughout the pulmonary cavity. To excavate the relationships among different regions within chest X-ray (CXR) images, we propose a cascaded pyramid Graph Neural Network (GNN) model tailored to address the task of COVID-19 classification. The model's architecture is illustrated in Fig. 1 and consists of three principal components.

The initial component of the architecture comprises a Convolutional Neural Network (CNN) feature extractor, positioned at the inception of the model. This extractor first divides each CXR image into a predefined number of patches, subsequently extracting the rudimentary features from each patch. The second component of the architecture, constructed of pyramid GNN layers, interprets these patches as nodes. The graph structure facilitates the transfer of information between neighboring nodes that share a similar semantic context. The features from these nodes, after undergoing processing, are then fused and fed into a Multilayer Perceptron (MLP) classifier. The MLP classifier, which is the final component of our proposed model, receives the consolidated features from the pyramid GNN layers and performs the final classification task.

In the following sections, we will provide a comprehensive breakdown of these three integral components of our model.

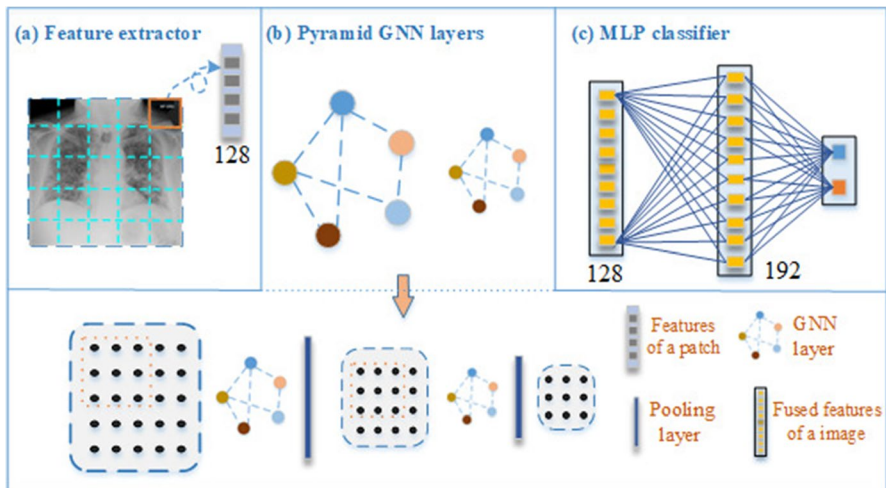


Fig. 1 The architecture of the proposed model consists of three components. The first is a CNN extractor at the head of the architecture, which obtains a shallow features of each patch in 16×16 size. The middle component is pyramid multi-layer GNNs, enhancing the representations features of a patch by fusing information from its neighborhood. A MLP component locates the tail of the architecture, which classifies the inputted image into a predicted category

3.1 Feature extractor

Let's consider a given CXR image I with dimensions $H \times W \times C$, where H , W , and C represent height, width, and channels, respectively. As depicted in Fig. 1a, we initially partition I into a patch set $P = p_1, p_2, \dots, p_{N^{(1)}}$, composed of $N^{(1)}$ patches, before the operation of the feature extractor F . Following [40], this model exploits patches of size 16×16 pixels from the image data, which allow for more granular feature extraction, preserving both local and global details in the images. Subsequently, the feature extractor F processes each patch p_i to convert it into a feature vector $x_i^{(1)} = F(p_i)$, which has a dimension of $C^{(1)}$. As a result, a feature set derived from the patch set P forms $X^{(1)} = x_1^{(1)}, x_2^{(1)}, \dots, x_{N^{(1)}}^{(1)}$. In our experiment, we employ the first five layers of the Resnet-50 architecture to function as the feature extractor F .

3.2 Pyramid GNN

The Pyramid Graph Neural Network (GNN) forms a key component of our model, enabling the transformation and fusion of patch features in a structured manner. This GNN is symbolized as $G = (P, E)$, where the node set, P , consists of the patches we've extracted from the input image, and E defines the edge set. Each element e_{ij} in the edge set represents the relationship between two nodes (patches) p_i and p_j . Initially, every node in the network is associated with a feature vector, forming the feature representation $X^{(1)} = x_1^{(1)}, x_2^{(1)}, \dots, x_{N^{(1)}}^{(1)}$ of the node set.

The structure of the pyramid GNN layer is delineated in Fig. 1b. Each layer of the GNN is designed to foster inter-node information exchange. Specifically, for every node $x_i^{(l)}$, the GNN layer identifies its K nearest neighbors $\mathcal{N}_K(x_i^{(l)})$ based on the weights of the edges $e^{(l)}_{ji}$ linking $x_i^{(l)}$ and all other nodes $x_j^{(l)} \in \text{Neighbor}(x_i^{(l)})$. Once the neighboring nodes have been identified, the GNN layer performs a graph convolution operation, which amalgamates the features of the node's neighbors.

Moving forward from a specific layer to the next, the l -th GNN layer initiates its operation from the feature maps in the preceding layer, $X^{(l)} \in \mathbb{R}^{N^{(l)} \times C^{(l)}}$. The operations performed by the l -th graph convolution layer can be expressed as:

$$X^{(l')} = \sigma(H_{\text{agg}}(X^{(l)})W^{(l)}) \quad (1)$$

In this equation, H_{agg} symbolizes an aggregation operation that combines the features of a node's neighboring nodes to compute a comprehensive representation for that node. We chose the element-addition operation for H_{agg} to limit computational cost. $W^{(l)}$ are the learnable weights in the l -th layer that transform the aggregated feature's dimension in a feature space. σ stands for a nonlinear activation function that introduces nonlinearity into the model, enhancing its capability to capture complex patterns. This process is detailed as follows:

$$x_i^{(l')} = \sigma(H_{\text{agg}}(x_i^{(l)}, \mathcal{N}_K(x_i^{(l)}))W^{(l)}) \quad (2)$$

In this process, $\mathcal{N}_{K^{(l)}}(x_i^{(l)})$ signifies the set of K nearest neighbors of $x_i^{(l)}$. We set $[K^{(1)}, K^{(2)}, K^{(3)}]$ as $[10, 8, 4]$ to configure the three GNN layers in our model.

Furthermore, as shown in Fig. 1b, a pooling layer is nestled between each pair of consecutive GNN layers. The purpose of this pooling layer is to shrink the size of the feature maps, reducing the computational load for subsequent operations. The correlation between the feature maps of the l -th and the $(l + 1)$ -th layers can be formalized as:

$$X^{(l+1)} = H_{\text{pool}}(X^{(l)}) \quad (3)$$

For this pooling operation, H_{pool} , we apply a 2×2 max pooling operation in our experiments. This operation selects the maximum value from a 2×2 window in the feature map, effectively reducing its size while preserving the most salient features.

3.3 MLP-based classifier

The model culminates in a three-layer MLP (Multilayer Perceptron)-based classifier M , serving as the final decision maker for the system. As illustrated in Fig. 1c, the classifier processes the output feature maps from the L -th GNN layers, transforming them into a flattened format that forms its input layer.

The second layer of M encompasses 192 hidden parameters. These parameters are the bridge between the input layer and the final layer, passing on processed information through a fully connected network. This information then reaches the third and final layer, which contains C output nodes. Each output node in this layer represents one of the C categories in the classification tasks, essentially indicating possible classifications for the input CXR image.

The output o of the classifier, representing the class prediction, is then calculated by applying the MLP M to the final GNN layer output $X^{(L)}$:

$$o = M(X^{(L)}) \quad (4)$$

3.4 The application of the proposed model

Algorithm 1 provides the pseudo-code for our Pyramid GNN model. We adopt Adam optimizer [58] to fine-tune parameters in a multi-class cross-entropy loss function defined by Eq. (5).

$$\text{Loss}(O, Y) = -\frac{1}{N} \sum_j^N \sum_c^M y_{jc} \log(o_{jc}) \quad (5)$$

wherein O is the outputted prediction of N training samples while Y is the set of their labels. M is the number of category. o_{jc} represents the probability that the j -th training sample belongs to c -th class. y_{jc} is a indicator function, which is assigned to 1 if the j -th training sample belongs to c -th class; otherwise 0.

Algorithm 1 One step training algorithm of the proposed model for CXR-based COVID-19 Classification

```

1: procedure
2:   Divide image  $I$  into  $N^{(1)}$  patch set  $P = \{p_1, p_2, \dots, p_{N^{(1)}}\}$ .
3:    $X^{(1)} = F(P)$  ▷ Node embedding
4:   Construct a graph  $G = (X^{(1)}, E)$ .
5:   for  $l = 1$  to  $3$  do ▷ Propagation steps
6:     for each node  $x_i^{(l)} \in X^{(l)}$  do
7:       Find  $K$  nearest neighbors of  $x_i^{(l)}$ , denoted as  $N_{K^{(l)}}(x_i^{(l)})$ 
8:        $x_i^{(l')} = \sigma(H_{agg}(x_i^{(l)}, \mathcal{N}_{K^{(l)}}(x_i^{(l)}))W^{(l)})$  ▷ Aggregate neighbors
9:     end for
10:     $X^{(l+1)} = H_{pool}(X^{(l')})$  ▷ Apply pooling operation to reduce feature maps
11:  end for
12:  Flatten the output from the last GNN layer and feed into a MLP-based classifier  $M$  to obtain the prediction  $o = M(X^{(L)})$ 
13:  Compute loss using multi-class cross-entropy loss function and perform parameter update
14: end procedure

```

4 Experiments

4.1 Databases

Due to the recency of the COVID-19 pandemic, publicly available X-ray image datasets are comparatively limited in size. For our experiments, we selected three commonly used datasets.

- X-ray COVID19 dataset [34]: This database is comprised of a total of 188 CXR images in two categories (COVID-19 and normal), where 148 images are in training set and 40 images in testing set.
- CoronaHack chest X-ray dataset [59]: This database contains 5911 images of cases with 4 classes (Normal, bacterial pneumonia, viral pneumonia and COVID-19). In our experiments, we merged the bacterial pneumonia and the viral pneumonia into a single category named pneumonia, converting the four-class classification into a three-class classification task with 3 classes (normal, pneumonia and COVID-19). It was divided into 5309 training samples and 603 testing samples.
- COVID-19 radiography dataset [60]: The database is a collaborative efforts by various universities in Asia. It consists of 3616 COVID-19 positive images, 10,192 normal images, 6012 Lung Opacity (non-COVID lung infection) and

Table 1 Characteristics of the three utilized CXR open-source datasets

| Datasets | Categories | Size | Train size | Test size |
|------------------|------------|--------|------------|-----------|
| Chest X-ray [59] | 3 | 5911 | 5309 | 603 |
| Radiography [60] | 4 | 21,165 | 16,932 | 4233 |
| X-ray [34] | 2 | 188 | 148 | 40 |

Table 2 Hyperparameters for the models used on all the datasets

| Model | Batchsize | Epochs | Optimizer | Activation function |
|--------------------|-----------|--------|-----------|---------------------|
| The proposed model | 32 | 200 | Adam | Leaky ReLU |
| Resnet-50 | 32 | 200 | Adam | ReLU |
| EfficientNet | 16 | 200 | Adam | Swish |
| VGG-16 | 32 | 200 | Adam | ReLU |
| Resnet-101 | 32 | 200 | Adam | ReLU |

1345 viral pneumonia images. In this dataset, 16,932 were used for training and 4233 for testing.

In addition, we aggregated images from all three databases and classified them into two categories: COVID-19 and non-COVID-19. The distribution of training and testing samples was kept consistent throughout the process. This has resulted in a new dataset that we referred to as Combined-X-ray. We utilized this dataset to conduct an additional accuracy test.

For convenience, we'll refer to the CoronaHack chest X-ray dataset, COVID-19 radiography dataset, and X-ray COVID19 dataset as Chest X-ray, Radiography, and X-ray, respectively. Table 1 provides an overview of the total number of training and testing samples in each dataset. As can be seen, these open-source datasets typically struggle with a limited scale of data, a common issue that we seek to address in our experiment.

4.2 Settings

In addition to the proposed model, we also evaluate the performance of four existing pretraining deep learning models: Resnet-50, Resnet-101, VGG-16, and EfficientNet on the aforementioned three datasets. All models' implementations are based on the PyTorch framework, and the computations are conducted on four Nvidia RTX3090 GPUs.

The input size of all models is resized to 224×224 pixels. We employ the Adam optimizer [58] to fine-tune the model parameters by minimizing a multi-class cross-entropy loss function. During the training phase for all models, the initial learning rates are uniformly set to 0.005, and then, they are scaled down by a factor of 0.1 every 10 epochs. The other hyperparameters of all models are listed in Table 2.

5 Results

For a thorough quantitative evaluation of the proposed model, alongside the four existing models, we assess each model based on multiple widely recognized metrics. These metrics include accuracy, precision, sensitivity, F1-score, the Receiver Operating Characteristic (ROC) curve, and the Area Under the Curve (AUC).

The proposed model's effectiveness was assessed using the aforementioned three datasets. Table 3 lists the accuracy of the models on the three test sets. From this table, it is apparent that despite having the smallest sample size, the X-ray dataset offers the highest accuracy for COVID-19 classification compared to the other two datasets. This can be attributed to the fact that the X-ray dataset only involves a relatively straightforward binary classification task. Additionally, the chest X-ray images in the X-ray dataset have been carefully selected to ensure high quality and minimal background noise. Interestingly, even though the Combined-X-ray dataset has the largest number of training samples, all models' performances on this dataset are the least optimal compared to other datasets. The primary reason for this observation is the high diversity within the non-COVID-19 class in the Combined-X-ray dataset, which introduces more complex patterns to be learned by the models.

The Chest X-ray and Radiography datasets, on the other hand, present two inherent challenges: inconsistent image quality and unbalanced class distribution. Table 3 also reveals that the proposed model outperforms the other models in terms of prediction accuracy across all datasets. It is noteworthy that all models exhibit relatively poorer performance on the Radiography dataset compared to the other two datasets. This can be traced back to two primary factors. Firstly, we consolidated the bacterial pneumonia and viral pneumonia categories into a single pneumonia category, which naturally incorporates a broader range of features. Secondly, the number of COVID-19 cases is insufficient when compared to the pneumonia and normal categories.

Tables 4, 5, and 6 present the precision and sensitivity of each class (COVID-19 and normal) for all five models across the three datasets. In Table 4, the proposed model and the Resnet-50 model rank the highest in the X-ray dataset, a finding consistent with the results in Table 3.

In Table 5, we calculate the individual precision and sensitivity for each class (normal, pneumonia, and COVID-19) within the chest X-ray dataset. Here, the proposed model significantly outperforms the others overall. Interestingly, while the

Table 3 The accuracy (all values in percentages) of models on all forementioned datasets for the classification task

| Model | X-ray | Chest X-ray | Radiography | Combined-X-ray |
|--------------------|---------------|-------------|-------------|----------------|
| The proposed model | 100.00 | 88.5 | 95.0 | 76.4 |
| Resnet-50 | 100.00 | 85.7 | 94.6 | 73.2 |
| EfficientNet | 95.0 | 79.7 | 81.5 | 63.8 |
| VGG-16 | 97.5 | 81.4 | 90.5 | 69.6 |
| Resnet-101 | 97.5 | 85.4 | 91.1 | 73.4 |

Bold represent the best performance achieved on the respective database

Table 4 Precision and sensitivity performance comparison (all values in percentages) of various models for each class on the X-ray dataset

| Model | Normal | | COVID-19 | |
|--------------------|---------------|---------------|---------------|---------------|
| | Prec | Sens | Prec | Sens |
| The proposed model | 100.00 | 100.00 | 100.00 | 100.00 |
| Resnet-50 | 100.00 | 100.00 | 100.00 | 100.00 |
| EfficientNet | 95.0 | 95.0 | 95.0 | 95.0 |
| VGG-16 | 100.00 | 95.0 | 95.0 | 100.00 |
| Resnet-101 | 100.00 | 95.0 | 95.2 | 100.00 |

Prec precision, *Sens* sensitivity

Bold represent the best performance achieved on the respective database

Table 5 Precision and sensitivity performance comparison (all values in percentages) of various models for each class on the chest X-ray dataset

| Model | Normal | | COVID-19 | | Pneumonia | |
|--------------------|------------|-------------|-------------|-------------|-------------|-------------|
| | Prec | Sens | Prec | Sens | Prec | Sens |
| The proposed model | 98.3 | 74.9 | 73.4 | 73.9 | 79.9 | 98.0 |
| Resnet-50 | 94.6 | 44.6 | 62.9 | 60.4 | 63.2 | 97.1 |
| EfficientNet | 86.8 | 47.6 | 54.9 | 44.9 | 59.8 | 92.5 |
| VGG-16 | 100 | 34.0 | 44.6 | 81.8 | 81.7 | 92.1 |
| Resnet-101 | 92.5 | 53.1 | 57.4 | 75.1 | 75.7 | 92.5 |

Prec precision, *Sens* sensitivity

Bold represent the best performance achieved on the respective database

Table 6 Precision and sensitivity performance comparison (all values in percentages) of various models for each class on the Radiography dataset

| Category | COVID-19 | | Normal | | Viral pneumonia | | Bacterial pneumonia | |
|--------------------|-------------|-------------|-------------|-------------|-----------------|-------------|---------------------|--------------|
| | Prec | Sens | Prec | Sens | Prec | Sens | Prec | Sens |
| The proposed model | 99.8 | 98.7 | 91.8 | 98.3 | 99.2 | 97.7 | 96.8 | 85.9 |
| Resnet-50 | 98.6 | 96.0 | 93.8 | 95.5 | 96.5 | 97.6 | 92.4 | 90.07 |
| EfficientNet | 82.5 | 82.9 | 81.5 | 85.0 | 83.8 | 82.5 | 78.0 | 72.2 |
| VGG-16 | 93.3 | 93.5 | 90.4 | 93.0 | 95.7 | 95.3 | 87.6 | 83.1 |
| Resnet-101 | 96.2 | 93.5 | 90.0 | 93.7 | 95.2 | 94.1 | 88.5 | 83.9 |

Prec precision, *Sens* sensitivity

Bold represent the best performance achieved on the respective database

VGG-16 model achieves the highest precision for the normal class, it also records the worst sensitivity for this class. This can be attributed to the fact that the normal class is the least represented. The imbalanced distribution of the three categories may cause a preference bias in the models.

Table 6 reveals that in the Radiography dataset, the proposed model achieves the highest precision in three classes (COVID-19, viral pneumonia, and bacterial pneumonia) but falls slightly short in the normal class, where the Resnet-50 model has a slight edge. The table also shows the sensitivity of the five models across the four classes. The proposed model demonstrates superior performance in three classes (COVID-19, normal, and viral pneumonia) when compared with the other four traditional models. It is worth noting that in comparison to the other four classic CNN classifiers, the proposed model performs significantly better in lesion categories such as COVID-19, viral pneumonia, and bacterial pneumonia. This shows our model's capability in accurately distinguishing COVID-19 from other respiratory conditions, which is crucial for precise diagnosis and subsequent treatment. The reason is that the characteristics of chest lesions vary greatly and are found in different locations within the pulmonary region, making a graph structure more effective than a grid or sequence for correlating these dispersed chest lesions.

Figure 2 shows confusion matrices of all evaluated models. It is clear that our proposed model consistently demonstrated superior performance across all three datasets. This is indicated by the higher counts of true positives and true negatives, coupled with a lower number of false positives and false negatives when compared to other models. The stronger performance of our model is especially pronounced on the Chest X-ray and Radiography datasets, where the disease manifestations are more complex and varied.

The ROC curves of the models are displayed in Fig. 3. The AUC values of the models are listed in Table 7, where the proposed model outperforms others on all three datasets.

As shown in Fig. 4, the convergence trends of the training loss across all models evaluated on the three datasets suggest a relatively smooth and rapid decline toward a minimal loss value. Compared to other baselines, our proposed model consistently reached a lower loss value faster than the other models. This suggests that our model was able to grasp the essential characteristics of the COVID-19 cases from the CXR images more effectively.

To further justify the design choices in our model, we conducted an ablation study that isolates the impact of each component. The first variant, A_1 , modifies the feature extractor of our proposed model. Instead of utilizing the first five layers of Resnet-50, it employs a four-layer CNN architecture with 3×3 kernels as the core of the feature extractor. The second variant, A_2 , substitutes the Pyramid GNN component of our model with the backbone of the Swin Transformer [42]. Similarly, A_3 replaces the Pyramid GNN with three-layer GraphSAGE [62], treating the extracted features of patches as nodes in a comparable manner to the Pyramid GNN. Finally, A_4 deviates from our MLP-based classifier and instead employs an SVM for the classification task.

Table 8 presents the average F1-scores for each of the four ablation models, obtained across the three datasets. It is evident that each modification in the model's structure leads to a decrease in performance, underscoring the importance of each component in the overall configuration. Model A_1 , which modifies the feature extractor, demonstrates a drop on the Chest X-ray and Radiography, illustrating the crucial role of Resnet-50 in efficient feature extraction from the complicated CXR images.

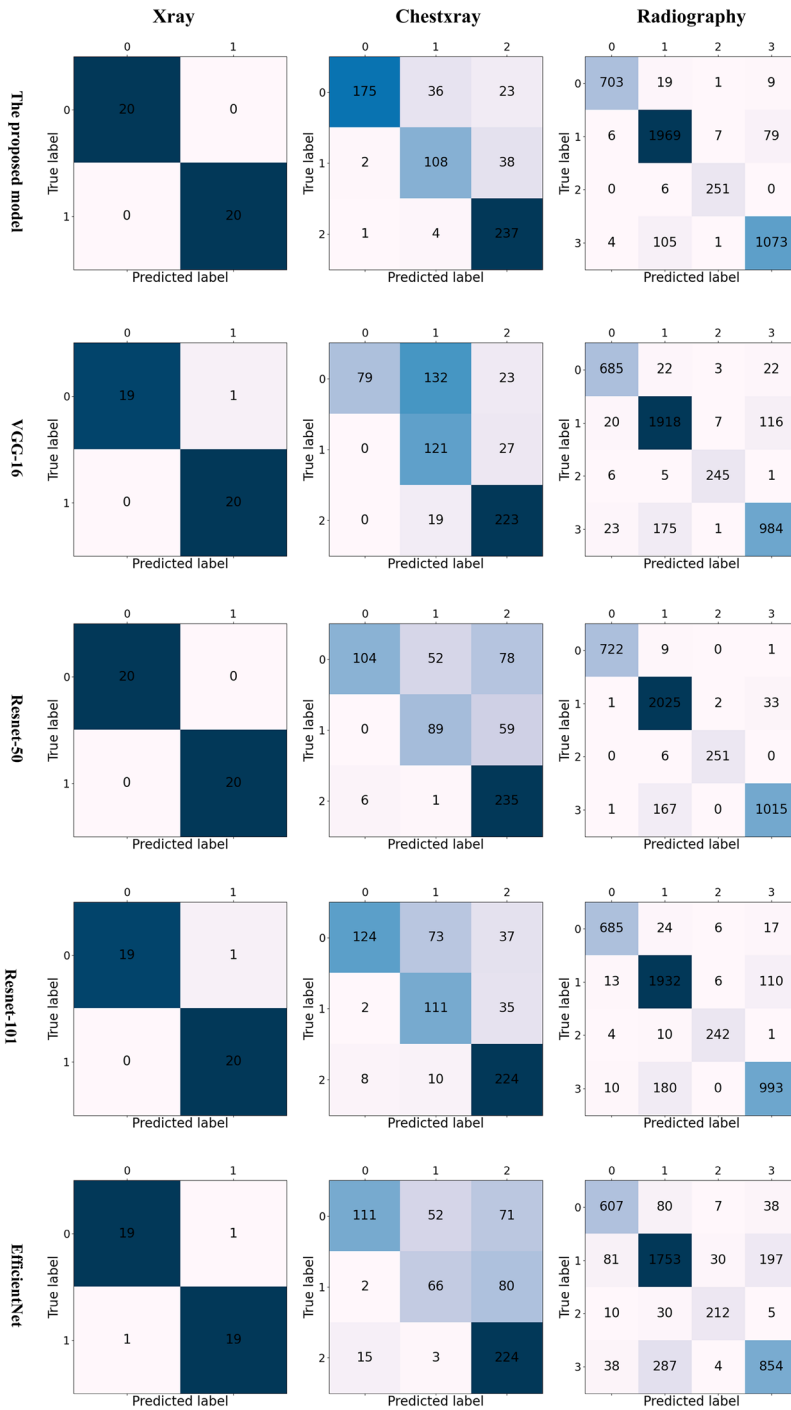


Fig. 2 Confusion matrices of all evaluated models, presented separately for each dataset. From left to right, the columns represent the results based on the X-ray, Chest x=X-ray, and Radiography datasets, respectively

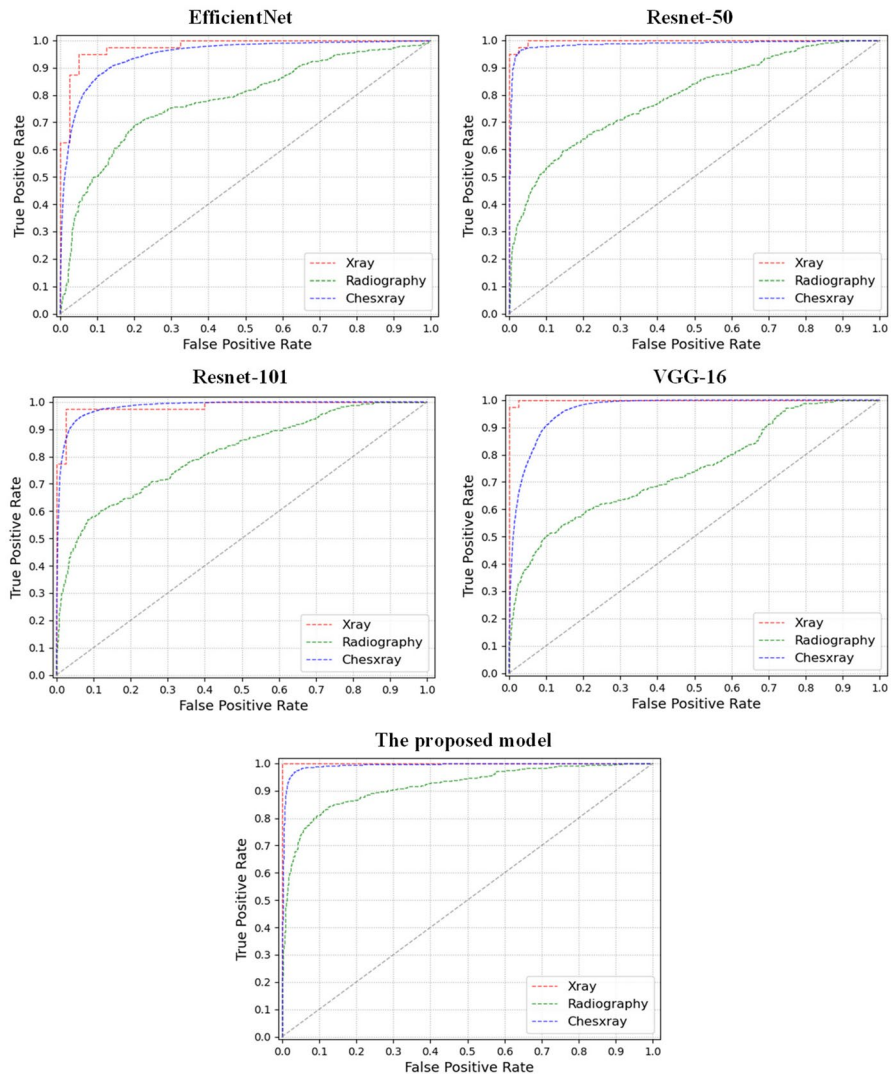


Fig. 3 Receiver Operating Characteristic (ROC) curves for all evaluated models across the three datasets. Each model's results are plotted separately for clear comparison and analysis

Similarly, A_2 and A_3 , which replace the Pyramid GNN with Swin Transformer and GraphSAGE, respectively, also experience a decrease in all performances, highlighting the effectiveness of the Pyramid GNN for dealing with feature aggregation of infected region in distant patches. Finally, model A_4 substitutes the MLP-based classifier with an SVM, and this change also results in a lower F1-score, emphasizing the suitability of MLP for the classification task at hand.

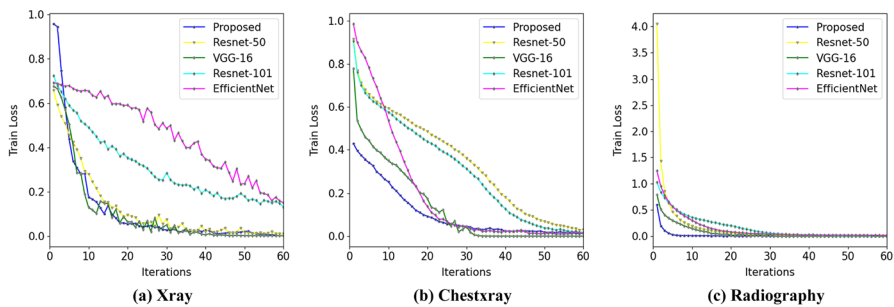


Fig. 4 The convergence trends of train loss for all models evaluated across the three datasets

Table 7 Area Under the Curve (AUC) performance comparison (all values in percentages) for all models on the three datasets in the classification task

| Model | The proposed model | VGG-16 | Resnet-50 | Resnet-101 | EfficientNet |
|------------------|--------------------|--------|-----------|------------|--------------|
| X-ray [59] | 100.00 | 99.9 | 99.8 | 98.5 | 97.8 |
| Chest X-ray [60] | 91.8 | 74.7 | 79.3 | 81.2 | 78.4 |
| Radiography [34] | 99.2 | 96.6 | 98.7 | 98.3 | 94.7 |

Bold represent the best performance achieved on the respective database

Table 8 The average F1-score (all values in percentages) of ablation models on the three datasets

| Model | X-ray | Chest X-ray | Radiography |
|--------------------|---------------|-------------|-------------|
| The proposed model | 100.00 | 82.2 | 95.9 |
| A ₁ | 100.0 | 79.5 | 94.6 |
| A ₂ | 99.5 | 76.8 | 92.4 |
| A ₃ | 99.5 | 77.3 | 91.6 |
| A ₄ | 99.0 | 78.1 | 92.8 |

Bold represent the best performance achieved on the respective database

6 Conclusion

In this study, we proposed a pyramid GNN model that leverages a graph structure to fuse extracted features, aiming to enhance the performance of classification tasks. Our extensive experimental results have demonstrated that the proposed model delivers promising results in terms of accuracy, sensitivity, and precision for individual classes in CXR-based COVID-19 classification tasks.

Moving forward, our research will concentrate on two primary objectives. Firstly, we aim to explore how the benefits of our proposed model can be utilized to enhance the performance of COVID-19 segmentation models. This would involve adapting

our model to handle more complex tasks, such as delineating the specific regions affected by COVID-19 in CXR images. Secondly, we are interested in further reducing the computational cost associated with the proposed model. Although our pyramid GNN model has shown great promise in the classification tasks, it also involves dense computations due to the nature of GNN operations. Thus, improving the computational efficiency without sacrificing the performance would be a crucial step toward making this model more practical for real-world applications. Through these future directions, we hope to continue contributing to the development of efficient and effective AI tools that can aid in the fight against COVID-19.

Author contributions JC and HR wrote the main manuscript text; JC and YS prepared the dataset and pre-processed the CXR images; JC and YT designed the proposed model. All authors reviewed the manuscript.

Funding This work was supported by the Natural Science Foundation of Anhui Province [Grant number 2108085MF205], the Project of School-enterprise Cooperative Practice Education Base [Grant Number 2022xqhzsjjd01], the Anhui Provincial Humanities and Social Science Foundation of China [Grant Numbers SK2020A0380, SK20210466, SK2021A0468], the Key Project of Anhui University Outstanding Young Talents Support Plan [Grant Numbers gxyqZD2016180]; Wannan Medical College teaching Quality and Teaching reform project [Grant Numbers 2019ylzy01, 2019kcbz02], the Quality Engineering Teaching Research Project in Wannan Medical College [Grant Numbers 2021ylkc03, 2022jyxm08], and Collaborative Innovation Project of Universities in Anhui Province [Grant Number GXXT-2021-087].

Data availability The datasets used in this research work are available at various repositories such as the Kaggle COVID-19 radiography dataset, Kaggle X-ray COVID-19 dataset and Kaggle CoronaHack chest X-ray dataset.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Ethical approval Not applicable.

References

1. Hopkins J et al (2020) Coronavirus resource center. Im Internet (Stand: 19.04.2020). <https://coronavirus.jhu.edu/data>
2. Leung T, Chan A, Chan E, Chan V, Chui C, Cowling B, Gao L, Ge M, Hung I, Ip M et al (2020) Short-and potential long-term adverse health outcomes of COVID-19: a rapid review. *Emerg Microbes Infect* 9(1):2190–2199
3. Manna S, Wruble J, Maron SZ, Toussie D, Voutsinas N, Finkelstein M, Cedillo MA, Diamond J, Eber C, Jacobi A et al (2020) COVID-19: a multimodality review of radiologic techniques, clinical utility, and imaging features. *Radiol Cardiothorac Imaging* 2(3):e200210
4. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
5. 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
6. Szegedy C, Ioffe S, Vanhoucke V, Alemi A (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol 31

7. 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* 24(3):1207–1220
8. Goel T, Murugan R, Mirjalili S, Chakrabartty DK (2022) Multi-COVID-Net: multi-objective optimized network for COVID-19 diagnosis from chest X-ray images. *Appl Soft Comput* 115:108250
9. Tharwat A, Houssein EH, Ahmed MM, Hassanien AE, Gabel T (2018) Mogo algorithm for constrained and unconstrained multi-objective optimization problems. *Appl Intell* 48(8):2268–2283
10. Fitriasari HI, Rizkinia M (2021) Improvement of Xception-ResNet50V2 concatenation for COVID-19 detection on chest X-ray images. In: 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT). IEEE, pp 343–347
11. Karacı A (2022) VGGCOV19-NET: automatic detection of COVID-19 cases from X-ray images using modified VGG19 CNN architecture and YOLO algorithm. *Neural Comput Appl* 34(10):8253–8274
12. Al-Antari MA, Hua C-H, Bang J, Lee S (2021) Fast deep learning computer-aided diagnosis of COVID-19 based on digital chest X-ray images. *Appl Intell* 51(5):2890–2907
13. Chola C, Mallikarjuna P, Muaad AY, Bibal Benifa J, Hanumanthappa J, Al-antari MA (2021) A hybrid deep learning approach for COVID-19 diagnosis via CT and X-ray medical images. In: *Computer Sciences & Mathematics Forum*, vol 2. MDPI, p 13
14. Ukwuoma CC, Qin Z, Heyat MBB, Akhtar F, Bamisile O, Muaad AY, Addo D, Al-Antari MA (2023) A hybrid explainable ensemble transformer encoder for pneumonia identification from chest X-ray images. *J Adv Res* 48:191–211
15. Ukwuoma CC, Qin Z, Heyat MBB, Akhtar F, Smahi A, Jackson JK, Furqan Qadri S, Muaad AY, Monday HN, Nneji GU (2022) Automated lung-related pneumonia and COVID-19 detection based on novel feature extraction framework and vision transformer approaches using chest X-ray images. *Bioengineering* 9(11):709
16. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
17. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 25
18. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) Mobilenets: efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*
19. Tan M, Le, Q (2019) Efficientnet: rethinking model scaling for convolutional neural networks. In: *International Conference on Machine Learning*. PMLR, pp 6105–6114
20. Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: towards real-time object detection with region proposal networks. *Adv Neural Inf Process Syst* 28
21. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 3431–3440
22. Hemdan EED, Shouman MA, Karar ME (2020) Covidx-net: a framework of deep learning classifiers to diagnose COVID-19 in X-ray images. *arXiv preprint arXiv:2003.11055*
23. Nigam B, Nigam A, Jain R, Dodia S, Arora N, Annappa B (2021) COVID-19: automatic detection from X-ray images by utilizing deep learning methods. *Expert Syst Appl* 176:114883
24. 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
25. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 779–788
26. Chouhan V, Singh SK, Khamparia A, Gupta D, Tiwari P, Moreira C, Damaševičius R, De Albuquerque VHC (2020) A novel transfer learning based approach for pneumonia detection in chest X-ray images. *Appl Sci* 10(2):559
27. Xia X, Togneri R, Soheli F, Huang D (2018) Auxiliary classifier generative adversarial network with soft labels in imbalanced acoustic event detection. *IEEE Trans Multimed* 21(6):1359–1371
28. Waheed A, Goyal M, Gupta D, Khanna A, Al-Turjman F, Pinheiro PR (2020) Covidgan: data augmentation using auxiliary classifier gan for improved covid-19 detection. *IEEE Access* 8:91916–91923
29. Haritha D, Praneeth C, Pranathi MK (2020) Covid prediction from X-ray images. In: 2020 5th International Conference on Computing, Communication and Security (ICCCS). IEEE, pp 1–5
30. 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

31. Jyoti K, Sushma S, Yadav S, Kumar P, Pachori RB, Mukherjee S (2023) Automatic diagnosis of COVID-19 with MCA-inspired TQWT-based classification of chest X-ray images. *Comput Biol Med* 152:106331
32. Chamseddine E, Mansouri N, Soui M, Abed M (2022) Handling class imbalance in COVID-19 chest X-ray images classification: using smote and weighted loss. *Appl Soft Comput* 129:109588
33. Barshooi AH, Amirkhani A (2022) A novel data augmentation based on Gabor filter and convolutional deep learning for improving the classification of COVID-19 chest X-ray images. *Biomed Signal Process Control* 72:103326
34. 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
35. Afshar P, Heidarian S, Naderkhani F, Oikonomou A, Plataniotis KN, Mohammadi A (2020) Covid-caps: a capsule network-based framework for identification of covid-19 cases from X-ray images. *Pattern Recogn Lett* 138:638–643
36. Montalbo FJP (2021) Diagnosing covid-19 chest X-rays with a lightweight truncated densenet with partial layer freezing and feature fusion. *Biomed Signal Process Control* 68:102583
37. Feki I, Ammar S, Kessentini Y, Muhammad K (2021) Federated learning for COVID-19 screening from chest X-ray images. *Appl Soft Comput* 106:107330
38. Bhattacharya A, Gawali M, Seth J, Kulkarni V (2022) Application of federated learning in building a robust COVID-19 chest X-ray classification model. *arXiv preprint [arXiv:2204.10505](https://arxiv.org/abs/2204.10505)*
39. Han K, Wang Y, Chen H, Chen X, Guo J, Liu Z, Tang Y, Xiao A, Xu C, Xu Y et al (2022) A survey on vision transformer. *IEEE Trans Pattern Anal Mach Intell* 45(1):87–110
40. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Mindecker M, Heigold G, Gelly S et al (2020) An image is worth 16 × 16 words: transformers for image recognition at scale. *arXiv preprint [arXiv:2010.11929](https://arxiv.org/abs/2010.11929)*
41. Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S (2020) End-to-end object detection with transformers. In: *European Conference on Computer Vision*. Springer, pp 213–229
42. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B (2021) Swin transformer: hierarchical vision transformer using shifted windows. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp 10012–10022
43. Zhang L, Wen Y (2021) A transformer-based framework for automatic COVID19 diagnosis in chest CTs. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp 513–518
44. Zhang L, Wen Y (2021) MIA-COV19D: a transformer-based framework for COVID19 classification in chest CTs. *arXiv*
45. Anwar T (2021) COVID19 diagnosis using AutoML from 3D CT scans. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp 503–507 (2021)
46. Krishnan KS, Krishnan KS (2021) Vision transformer based COVID-19 detection using chest X-rays. In: *2021 6th International Conference on Signal Processing, Computing and Control (ISPCC)*. IEEE, pp 644–648
47. Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y (2017) Graph attention networks. *arXiv preprint [arXiv:1710.10903](https://arxiv.org/abs/1710.10903)*
48. Kipf TN, Welling M (2016) Semi-supervised classification with graph convolutional networks. *arXiv preprint [arXiv:1609.02907](https://arxiv.org/abs/1609.02907)*
49. Hamilton WL, Ying R, Leskovec J (2017) Representation learning on graphs: methods and applications. *arXiv preprint [arXiv:1709.05584](https://arxiv.org/abs/1709.05584)*
50. Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE (2017) Neural message passing for quantum chemistry. In: *International Conference on Machine Learning*, pp 1263–1272
51. Li X-S, Liu X, Lu L, Hua X-S, Chi Y, Xia K (2022) Multiphysical graph neural network (MP-GNN) for COVID-19 drug design. *Brief Bioinform* 23(4):231
52. Cheung M, Moura JM (2020) Graph neural networks for COVID-19 drug discovery. In: *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, pp 5646–5648
53. Wang L, Ben X, Adiga A, Sadilek A, Tendulkar A, Venkatramanan S, Vullikanti A, Aggarwal G, Talekar A, Chen J et al (2020) Using mobility data to understand and forecast covid19 dynamics. *medRxiv*
54. Xie H, Li D, Wang Y, Kawai Y (2022) Visualization method for the spreading curve of COVID-19 in universities using GNN. In: *2022 IEEE International Conference on Big Data and Smart Computing (BigComp)*. IEEE, pp 121–128
55. Yu Z, Zheng X, Yang Z, Lu B, Li X, Fu M (2021) Interaction-temporal GCN: a hybrid deep framework for COVID-19 pandemic analysis. *IEEE Open J Eng Med Biol* 2:97–103

56. Song X, Li H, Gao W, Chen Y, Wang T, Ma G, Lei B (2021) Augmented multicenter graph convolutional network for COVID-19 diagnosis. *IEEE Trans Ind Inf* 17(9):6499–6509
57. Lu S, Zhu Z, Gorriz JM, Wang S-H, Zhang Y-D (2022) NAGNN: classification of COVID-19 based on neighboring aware representation from deep graph neural network. *Int J Intell Syst* 37(2):1572–1598
58. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)
59. Cohen JP, Morrison P, Dao L (2020) COVID-19 image data collection. arXiv preprint [arXiv:2003.11597](https://arxiv.org/abs/2003.11597)
60. Chowdhury ME, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, Islam KR, Khan MS, Iqbal A, Al Emadi N et al (2020) Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access* 8:132665–132676
61. Lisa M, Bot H. My research software. <https://doi.org/10.5281/zenodo.1234>. <https://github.com/github/linguist>
62. Hamilton W, Ying Z, Leskovec J (2017) Inductive representation learning on large graphs. *Adv Neural Inf Process Syst* 30 (2017)

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.

Authors and Affiliations

Chang Jie¹ · Chen Jiming² · Shao Ying² · Tong Yanchun³ · Ren Haodong⁴

✉ Tong Yanchun
yctong@zknv.edu.cn

✉ Ren Haodong
rhd@zjnu.edu.cn

Chang Jie
cjfuture2022@gmail.com

Chen Jiming
yjsyycjm@126.com

Shao Ying
shaoying@yjsyy.com

¹ Department of Medical Information, Wannan Medical College, Wuhu 241000, Anhui, China

² Medical Image Center of Yi-Ji Shan Hospital, Wannan Medical College, Wuhu 241000, Anhui, China

³ College of Mathematics and Statistics, Zhoukou Normal University, Zhoukou 466000, Henan, China

⁴ Zhejiang Normal University Library, Zhejiang Normal University, Jinhua 321000, Zhejiang, China