

# A Distance Transformation Deep Forest Framework With Hybrid-Feature Fusion for CXR Image Classification

Qingqi Hong<sup>✉</sup>, *Member, IEEE*, Lingli Lin, Zihan Li<sup>✉</sup>, Qingde Li<sup>✉</sup>, Junfeng Yao<sup>✉</sup>, Qingqiang Wu, Kunhong Liu<sup>✉</sup>, *Member, IEEE*, and Jie Tian<sup>✉</sup>, *Fellow, IEEE*

**Abstract**—Detecting pneumonia, especially coronavirus disease 2019 (COVID-19), from chest X-ray (CXR) images is one of the most effective ways for disease diagnosis and patient triage. The application of deep neural networks (DNNs) for CXR image classification is limited due to the small sample size of the well-curated data. To tackle this problem, this article proposes a distance transformation-based deep forest framework with hybrid-feature fusion (DTDF-HFF) for accurate CXR image classification. In our proposed method, hybrid features of CXR images are extracted in two ways: hand-crafted feature extraction and multigrained scanning. Different types of features are fed into different classifiers in the same layer of the deep forest (DF), and the prediction vector obtained at each layer is transformed to form distance vector based on a self-adaptive scheme. The distance vectors obtained by different classifiers are fused and concatenated with the original features, then input into the corresponding classifier at the next layer. The cascade grows until DTDF-HFF can no longer gain benefits from the new layer. We compare the proposed method with other methods on the public CXR datasets, and the experimental results show that the proposed method can achieve state-of-the-art (SOTA) performance. The code will be made publicly available at <https://github.com/hongqq/DTDF-HFF>.

**Index Terms**—Chest X-Ray (CXR), classification, coronavirus disease 2019 (COVID-19), deep forest (DF), hybrid feature fusion.

## I. INTRODUCTION

PNEUMONIA is one of the leading causes of death among children and elderly people worldwide, with about 4 million patients at risk of dying each year. According to World Health Organization (WHO), pneumonia accounts for 15% of all deaths in children under five years of age [1]. It is an infection caused by viruses, bacteria, or fungi, which will result in inflammation in the lungs and adversely affect the alveoli. Particularly, in the past year, Coronavirus disease 2019 (COVID-19), caused by a novel corona-virus (SARS-CoV-2), has become the most severe epidemic disease in the world. So far, 215 countries and territories have been affected by the COVID-19 epidemic. As of March 21, 2023, in the worldwide, more than 761 million cases of COVID-19 have been confirmed, and more than 6 million deaths have been recorded [2]. It has had a severe impact on human's health and global economy.

Early diagnosis of pneumonia, especially COVID-19, is of great importance in providing the opportunity of performing patient triage in time, planning early interventions, and offering effective treatments. With the continuous development of imaging technology, chest radiography imaging has become a routine tool to estimate whether a patient has pneumonia. Chest computed tomography (CT) and chest X-ray (CXR) have been widely used in clinical practice, which can provide diagnostic basis for screening, diagnosis and differentiation of early pulmonary infection.

Compared to CXR, CT scan has a higher resolution in space and density, which is able to facilitate a quantitative assessment for pulmonary changes. However, CT is not suitable for large-scale screening of pneumonia due to the relatively high cost and high radiation dose [3]. In addition, it requires strict separation of patients with COVID-19 from other general patients to avoid cross infection. With the pandemic of COVID-19, the burden of CT examination in radiology department is increasing along with the potential risk of infection. On the other hand, CXR examination is simple to operate. It requires less imaging time and lower cost. As a supplement to CT, CXR can be utilized to screen patients' initial diagnosis and confirm the priority of patients' treatments, which is helpful for the saturated healthcare system in the pandemic situation [4]. However, CXR images have

Manuscript received 24 October 2021; revised 11 April 2022, 19 October 2022, 24 December 2022, and 28 March 2023; accepted 17 May 2023. This work was supported in part by the Natural Science Foundation of Fujian Province of China under Grant 2020J01006; in part by the Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, under Grant VRLAB2022AC04; in part by the ITC-InnoHK Projects at Hong Kong Centre for Cerebro-Cardiovascular Health Engineering (COCHE); in part by the National Natural Science Foundation of China under Grant 61772023 and Grant 62072388; and in part by the National Key Research and Development Program of China under Grant 2019QY1803. (Corresponding authors: Qingqiang Wu; Kunhong Liu.)

Qingqi Hong is with the Department of Digital Media Technology, Xiamen University, Xiamen 361005, China, and also with the Hong Kong Centre for Cerebro-Cardiovascular Health Engineering (COCHE), Hong Kong (e-mail: hongqq@xmu.edu.cn).

Lingli Lin and Zihan Li are with the School of Informatics, Xiamen University, Xiamen 361005, China.

Qingde Li is with the School of Computer Science, University of Hull, HU6 7RX Hull, U.K. (e-mail: Q.li@hull.ac.uk).

Junfeng Yao and Qingqiang Wu are with the School of Informatics, the Key Laboratory of Digital Protection and Intelligent Processing of Intangible Cultural Heritage of Fujian and Taiwan Ministry of Culture and Tourism, and the School of Film, Xiamen University, Xiamen 361005, China (e-mail: yao0010@xmu.edu.cn; wuqq@xmu.edu.cn).

Kunhong Liu is with the School of Film, Xiamen University, Xiamen 361005, China (e-mail: lkhqz@xmu.edu.cn).

Jie Tian is with the CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: tian@iee.org).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TNNLS.2023.3280646>.

Digital Object Identifier 10.1109/TNNLS.2023.3280646

many very subtle and similar features with low sensitivity, which makes them a cumbersome task for visual inspection. Generally, it is not an easy task to interpret such subtle abnormalities, even for experienced radiologists. Furthermore, the number of suspected infected patients is increasing rapidly, while the number of specialist radiologists is very limited. Hence, there is an urgent need to develop automated methods for identifying those subtle abnormalities in CXR images, which could aid the diagnostic procedure and improve early diagnosis rates with high accuracy. Artificial intelligence (AI) and machine learning solutions are potentially powerful tools for solving such problems [5].

Consequently, researchers in the field of machine/deep learning have been actively exploring approaches to classify CXR images [6], [7], [8], [9], [10], [11], [12], [13]. At present, deep learning methods used in medical image classification are mostly based on deep neural networks (DNNs). Although DNNs are capable of accomplishing the task well in most cases, they require a large number of samples for training and consume a lot of computational resources. DNNs map the original data into the feature space through the convolution layer, pooling layer, and activation function layer. However, DNNs have no obvious advantage in processing features extracted by traditional methods. In recent years, as a new deep learning model integrating trees and forests, deep forest (DF) [14] has been gradually applied in many fields, which can accomplish the corresponding tasks under the conditions of fewer required parameters and computing resources. Compared with the feature space mapped by the convolution layer and pooling layer in DNNs, DF can directly handle various types of features. In the case of COVID-19 pandemic, it is difficult to have a collection of a large number of filtered CXR datasets. Therefore, under the condition of limited training datasets, this article develops an improved DF framework based on mixed feature fusion and distance transformation, i.e. distance transformation-based DF framework with the hybrid-feature fusion (DTDF-HFF), to achieve accurate pneumonia classification on CXR images. Our proposed model uses the traditional feature extraction method to extract the image features and uses multigrained scanning to provide the image information at different resolutions, enriching the feature input and enhancing the robustness of the model.

The overall process of the method proposed in this article is as follows. First, the CXR images are segmented to obtain the lung mask. Two feature extraction methods are carried out on the image of the mask part to generate diverse feature subsets: hand-crafted feature extraction and multigrained scanning. Different types of features are fed into different classifiers in the same layer, and the prediction vectors of each layer are transformed by a self-adaptive method to form distance vectors. The distance vectors obtained by different classifiers are fused and concatenated with the original features and then fed to the corresponding classifiers at the next level. Cascading forests are constructed in this way until they converge. The outputs of the final layer are averaged as the classification result. This framework can be extended to classification tasks of various medical images. We compare the proposed method with other methods on the public CXR data,

and the experimental results show that the proposed method can achieve superior performance to other state-of-the-art deep learning methods.

To the best of our knowledge, our DTDF-HFF is the first framework to exploit DF in the task of CXR image classification. The main contributions of this article are as follows.

- 1) A new DF framework (DTDF-HFF) based on hybrid-feature fusion and distance transformation (DT) is proposed for the accurate classification of CXR images. This framework can be extended to classification tasks of various medical images.
- 2) Multiple types of features are fed to different classifiers, and the sliding window is used to get multiresolution image features to avoid global information loss under a single resolution. Therefore, hybrid features support the high diversity among different classifiers, so as to enhance the discriminative ability of the proposed model.
- 3) A new DT scheme is designed for the DF framework. It replaces the augment vector in the original DF with the distance vector transformed by the adaptive method, aiming to accelerate the convergence.

## II. RELATED WORK

### A. Machine Learning Techniques for Medical Image Classification

During the past few years, a considerable amount of machine-learning techniques have been proposed for medical image classification. Most of these methods can be roughly classified into two groups: classical machine learning-based methods and deep learning-based methods.

Generally, the classical machine learning-based methods require two stages for the classification task. In the first stage, multiple features of medical images are extracted, and then the features are fed to classifiers, such as support vector machine (SVM), Markov random fields, random forests (RF), neural networks, in the second stage for classification [15], [16], [17], [18], [19].

Under the influence of the COVID-19 pandemic, recent studies have mostly focused on the automatic diagnosis of pneumonia using chest radiography images. Zargari et al. [20] used global features of the whole CXR images to build machine learning classifier. The model is able to distinguish COVID-19 cases from non-COVID-19 cases with high accuracy and sensitivity. Dey et al. [21] employed the ensemble feature scheme (EFS) for pneumonia detection in CXR, which combined the hand-crafted features with the deep features and tested in different classifiers.

Recently, more and more studies have focused on medical image classification based on the deep learning technique, which has a strong representation learning ability via training multilayer artificial neural networks [8], [10], [22], [23], [24], [25], [26]. Pun and Agarwal [27] proposed a loss function approach for transfer learning on binary classification of posteroanterior CXR images. Rajpurkar et al. [28] developed a deep learning network termed Chexnet based on ChestX-ray14

for pneumonia detection. The model achieved state-of-the-art results on all 14 diseases in CXRs. Nour et al. [29] proposed a deep convolutional neural network (CNN) model which was trained from scratch to extract discriminative features on CXR images. Rubin et al. [30] proposed a novel DuaNet for processing both frontal and lateral CXR images. Their model improved performance in recognizing findings in CXR images. Lakhani and Sundaram [31] built a model that integrated the AlexNet and GoogleNet for detecting tuberculosis in CXR. Wang et al. [32] utilized a natural language processing method to mine disease image tags, demonstrating that these common chest diseases can be detected by a unified weakly supervised multilabel image classification and disease localization framework.

On account of the COVID-19 pandemic, more researchers have been applying deep learning to pneumonia diagnosis. Ozturk et al. [33] presented the DarkCovidNet for automatic COVID-19 detection in CXR images. Chen et al. [34] designed a system based on CT for the diagnosis of COVID-19 pneumonia. Xu et al. [35] established deep learning models for the early screening of COVID-19 on CT datasets. Oh et al. [4] proposed a patch-based CNN approach with a relatively small number of trainable parameters for COVID-19 diagnosis. The Covid-Net, an open-source deep CNN platform, was proposed in [6] for detecting chest radiographs of COVID-19 cases. Ozturk et al. [33] proposed a model for the automatic detection of COVID-19 for binary classification diagnosis. Jadon [36] proposed a custom few-shot learning method to detect COVID-19 using Siamese networks. Chaddad et al. [37] fed GMM-CNN features into a robust classifier to distinguish COVID-19 from other cases of pneumonia. CoVIRNet was proposed by Almalki et al. [38] to automatically diagnose COVID-19 patients from CXR images. Dong et al. [39] proposed a learning model called RCoNet, which can distinguish COVID-19 from ordinary pneumonia more accurately and robustly. Irfan et al. [40] proposed a hybrid DNN (HDNNs) to predict the risk of developing COVID-19 in patients using CT and X-ray imaging. A deep uncertainty-aware transfer learning framework was proposed by Shamsi et al. [41] for the detection of COVID-19-infected patients using CT and CXR images.

Despite more other works in literature works, the works [42], [43], [44], [45], [46], [47], [48] report excellent performance in detecting COVID-19, most deep learning techniques for medical image classification are based on deep CNNs, which generally require a large number of training samples or pretrained models. Currently, collecting sufficient data samples with proper annotations is still one of the main challenges for effective training of DNN models.

### B. Deep Forest

Inspired by the layer-to-layer of DNNs which abstract information layer by layer, Zhou and Feng [14] put forward a novel decision tree-based ensemble deep learning framework with layers of cascade structure. Each layer contains two different types of forests with the aim of enhancing diversity. Although the structures similar to DNNs, DF does not rely on back

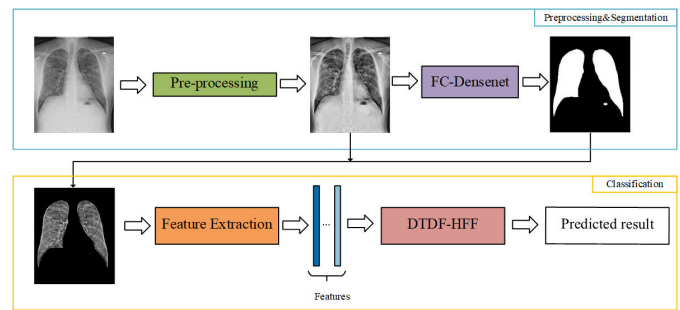


Fig. 1. Overall architecture of the proposed method. 1) The CXR images are preprocessed and segmented to generate the mask of lung regions. 2) Multiple types of features are extracted and input into our DTDF-HFF for classification.

propagation. Compared with DNNs, DF requires much less training data and is easier to be trained with fewer parameters.

A number of studies have begun to develop variant models based on DF, including the deep multifeature fusion for hyperspectral image feature fusion and classification [49], the Deep-Resp-Forest for the classification of the anticancer drug response [50], the multilabel based DF (MLDF) for the multilabel problems [51], the weighted DF for Schizophrenia data classification [52], and the adaptive feature selection guided DF for COVID-19 classification with chest CT [53]. These studies have demonstrated that DF can well handle a set of challenges in a variety of fields [54], [55], [56], [57]. Given the limited medical image datasets, this study tries to develop the DF with hybrid feature fusion, aiming to improve the diagnostic accuracy of lung diseases.

## III. PROPOSED METHOD

The overall architecture of the proposed method is shown in Fig. 1. The CXR images are first preprocessed and segmented by FC-DenseNet [58] to generate the mask of lung regions, which is subsequently processed by two feature extraction schemes: hand-crafted feature extraction and multigrained scanning. These features are fed to the DTDF-HFF, which is a new framework for accurate CXR image classification. The details are given as follows.

### A. Preprocessing and Segmentation

CXR images mainly contain left lung, right lung, heart, and background. The segmentation technique is used to extract the lung region to remove the negative impact from the irrelevant features based on FC-DenseNet103. Since the original images are collected from different datasets under various conditions, the datasets are processed and normalized to ensure that the images can be normally segmented.

- 1) Adopt the contrast-first adaptive histogram equalization algorithm (CLAHE) to make the histogram strength of data uniform.
- 2) Adjust all images to the same size before inputting into FC-DenseNet103 for image segmentation.
- 3) Detect the contour shape after image segmentation and calculate the contour area.
- 4) Keep the top two contour images in area size and remove the rest contours with noise in the image.



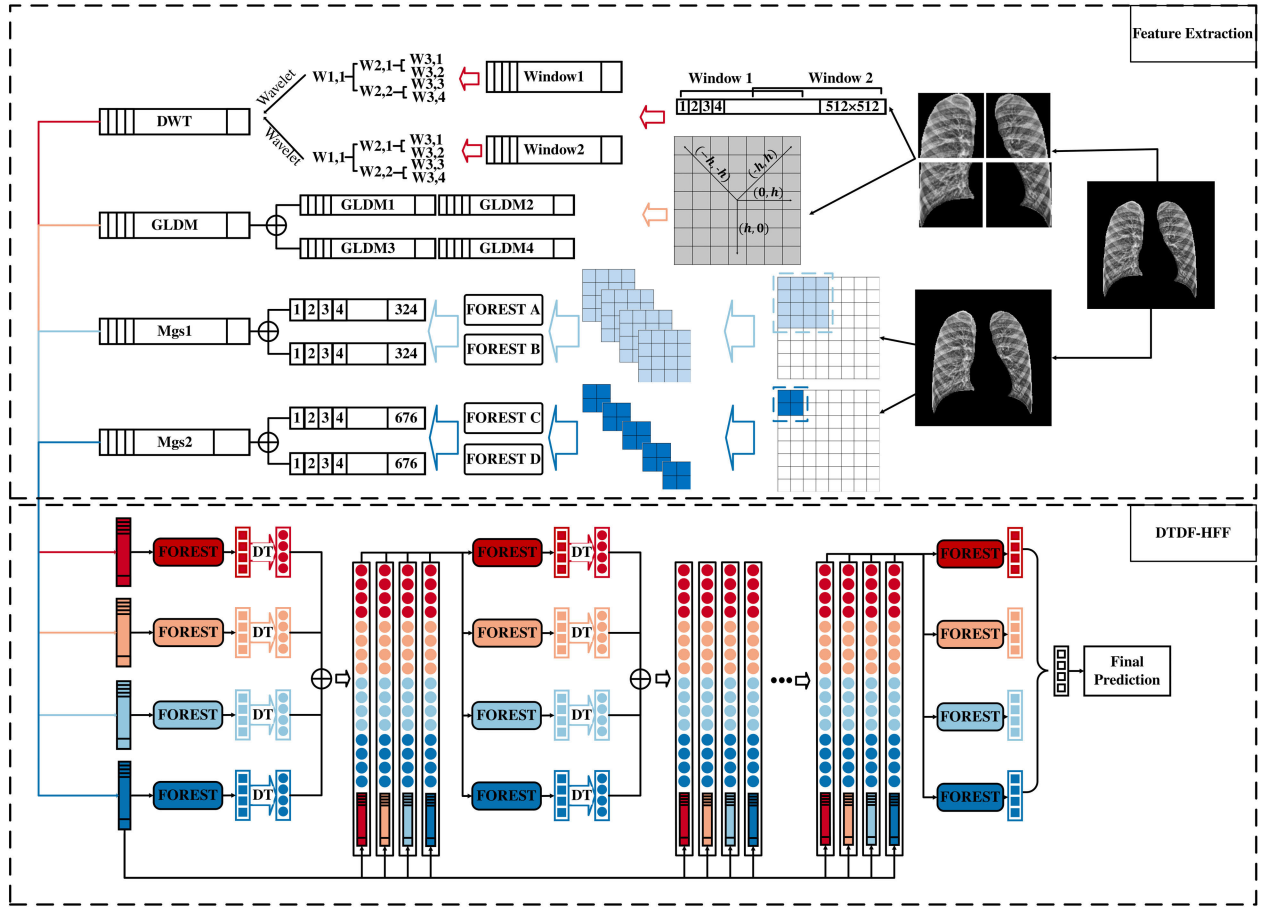


Fig. 2. Structure of our DTDF-HFF. 1) We extract hand-crafted features, respectively, for the upper and lower parts of the left and right lungs and generate multigrained scanning features for the feature re-representation. 2) Different colors represent different input features. 3) DT vector is calculated by DT of the prediction probability. 4) The inputs in the second and subsequent layers are concatenation of feature and DT vectors.

### B. Hybrid Feature Extraction

Owing to the rich information of pathological features in CXR images, multiple features obtained from different processing techniques are applied to mine image features and their rich internal information fully [54]. The common hand-crafted features obtained from discrete wavelet transform (DWT) and gray level difference method (GLDM), as well as multigrained scanning features are applied in our framework.

1) *Hand-Crafted Feature*: There is no texture information in the masked image except for the lung areas. Therefore, it is necessary to reduce the useless information of these untextured regions in the extraction of hand-crafted features. As shown in Fig. 2, we clip most of the untextured areas around the lung areas and scale the image to a uniform size, and then extract features, respectively, for the upper and lower parts of the left and right lungs.

The features of the low-frequency components, horizontal high-frequency components, vertical high-frequency components, and diagonal high-frequency components are obtained by 2-D DWT. Assuming that  $\psi(r)$  is a wavelet centered at  $r = 0$ . Let

$$\psi_{a,b}(r) = \frac{1}{\sqrt{a}} \psi\left(\frac{r-b}{a}\right) \quad (1)$$

where  $a$  and  $b$  are scale factors and shift factors defined discretely, respectively. And according to certain given base scale factor  $a_0$  and shift factor  $b_0$ ,  $a$ , and  $b$  are defined as  $a = a_0^\alpha$ ,  $a_0 > 0$ ,  $\alpha \in \mathbb{Z}$ ;  $b = \beta b_0 a_0^\alpha$ ;  $\beta \in \mathbb{R}$ ,  $n \in \mathbb{Z}$ . More specifically, for each pair of  $(\alpha, \beta)$ , the corresponding wavelet is

$$\psi_{\alpha,\beta}(r) = a_0^{-\frac{\alpha}{2}} \psi(a_0^{-\alpha} r - \beta b_0). \quad (2)$$

The DWT of a function  $f(r)$  is defined as

$$\text{DWT}(\alpha, \beta) = \int_{\mathbb{R}} f(r) \psi_{\alpha,\beta}(r) dr. \quad (3)$$

The features based on the GLDM are block-based localized features, which are derived from the  $P$  function [59]. These features are calculated by the  $P$  function for eight gray levels in four different directions. For any given displacement vector  $h = (dx, dy)$ , the probability density function of gray level  $g$  of image  $I$  is defined as

$$P(g | h) = \text{Prob}(\text{dif } f(\theta, \rho) = g) \quad (4)$$

where  $\text{dif } f(\theta, \rho) = |I(\theta, \rho) - I(\theta + dx, \rho + dy)|$ .

GLDM features are extracted from four directions of each part and DWT features are extracted from 8 wavebands of each part. The features obtained from each part are calculated by

statistical methods, including Mean, Std, Skewness, Kurtosis, Energy, Entropy, Max, Min, Mean Deviation, Median, Range, rms, Uniformity, MeanGradient, and StdGradient. Eventually, 448-D features are extracted from DWT and 224-D features are extracted from GLDM, and they are concatenated to form a 672-D vector for each sample.

2) *Multigrained Scanning Features (Mgs)*: Besides the extracted hand-crafted features, the sliding window scanning is deployed for feature re-representation. As shown in the bottom half of Fig. 1, the images are processed by the sliding windows to input extra-trees (ETs). The feature dimension  $G$  generated for each ET is

$$G = C \times \left( \left\lceil \frac{(z_l - w)}{s} \right\rceil + 1 \right) \times \left( \left\lceil \frac{(z_w - w)}{s} \right\rceil + 1 \right) \quad (5)$$

where  $C$  is the class number, and  $z_l, z_w$  are the image length and width dimensions;  $w$  is the size of a scan window, and  $s$  is the length of stride.

In this way, our proposed framework accommodates four different types of features, which are fed to different classifiers to enrich the flexibility and versatility of the model.

### C. DT Scheme

The DT scheme is designed to enhance the discriminative ability of the DF framework, as shown in Fig. 2.

In the original cascading structure, different classifiers in the same layer receive the same features as inputs. Their outputs are combined to form the augment vector, which is fused with the original feature set to form the input for the next layer. While in our algorithm, hybrid features are input to train the corresponding ETs, so as to enhance the diversity among them. As shown in Fig. 2, the four ETs in each layer output four probability vectors, which are combined as the augment vector.

After the preprocessing, there would still exist noise in the image data, so that the probabilities corresponding to different classes could still be biased toward the major classes. Furthermore, the inconspicuous lesions as well as characterization of different diseases may have similar appearance on some regions of the samples. As a result, the probability values of two classes in the prediction vector may not be of great difference. When the output probabilities of different classes of the model are close, the diseases represented by the images may be misclassified.

To optimize the information from the probability vector, the confusion matrix is deployed to measure the correlation strength of various classes by a self-adaptive method. The confusion matrix reflects the relationship between the ground-truth labels and the prediction results. Assume that for the classification task of  $C$  classes, the identification dataset  $S$  contains  $n$  samples. Let  $cm_{i,j}$  represent the number of samples of class  $i$  judged by the  $m$ th ET as class  $j$ . Then, the confusion

matrix  $CM$  of ET  $m$  is a  $C \times C$  matrix

$$CM(S, m) = \begin{bmatrix} cm_{1,1} & \cdots & cm_{1,j} & \cdots & cm_{1,C} \\ cm_{2,1} & \cdots & cm_{2,j} & \cdots & cm_{2,C} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ cm_{i,1} & \cdots & cm_{i,j} & \cdots & cm_{i,C} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ cm_{C,1} & \cdots & cm_{C,j} & \cdots & cm_{C,C} \end{bmatrix}. \quad (6)$$

The main steps of the distance vector mapping are given as follows.

① Generate the output probability vectors with the corresponding ETs. Suppose there are  $M$  ETs in the  $l$ th layer, and  $T$  trees in each ET, then the probability vector of the  $n$ th sample is

$$v_{m,l}^n = \frac{1}{T} \sum_{t=1}^T p_{m,l}^{n,t} \quad (7)$$

where  $p_{m,l}^{n,t}$  is the prediction probability of the  $n$ th sample in tree  $t$  of the  $m$ th ET in layer  $l$ .

② Calculate the distance between each row of the normalized confusion matrix  $CM(S, m)$  and the probability vector  $v_{m,l}^n$ . The distance evaluation method adopted here is Euclidean distance, and the distance between the probability vector and a row vector  $C_n$  of the normalized confusion matrix is calculated by

$$d_{m,l}^{n,c} = \sqrt{\sum_{k=1}^C (v_{m,l}^{n,k} - cm_{c,k})^2}. \quad (8)$$

③ Generate the input vectors for the corresponding classifiers of the next layer. After calculating the distance, the  $N$ -dimensional enhancement vectors are fused and concatenated with the corresponding original features and input to the corresponding classifiers of the next layer. Let  $D_{m,l}^n = (d_{m,l}^{n,1}, \dots, d_{m,l}^{n,C})$  be the  $n$ th sample's DT vector of  $m$  ET in  $l$ th layer. Let  $V_l^n = (D_{1,l}^n, \dots, D_{M,l}^n)$  be the concatenation vectors of distance vectors. The training feature of sample  $n$ th for ET  $m$  in the next layer is

$$X_{m,l}^n = (S_m^n, V_l^n) \quad (9)$$

where  $S_m^n$  is the  $m$ th feature of  $n$ th sample.

④ Calculate the final prediction result at the last layer. The layer obtaining the highest performance on the validation set is deployed as the output layer. And after obtaining the prediction vectors of all ETs in the last layer, the  $n$ th sample is assigned to the class corresponding to the maximum value of the averaged prediction vector, as given below

$$V^n = \frac{1}{M} \sum_{m=1}^M (v_{m,l}^n). \quad (10)$$

Algorithm 1 presents the process of training and prediction based on our DT DF with hybrid-feature fusion (DTDF-HFF). The dataset  $S$  is divided into the training set  $S_{Tr} = (S_{Tr}^1, \dots, S_{Tr}^M)$  and the test set  $S_{Test} = (S_{Test}^1, \dots, S_{Test}^M)$ . Samples in  $S_{Test}$  are regarded as unknown samples. At the training phase, after each layer is generated, the overall performance

**Algorithm 1** Distance Transformation Deep Forest With Hybrid-Feature Fusion

---

```

1:  $V_l^{X_{Tr}} = [], V_l^{X_{Test}} = []$ 
2: //Training stage
3: for  $l = 1 : L$  do
4:   if  $i = 1$  then
5:      $\text{let } X_{Tr} = S_{Tr}$ 
6:   else
7:      $\text{let } X_{Tr} = \text{concat}(S_{Tr}, V_{l-1}^{X_{Tr}})$ 
8:   end if
9:   for each  $m = 1 : M$  do
10:    train  $ET(m)$  using  $X_{Tr}$  to get the  $V_{m,l}^{X_{Tr}}$  and  $CM(X_{Tr}, m)$ 
11:    calculate  $D_{m,l}^{X_{Tr}}$  by formula 8
12:   end for
13:    $V_l^{X_{Tr}} = \text{concat}(D_{1,l}^{X_{Tr}}, \dots, D_{m,l}^{X_{Tr}})$ 
14: end for
15: //Testing stage
16: for  $l = 1 : L - 1$  do
17:   if  $i = 1$  then
18:      $\text{let } X_{Test} = S_{Test}$ 
19:   else
20:      $\text{let } X_{Test} = \text{concat}(S_{Test}, V_{l-1}^{X_{Test}})$ 
21:   end if
22:   for each  $m = 1 : M$  do
23:    feed  $X_{Test}$  into  $ET(m)$  to get the  $V_l^{X_{Test}}$  and  $CM(X_{Tr}, m)$ 
24:    calculate  $D_{m,l}^{X_{Test}}$  by formula 8
25:   end for
26:    $V_l^{X_{Test}} = \text{concat}(D_{1,l}^{X_{Test}}, \dots, D_{m,l}^{X_{Test}})$ 
27: end for
28: if  $l = L$  then
29:   for each  $m = 1 : M$  do
30:    feed  $X_{Test}$  into  $ET(m)$  to get  $v_{m,L}^{X_{Test}}$ 
31:   end for
32:   calculate  $V^{X_{Test}}$  by formula 10
33:    $Y_{Test} = \text{argmax}(V^{X_{Test}})$ 
34: end if

```

---

of the model is calculated to determine whether the cascade forest continues to grow. At the testing phase, the outputs of the last layer are combined together as the output vectors, and the corresponding unknown sample in the test set is assigned to the class represented by the highest probability in the averaged output vector. The performance of our algorithm is verified in the following experiments.

#### IV. EXPERIMENTAL RESULTS AND DISCUSSION

##### A. Data Source and Processing

The pandemic of COVID-19 has put limited medical resources under serious and continuous strain, making the pneumonia-related diseases diagnosed to become an urgency. In general, most community-acquired pneumonia (CAP) patients suffered from bacterial infection. Studies have shown that *Streptococcus pneumoniae*, *Haemophilus influenza*, and virus pneumonia are the common causes of CAP. Even in the

TABLE I  
PUBLIC CXR DATASETS USED IN EXPERIMENTS

Datasets	Class	Number	Mask
JSRT/SCR	Normal/Nodule	227	Left lung and right lung
NLM	TB	58	—
CoronaHack	bacteria/virus /normal/covid	426	—
Cohen et al	bacteria/covid	216	—

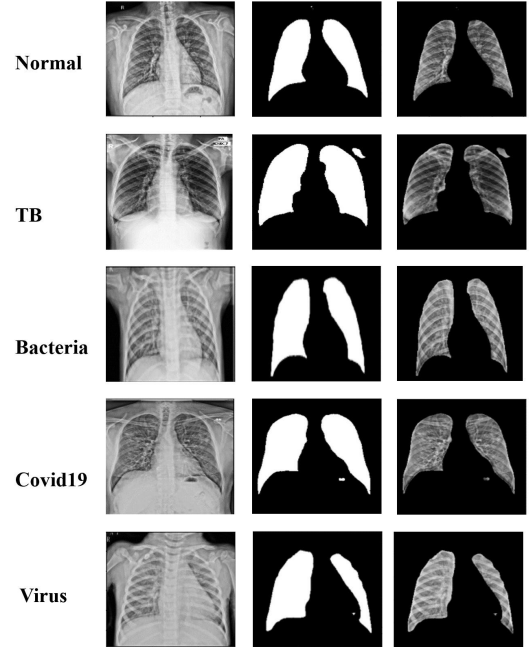


Fig. 3. Example of the preprocessed images (left), together with their corresponding segmentation results (middle), and the extracted lung areas (right) for different categorical classes of CXR images.

case of the COVID-19 pandemic, the incidence of bacterial pneumonia and tuberculosis remains high. Therefore, in order to distinguish between common lung diseases and COVID-19, we have conducted experiments based on the public CXR datasets as shown in Table I.

The JSRT [60] and SCR [61] datasets are considered for segmentation network training. Corresponding segmentation masks were collected from the SCR dataset. To compare COVID-19 with normal and different lung diseases, the CXR images were also collected from different dataset sources, such as NLM [62], CoronaHack [63], and Cohen et al. [64], which are fully open to any research community. These images have different dimensions and are resized to 1024 pixels. In our experiments, CXR images are divided into four classes: normal, tuberculosis (TB), COVID-19, and pneumonia (bacteria and virus), and the COVID-19 label is distinguished from other pneumonia. The combined dataset includes 218, 58, 204, and 220 images for normal, TB, pneumonia (bacteria and virus), and COVID-19, respectively. All our experimental data are based on the ten-fold cross-validation. That is, the dataset is divided into training set and test set, and test

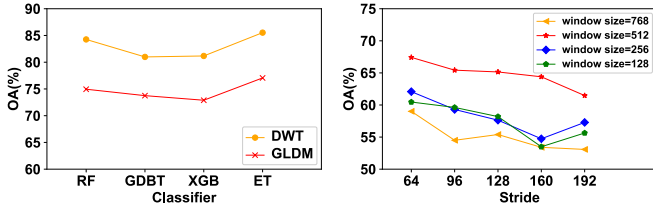


Fig. 4. Accuracy of GLDM and DWT on different base classifiers (left). Accuracy of different sliding window sizes on ET (right).

set in each fold is one-tenth of the total data, which is a conventional division way for small datasets. In the training phase, cross-validation will be conducted after the generation of a new layer. In other words, part of the data in the training set will be used as the validation set to generate evaluation indicators to determine whether DTDF-HFF continues to grow. The training data used in the segmentation are all JSRT/SCR datasets. The other datasets are segmented by the trained FC-Densenet model, and the obtained images after applying the mask are used for the training and test of the classification model. After segmentation, the mask is restored to the same original size of CXR images. Fig. 3 shows an example of the preprocessed images, together with their corresponding segmentation results, and the extracted lung areas for different classes of CXR images.

### B. Performance Metrics and Experimental Environments

For the performance comparisons, six measurements are used based on true-positive (TP), true-negative (TN), false-positive (FP), false-negative (FN), as given as follows:

$$1) \text{ Accuracy} = (TN + TP) / (TN + TP + FN + FP) \quad (11)$$

$$2) \text{ Precision} = TP / (TP + FP) \quad (12)$$

$$3) \text{ Recall} = \text{sensitivity} = TP / (TP + FN) \quad (13)$$

$$4) \text{ F1 score} = 2(\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (14)$$

$$5) \text{ Specificity} = TN / (TN + FP) \quad (15)$$

$$6) \text{ Youden Index} = \text{Recall} + \text{Specificity} - 1. \quad (16)$$

The experimental environments in this article are as follows: DTDF-HFF and other decision tree-based methods run on the 32 Intel(R) Xeon(R) CPU E5-2665 0 @ 2.40 GHz, and DNN-based methods run on VGA compatible controller: NVIDIA Corporation GP102 [GeForce GTX 1080 Ti].

### C. Feature Extraction and Base Classifier Selection

Since the size of each sample fed into the DTDF-HFF needs to be the same, multiple statistical indicators such as the mean, maximum, and minimum values of the DWT and GLDM features are calculated and concatenated as the input features. The accuracy of these features on different base classifiers is observed in Fig. 4 (left), which shows that ET achieves the best performance compared with other base classifiers, such as XGBoost (XGB), RF, and gradient boosting decision tree (GDBT).

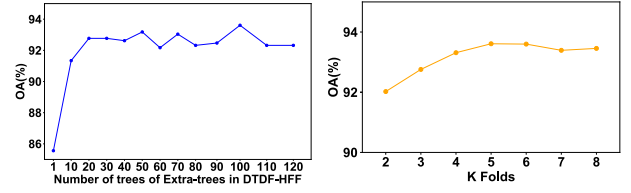


Fig. 5. Impact of the number of ETs on the performance of DTDF-HFF (left). K-value of the  $k$  fold cross validation of ET in DTDF-HFF (right).

TABLE II  
HYPERPARAMETERS OF DTDF-HFF

Hyperparameter	value
Number of ET in each layer	4
Number of trees in each ET	100
K-Fold cross validation	5
Early stop rounds	3

As for the multigrained scanning, the window size may affect the generation of features, and the accuracy is observed by using ET to determine the relevant parameters. Two multigrained scanning features with different parameters (consistent with the number of hand-crafted features) are selected for the training of our proposed model. The features obtained by two sliding window sizes reflect the image information under different resolutions. However, as shown in Fig. 4 (right), the accuracy using the features with a different stride of 768 window sizes is low, so the features with higher accuracy corresponding to other window sizes are selected to participate in our model training. The sliding window sizes of the two types of features are set to 512, 256, respectively, and the stride sizes are both set to 64. The corresponding feature dimensions generated by multigranularity scanning are 648 (Mgs1), 1352 (Mgs2). It should be noted that, when the image resolution is high and the stride size is small, the operation of a sliding window is kind of memory-consuming. However, it is not necessary to use a very small stride size of sliding window for high-resolution image, which has a little influence on the performance of the proposed framework.

### D. Hyperparameter Analysis

As the advantage of the DF framework, the number of hyperparameters of DTDF-HFF is smaller than those of neural networks, which are summarized in Table II. As the analysis above, ET is adopted as the base classifier in our model for training and prediction. The number of trees in each ET is a key parameter. The accuracy is observed with the changes of tree numbers in each ETs of DTDF-HFF. As shown in Fig. 5 (left), the accuracy generally increases with the increase of the tree number. When the number of trees in the ETs is less than 20, the performance of the model is obviously poor. This is because the learning ability of the model is greatly limited by the small scale of ETs. When the number of ETs is greater than 50, the performances of the model do not vary much. It is observed that the accuracy of DTDF-HFF with 110 ETs is even slightly lower than that using 100 ETs, and more ETs



TABLE III  
ACCURACY OF DIFFERENT DECISION TREE-BASED METHODS

Feature	RF	GDBT	XGB	ET	DF	DTDF-HFF
Mgs1	65.79	64.26	65.94	67.42	71.39	<b>73.05</b>
Mgs2	61.73	60.63	59.89	61.57	64.66	<b>65.25</b>
DWT	84.26	80.99	81.17	85.53	86.48	<b>86.91</b>
GLDM	74.95	73.75	72.89	75.98	77.06	<b>81.02</b>
Mgs1+Mgs2 +DWT+GLDM	86.64	86.81	86.51	88.37	88.49	<b>93.61</b>

TABLE IV  
CLASSIFICATION RESULTS OF DIFFERENT DECISION TREE-BASED METHODS

Method	Accuracy	Precision	Recall	F1 score	Specificity	Youden Index
RF	86.64	89.31	87.82	87.88	95.28	83.10
GDBT	86.81	89	87.76	87.41	95.34	83.09
XGB	86.51	89.44	87.80	87.84	95.20	83.00
ET	88.37	89.63	89.29	89.06	95.89	85.18
DF	88.48	90.41	89.32	89.32	95.76	85.08
DTDF-HFF	<b>93.61</b>	<b>95.01</b>	<b>93.66</b>	<b>93.93</b>	<b>97.74</b>	<b>91.40</b>

means higher computational cost. Thus, the number of ETs in our model is set to 100.

The cascade structure is key to the excellent performance of deep learning methods. The early stop parameter of the DTDF-HFF is set to 3. That is, the training process terminates when the accuracy of the  $i$ th layer is not larger than that of the  $(i - 3)$ th layers. Then the number of layer is set to  $i$ . This accuracy is based on the training set. As shown in Fig. 5 (right), our DTDF-HFF can achieve the best classification performance, when the fivefold cross-validation is adopted in the training phase. When  $K$  is greater than 5, increasing the number of  $K$  folds increases the training time of the model, but does not improve the classification performance significantly. Cascade forest evaluates whether the model will continue learning or not based on pre-set evaluation indicators, which are derived from the training set. Therefore, when evaluating whether DTDF-HFF continues learning or not, evaluation indicators is calculated through fivefold cross validation. Usually the number of layers increases, the accuracy of the model is improved. The observation of experimental results shows that our model grows no more than five layers in most cases.

#### E. Comparison With Different Decision Tree-Based Methods

In order to verify the effectiveness of model optimization, hand-crafted features, and multigrained features are fed to original DF and other decision tree-based methods (including RF, GDBT, XGB, and ET) for comparison.

Table III shows the accuracy of single and fused features on different decision tree-based methods. In accordance with the previous observation, the performance of ET for a single feature is slightly lower than that of DF and our method, but superior to other decision tree-based methods. Therefore, ET is selected as the classifier of each layer in our model. As shown

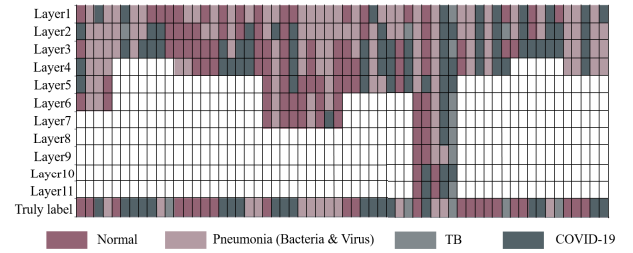


Fig. 6. Typical training results in different layers of DTDF-HFF without the DT (the colored area represents the predicted result of the corresponding layer sample, and the white area represents the stop of the model).

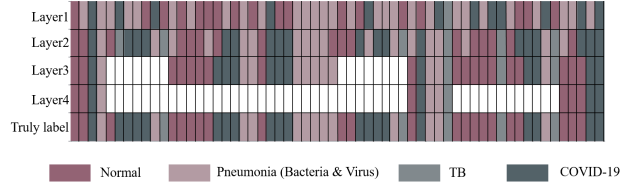


Fig. 7. Typical training results in different layers of DTDF-HFF with the DT (the colored area represents the predicted result of the corresponding layer sample, and the white area represents the stop of the model).

in the table, the results obtained using DWT is the best in each model when compared with those using a single type of feature. And it is obvious that the results based on the ensemble methods are much better than those using a single learner, and the performance of DF and our model are much better than that of other methods.

As for the results obtained based on the hybrid features, since except for our model, other models cannot handle multiple types of features at the same time, different types of features are concatenated to serve as their inputs. It can be seen that the change of inputs offer slightly higher discriminative ability. On the other hand, the input of multiple features to our model greatly improves its performance. To some extent, the diversity of these features maintains high generalization ability of our model.

Although it can be seen from the comparison of the single feature-based classification performances of DF and DTDF-HFF in Table III, the DT DF has achieved better accuracy. But the effect is not significant and is not compared in the case of multiple features. Hence, in order to verify the optimization effect of the predicted probability vector, we compare and visualize the predicted results of the model with and without the self-adaption method layer by layer. Fig. 6 shows the prediction results of the samples at different layers without the self-adaption method in the DTDF-HFF. By comparing the prediction results with the self-adaption method in Fig. 7, it can be seen that the self-adaption method gets the correct prediction at earlier levels. In other words, the enhancement vector can optimize the training of the model, and the DT vector corrects the misclassified samples of the previous layer in the training process of the subsequent layers. In addition, it is obvious that the model trained by distance vectors can converge earlier.

Besides accuracy, other performance metrics of the different decision tree-based methods are presented in Table IV, which



TABLE V  
CLASSIFICATION RESULTS OF DIFFERENT SOTA METHODS

Method	Accuracy	Preci-sion	Recall	F1 score	Specifi-city	Youden Index
Resnet34	88.12	89.55	89.57	88.84	95.44	85.01
CovidCXR	88.99	90.66	90.34	89.99	96.19	86.53
Covid-classifier	83.33	85.97	82.86	83.31	82.14	65.02
DTDF-HFF	<b>93.61</b>	<b>95.01</b>	<b>93.66</b>	<b>93.93</b>	<b>97.74</b>	<b>91.40</b>

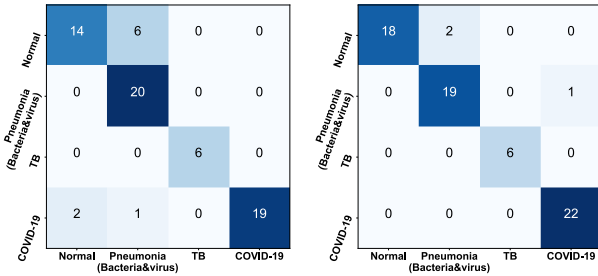


Fig. 8. Confusion matrices for CovidCXR (left) and DTDF-HFF (right).

shows that the proposed DTDF-HFF can achieve the best performances when compared with other decision tree-based methods.

#### F. Comparison With Other SOTA Methods

The performance of DTDF-HFF is compared with other state-of-the art (SOTA) approaches, and the overall performances of the proposed method and the other SOTA approaches on the above datasets are presented in Table V. The CovidCXR method segments the lung area from CXR image first, and then focuses on the random patches of the lung area for model training. The covid-classifier directly uses a variety of hand-crafted features for the training process. As shown in Table V, our method is superior to the other SOTA methods. In detail, compared with Resnet34, CovidCXR [4], and covid-classifier [20], our method gets higher scores with 5.49%, 4.62%, and 10.28% accuracy scores, and 5.09%, 3.94%, and 10.62% F1-score scores. As the accuracy results indicate the overall performance of the method on all test samples, and the F1-score results reveal the balance between the Precision and Recall measurements, these higher results confirm that the performance of our method is much better than other SOTA methods in general. The confusion matrices for CovidCXR (left) and DTDF-HFF (right) are shown in Fig. 8, which verifies that our method can better recognize the COVID-19 samples.

The detailed results of different classes are tabulated in Table VI. It is observed that, our method is able to achieve remarkable classification performances in most metrics of each class. Especially, compared with Resnet34, CovidCXR, and covid-classifier, the proposed DTDF-HFF has got higher scores with 1.98%, 0.55%, and 9.66% F1-score scores on the COVID-19 class.

In addition, DTDF-HFF has performed well in terms of the sensitivity on different classes. Table VII presents the

TABLE VI  
DETAIL CLASSIFICATION PERFORMANCES OF EACH CLASS OF THE PROPOSED METHOD AND THE OTHER SOTA APPROACHES

Methods	Categories	Preci-sion	Recall	F1 score	Specifi-city	Youden Index
Resnet34	Normal	86.55	72	76.74	95.63	68.63
	Bacteria & virus	76.29	89	88.41	92.09	81.09
	TB	93.58	<b>100</b>	96.52	96.75	96.75
	COVID-19	<b>98</b>	94.28	93.39	97.3	91.57
Covid CXR	Normal	87.41	80	83.86	94.89	74.89
	Bacteria & virus	83.16	87	84.07	92.51	79.52
	TB	96.6	<b>100</b>	<b>98.18</b>	99.33	<b>97.33</b>
	COVID-19	96.39	94.36	94.82	<b>98</b>	92.364
Covid-classifier	Normal	88.89	80	84.21	85.71	60.13
	Bacteria & virus	80	80	80	<b>95.23</b>	71.04
	TB	<b>100</b>	71.4	83.33	71.42	95.34
	COVID-19	75	<b>100</b>	85.71	76.19	77.60
DTDF-HFF	Normal	<b>96.89</b>	<b>85.82</b>	<b>90.57</b>	<b>98.58</b>	84.41
	Bacteria & virus	<b>91.27</b>	<b>93.09</b>	<b>91.91</b>	95.94	<b>89.0</b>
	TB	99.09	97.14	97.85	<b>99.84</b>	96.98
	COVID-19	92.8	98.56	<b>95.37</b>	96.60	<b>95.16</b>

TABLE VII  
COMPARISON OF SENSITIVITY OF COVIDCXR AND DTDF-HFF

Methods	Normal	Bacteria&virus	TB	COVID-19
CovidCXR	0.8	0.87	<b>1</b>	0.94
DTDF-HFF	<b>0.86</b>	<b>0.93</b>	0.97	<b>0.99</b>

TABLE VIII  
COMPARISON OF RUNNING TIME OF COVIDCXR AND DTDF-HFF

Methods	Feature extraction	Classification
CovidCXR	—	56 m52 s
DTDF-HFF	15 m36 s	<b>7 m94 s</b>

comparison of sensitivity between CovidCXR and DTDF-HFF. Although the sensitivity to the TB class of our model is slightly lower than that of CovidCXR, our method has significantly improved the sensitivity to the other three classes when compared to CovidCXR.

Since our proposed method is a decision tree-based architecture without the backpropagation algorithm, our method spends less than half the time required by CovidCXR (Table VIII). In short, our method has achieved excellent results on the public CXR datasets, when compared to other SOTA methods. The structure of our proposed method can not only bring advantages in reducing computational cost but also guarantee the high performance on small datasets.

#### G. Clinical Application

We are conducting a case study of deploying the proposed framework for the assessment and detection of pneumonia in clinical applications. Fig. 9 presents a typical clinical application scenario of this study. First, the proposed framework is trained with well-curated data. Then, the trained model is

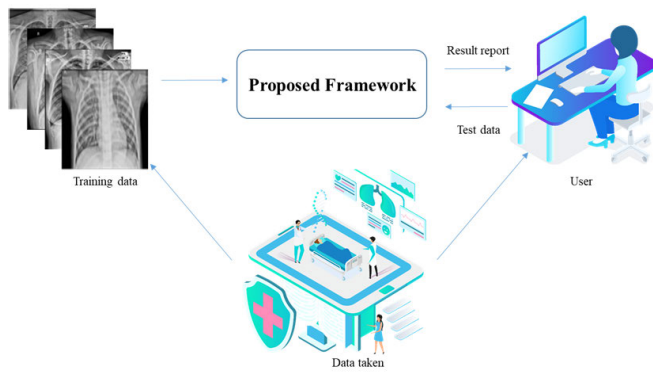


Fig. 9. Typical practical application scenario of this study for the assessment and detection of pneumonia in clinical application.

deployed in hospitals using a web server remotely. The new X-ray images obtained from the patients will be sent to the server and input into the trained model on the server. The model will make a prediction of the pneumonia type for the patient and send a report to the user on the client.

## V. CONCLUSION

In this article, a new method named DTDF-HFF has been proposed for accurate CXR image classification, which extends the structure of the original DF model to make better use of fusion feature information. The hybrid features are input into the proposed method to increase the diversity of feature information among different forests in the model. In our framework, in order to take full advantage of output information at each layer, the output vectors are transformed through the self-adaptive method and combined with the original input characteristics to create enhanced features. The performance of our method is verified on the public available datasets. And the experimental results confirm that our DTDF-HFF can achieve superior performance to other SOTA deep learning methods. Furthermore, unlike the neural network-based deep learning techniques, our method requires much less computational cost.

Generally, our method can achieve high accuracy in the case of small training samples. If another epidemic pneumonia breaks out next time, it would be quite difficult to collect a large number of well-curated data in a short time to train DNN-based models. However, our framework can be easily extended to the new classification tasks in the case of limited training data.

Although our proposed method has shown excellent performance on CXR image classification, it should be noted that the overall performance of the model could be slightly affected by the dataset variance. Further work can be done to enhance the applicability of the framework. First, the hierarchical structure of DTDF-HFF is inspired by the DNN, but there is no strategy in place to utilize errors for the modification of the generated structure, which may limit the performance of our model. Therefore, our future work will focus on designing new strategies to feed errors back to the previous layer or forests, aiming to further improve the performance of the overall structure. Second, the feature size deployed in our model is relatively large, which could result in high memory

consumption. Therefore, feature reduction techniques should be employed to decrease the feature size, particularly by eliminating redundant multigrained scanning information.

## REFERENCES

- [1] WHO. (2019). *Pneumonia*. Accessed: Jul. 19, 2021. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/pneumonia>
- [2] *Who Coronavirus Disease (COVID-19) Dashboard*. Accessed: Mar. 26, 2023. <https://covid19.who.int/>
- [3] J. Li et al., "Multi-task contrastive learning for automatic CT and X-ray diagnosis of COVID-19," *Pattern Recognit.*, vol. 114, Jun. 2021, Art. no. 107848.
- [4] Y. Oh, S. Park, and J. C. Ye, "Deep learning COVID-19 features on CXR using limited training data sets," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2688–2700, Aug. 2020.
- [5] S. Minaei, R. Kafieh, M. Sonka, S. Yazdani, and G. Jamalipour Soufi, "Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101794.
- [6] L. Wang, Z. Q. Lin, and A. Wong, "COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Sci. Rep.*, vol. 10, no. 1, p. 19549, Nov. 2020.
- [7] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," *Pattern Anal. Appl.*, vol. 24, no. 3, pp. 1207–1220, Aug. 2021.
- [8] I. D. Apostolopoulos and T. A. Mpesiana, "COVID-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks," *Phys. Eng. Sci. Med.*, vol. 43, no. 2, pp. 635–640, Apr. 2020.
- [9] E. El-Din Hemdan, M. A. Shouman, and M. E. Karar, "COVIDX-Net: A framework of deep learning classifiers to diagnose COVID-19 in X-ray images," 2020, *arXiv:2003.11055*.
- [10] I. D. Apostolopoulos, S. I. Aznaouridis, and M. A. Tzani, "Extracting possibly representative COVID-19 biomarkers from X-ray images with deep learning approach and image data related to pulmonary diseases," *J. Med. Biol. Eng.*, vol. 40, no. 3, pp. 462–469, May 2020.
- [11] M. Farooq and A. Hafeez, "COVID-ResNet: A deep learning framework for screening of COVID19 from radiographs," 2020, *arXiv:2003.14395*.
- [12] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, and A. Mohammadi, "COVID-CAPS: A capsule network-based framework for identification of COVID-19 cases from X-ray images," *Pattern Recognit. Lett.*, vol. 138, pp. 638–643, Oct. 2020.
- [13] Z. Lin et al., "AANet: Adaptive attention network for COVID-19 detection from chest X-ray images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4781–4792, Nov. 2021.
- [14] Z.-H. Zhou and J. Feng, "Deep forest: Towards an alternative to deep neural networks," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 3553–3559.
- [15] S. A. Korkmaz and H. Binol, "Classification of molecular structure images by using ANN, RF, LBP, HOG, and size reduction methods for early stomach cancer detection," *J. Mol. Struct.*, vol. 1156, pp. 255–263, Mar. 2018.
- [16] N. Bharanidharan, H. Rajaguru, and V. Geetha, "Performance analysis of KNN classifier with and without GLCM features in brain tumor detection," *Int. J. Innov. Technol. Exploring Eng.*, vol. 8, pp. 103–106, Dec. 2018.
- [17] K. Kumar, M. D. Thiagarajan, and M. Shanmugam, "An efficient method for brain tumor detection using texture features and SVM classifier in MR images," *Asian Pacific J. Cancer Prevention*, vol. 19, pp. 2789–2794, Oct. 2018.
- [18] S. Wan et al., "Integrated local binary pattern texture features for classification of breast tissue imaged by optical coherence microscopy," *Med. Image Anal.*, vol. 38, pp. 104–116, May 2017.
- [19] Z. Zhang, B. Chen, and Y. Luo, "A deep ensemble dynamic learning network for corona virus disease 2019 diagnosis," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Sep. 2, 2022, doi: [10.1109/TNNLS.2022.3201198](https://doi.org/10.1109/TNNLS.2022.3201198).
- [20] A. Z. Khuzani, M. Heidari, and S. A. Shariati, "COVID-classifier: An automated machine learning model to assist in the diagnosis of COVID-19 infection in chest X-ray images," *Sci. Rep.*, vol. 11, 2021, Art. no. 9887, doi: [10.1038/s41598-021-88807-2](https://doi.org/10.1038/s41598-021-88807-2).

- [21] N. Dey, Y.-D. Zhang, V. Rajinikanth, R. Pugalenth, and N. S. M. Raja, "Customized VGG19 architecture for pneumonia detection in chest X-rays," *Pattern Recognit. Lett.*, vol. 143, pp. 67–74, Mar. 2021.
- [22] O. Stephen, M. Sain, U. J. Maduh, and D.-U. Jeong, "An efficient deep learning approach to pneumonia classification in healthcare," *J. Healthcare Eng.*, vol. 2019, pp. 1–7, Mar. 2019.
- [23] M. Shorffuzzaman and M. S. Hossain, "MetaCOVID: A Siamese neural network framework with contrastive loss for n-shot diagnosis of COVID-19 patients," *Pattern Recognit.*, vol. 113, May 2021, Art. no. 107700.
- [24] M. Rahimzadeh, A. Attar, and S. M. Sakhaei, "A fully automated deep learning-based network for detecting COVID-19 from a new and large lung CT scan dataset," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102588.
- [25] N. E. M. Khalifa, M. H. N. Taha, A. E. Hassanien, and S. Elghamrawy, "Detection of coronavirus (COVID-19) associated pneumonia based on generative adversarial networks and a fine-tuned deep transfer learning model using chest X-ray dataset," 2020. [Online]. Available: <https://arxiv.org/abs/2004.01184>
- [26] C.-C. Hsu, G.-L. Chen, and M.-H. Wu, "Visual transformer with statistical test for COVID-19 classification," 2021. [Online]. Available: <https://arxiv.org/abs/2107.05334>
- [27] N. S. Punni and S. Agarwal, "Automated diagnosis of COVID-19 with limited posteroanterior chest X-ray images using fine-tuned deep neural networks," *Int. J. Speech Technol.*, vol. 51, no. 5, pp. 2689–2702, Oct. 2020.
- [28] P. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," 2017. [Online]. Available: <https://arxiv.org/abs/1711.05225>
- [29] M. Nour, Z. Cömert, and K. Polat, "A novel medical diagnosis model for COVID-19 infection detection based on deep features and Bayesian optimization," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 106580.
- [30] J. Rubin, D. Sanghavi, C. Zhao, K. Lee, A. Qadir, and M. Xu-Wilson, "Large scale automated reading of frontal and lateral chest X-rays using dual convolutional neural networks," 2018, *arXiv:1804.07839*.
- [31] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, 2017, Art. no. 162326.
- [32] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Los Alamitos, CA, USA, Jul. 2017, pp. 3462–3471.
- [33] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Comput. Biol. Med.*, vol. 121, Jun. 2020, Art. no. 103792.
- [34] J. Chen et al., "Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography," *Sci. Rep.*, vol. 10, no. 1, p. 19196, Nov. 2020.
- [35] X. Xu et al., "A deep learning system to screen novel coronavirus disease 2019 pneumonia," *Engineering*, vol. 6, no. 10, pp. 1122–1129, Oct. 2020.
- [36] S. Jadon, "COVID-19 detection from scarce chest X-ray image data using few-shot deep learning approach," in *Medical Imaging 2021: Imaging Informatics for Healthcare, Research, and Applications*. SPIE, Feb. 2021.
- [37] A. Chaddad, L. Hassan, and C. Desrosiers, "Deep radiomic analysis for predicting coronavirus disease 2019 in computerized tomography and X-ray images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 1, pp. 3–11, Jan. 2022.
- [38] Y. E. Almalki et al., "A novel method for COVID-19 diagnosis using artificial intelligence in chest X-ray images," *Healthcare*, vol. 9, no. 5, p. 522, Apr. 2021.
- [39] S. Dong, Q. Yang, Y. Fu, M. Tian, and C. Zhuo, "RCoNet: Deformable mutual information maximization and high-order uncertainty-aware learning for robust COVID-19 detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 8, pp. 3401–3411, Aug. 2021.
- [40] M. Irfan et al., "Role of hybrid deep neural networks (HDNNs), computed tomography, and chest X-rays for the detection of COVID-19," *Int. J. Environ. Res. Public Health*, vol. 18, no. 6, p. 3056, Mar. 2021.
- [41] A. Shamsi et al., "An uncertainty-aware transfer learning-based framework for COVID-19 diagnosis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 4, pp. 1408–1417, Apr. 2021.
- [42] S. R. Nayak, D. R. Nayak, U. Sinha, V. Arora, and R. B. Pachori, "Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study," *Biomed. Signal Process. Control*, vol. 64, Feb. 2021, Art. no. 102365.
- [43] H. Hosseinzadeh, "Deep multi-view feature learning for detecting COVID-19 based on chest X-ray images," *Biomed. Signal Process. Control*, vol. 75, May 2022, Art. no. 103595.
- [44] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, and M. Chowdhury, "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-rays images," *Comput. Biol. Med.*, vol. 132, no. 2, 2021, Art. no. 104319.
- [45] S. Kugunavar and C. J. Prabhakar, "Convolutional neural networks for the diagnosis and prognosis of the coronavirus disease pandemic," *Vis. Comput. Ind., Biomed., Art.*, vol. 4, no. 1, pp. 1–14, Dec. 2021.
- [46] J. Zhang et al., "Viral pneumonia screening on chest X-rays using confidence-aware anomaly detection," *IEEE Trans. Med. Imag.*, vol. 40, no. 3, pp. 879–890, Mar. 2021.
- [47] M. Yamaç, M. Ahishali, A. Degerli, S. Kiranyaz, M. E. H. Chowdhury, and M. Gabbouj, "Convolutional sparse support estimator-based COVID-19 recognition from X-ray images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 5, pp. 1810–1820, May 2021.
- [48] A. S. Azar et al., "Lightweight method for the rapid diagnosis of coronavirus disease 2019 from chest X-ray images using deep learning technique," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. (NSS/MIC)*, Oct. 2021, pp. 1–5.
- [49] X. Cao, R. Li, L. Wen, J. Feng, and L. Jiao, "Deep multiple feature fusion for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 10, pp. 3880–3891, Oct. 2018.
- [50] R. Su, X. Liu, L. Wei, and Q. Zou, "Deep-resp-forest: A deep forest model to predict anti-cancer drug response," *Methods*, vol. 166, pp. 91–102, Aug. 2019.
- [51] L. Yang, X.-Z. Wu, Y. Jiang, and Z.-H. Zhou, "Multi-label learning with deep forest," 2019, *arXiv:1911.06557*.
- [52] Y. Zhu, S. Fu, S. Yang, P. Liang, and Y. Tan, "Weighted deep forest for schizophrenia data classification," *IEEE Access*, vol. 8, pp. 62698–62705, 2020.
- [53] L. Sun et al., "Adaptive feature selection guided deep forest for COVID-19 classification with chest CT," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 10, pp. 2798–2805, Oct. 2020.
- [54] O. Guehairia, A. Ouamane, F. Dornaika, and A. Taleb-Ahmed, "Feature fusion via deep random forest for facial age estimation," *Neural Netw.*, vol. 130, pp. 238–252, Oct. 2020.
- [55] X. Zeng et al., "Network-based prediction of drug–target interactions using an arbitrary-order proximity embedded deep forest," *Bioinformatics*, vol. 36, no. 9, pp. 2805–2812, May 2020.
- [56] C. Ma, Z. Liu, Z. Cao, W. Song, J. Zhang, and W. Zeng, "Cost-sensitive deep forest for price prediction," *Pattern Recognit.*, vol. 107, Nov. 2020, Art. no. 107499.
- [57] Z. Jia, Z. Liu, Y. Gan, C. Vong, and M. Pecht, "A deep forest-based fault diagnosis scheme for electronics-rich analog circuit systems," *IEEE Trans. Ind. Electron.*, vol. 68, no. 10, pp. 10087–10096, Oct. 2021.
- [58] S. Jegou, M. Drozdal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers Tiramisu: Fully convolutional DenseNets for semantic segmentation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2017, pp. 1175–1183.
- [59] J. Kook Kim and H. Wook Park, "Statistical textural features for detection of microcalcifications in digitized mammograms," *IEEE Trans. Med. Imag.*, vol. 18, no. 3, pp. 231–238, Mar. 1999.
- [60] J. Shiraishi et al., "Development of a digital image database for chest radiographs with and without a lung nodule," *Amer. J. Roentgenology*, vol. 174, no. 1, pp. 71–74, Jan. 2000.
- [61] B. van Ginneken, M. B. Stegmann, and M. Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: A comparative study on a public database," *Med. Image Anal.*, vol. 10, no. 1, pp. 19–40, Feb. 2006.
- [62] S. Jaeger, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quant. Imag. Med. Surg.*, vol. 4, no. 6, pp. 475–477, Dec. 2014.
- [63] Praveen. CoronaHack—Chest X-Ray-Dataset. CoronaHack. Accessed: Aug. 19, 2020. [Online]. Available: <https://www.kaggle.com/praveengovi/coronahack-chest-xraydataset>
- [64] J. P. Cohen, P. Morrison, and L. Dao, "COVID-19 image data collection," 2020. [Online]. Available: <https://arxiv.org/abs/2003.11597>



**Qingqi Hong** (Member, IEEE) received the Ph.D. degree in computer science from the University of Hull, Hull, U.K., in 2012.

He is currently an Associate Professor with Xiamen University, Xiamen, China, and an Associate Research Scientist with the Hong Kong Centre for Cerebro-Cardiovascular Health Engineering (COCHE), Hong Kong. His main research interests include medical imaging processing, deep learning, machine learning, computer vision, and computer graphics.



**Junfeng Yao** received the B.S. and Ph.D. degrees from Central South University, Changsha, China, in 1995 and 2001, respectively.

He is currently a Professor with the School of Film and School of Informatics, Xiamen University, Xiamen, China. His research interests include computer graphics, virtual reality, intelligent algorithm research, and industrial process simulation (for more information, visit <https://cdmc.xmu.edu.cn/info/1010/1062.htm>).



**Lingli Lin** received the bachelor's degree in information management and information system from Wuhan University, Wuhan, China, in 2019, and the master's degree in software engineering from Xiamen University, Xiamen, China, in 2022.

Her research interests include medical image analysis, deep forest, and feature extraction.



**Qingqiang Wu** received the Ph.D. degree from the Chinese Academy of Sciences, Beijing, China, in 2008.

He is currently a Professor with the School of Informatics and School of Film, Xiamen University, Xiamen, China, focusing on information science, artificial intelligence, knowledge services, big data analysis, and quantitative venture investment. He has published more than 40 papers in international journals/conferences.



**Zihan Li** received the bachelor's degree from Xiamen University, Xiamen, China, in 2022, and the master's degree from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 2023. He is currently pursuing the Ph.D. degree with the University of Washington, Seattle, WA, USA.

His research interests include computer vision, self-supervised learning, medical image analysis, and vision-language model.



**Kunhong Liu** (Member, IEEE) received the B.Sc. and M.Sc. degrees from Fujian Normal University, Fuzhou, China, in 1999 and 2004, respectively, and the Ph.D. degree from the University of Science and Technology of China, Hefei, China, in 2008.

He served as the Post-Doctoral Research Fellow at the Department of Computing, Hong Kong Polytechnic University, Hong Kong, from 2014 to 2015. From 2021, he works as the Head of the Department of Digital Media Technology. He is a Professor with Xiamen University, Xiamen, China. His main

research interests include machine learning, data mining and evolutionary algorithm. In these areas he has published more than 100 peer-reviewed journal and conference papers.



**Qingde Li** was a Professor and the Deputy Head of Mathematics and Computer Science at Anhui Normal University, Wuhu, China. He has been a Lecturer in computer science at the University of Hull, Hull, U.K., since 2001. His research findings have been published in the world's most prestigious graphics journals including ACM TRANSACTIONS ON GRAPHICS, IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, and IEEE TRANSACTIONS ON MEDICAL IMAGING.



**Jie Tian** (Fellow, IEEE) received the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1993.

He is currently the Chief Scientist of the CAS Key Laboratory of Molecular Imaging, Beijing Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, China, and also with the Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Engineering Medicine, Beihang University, Beijing. His research interests include optical multi-

modality molecular imaging technology development and artificial intelligence (AI) in radiomics.