



A distance transformation deep forest framework with hybrid-feature fusion for CXR image classification

Journal:	<i>IEEE Transactions on Neural Networks and Learning Systems</i>
Manuscript ID	TNNLS-2021-P-19189.R2
Manuscript Type:	Regular Paper
Date Submitted by the Author:	19-Oct-2022
Complete List of Authors:	Hong, Qingqi; Xiamen University Lin, Lingli; Xiamen University Li, Zihan; Xiamen University Li, Qingde; University of Hull Yao, Junfeng; Xiamen University Wu, Qingqiang; Xiamen University Liu, Kunhong; Xiamen University Tian, Jie; Chinese Academy of Sciences Institute of Automation
Keywords:	Image Classification, Chest X-Ray, COVID-19, Hybrid Feature Fusion

SCHOLARONE™
Manuscripts

A distance transformation deep forest framework with hybrid-feature fusion for CXR image classification

Qingqi Hong, Lingli Lin, Zihan Li, Qingde Li, Junfeng Yao, Qingqiang Wu, Kunhong Liu, and Jie Tian, *Fellow, IEEE*

Abstract—Detecting pneumonia, especially 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 network for CXR image classification is limited due to the small sample size of the well-curated data. To tackle this problem, this paper proposes a distance transformation-based deep forest framework with the hybrid-feature fusion (DTDF-HFF) for accurate CXR image classification. In our proposed method, hybrid features of CXR images are extracted by two ways: hand-crafted feature extraction and multi-grained scanning. Different types of features are fed into different classifiers in the same layer of the deep forest, 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 the benefits from the new layer. We compare the proposed method with other methods on the public CXR data sets, and the experimental results show that the proposed method can achieve state-of-the-art performance. The code will be made publicly available at <https://github.com/hongqq/DTDF-HFF>.

Index Terms—Classification, Chest X-Ray, COVID-19, Deep Forest, 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. Especially 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 September 30, 2022, in the worldwide, more than 614 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.

Qingqi Hong, Lingli Lin, Zihan Li, Junfeng Yao, Qingqiang Wu, and Kunhong Liu are with Xiamen University, China (e-mail: hongqq@xmu.edu.cn; wuqq@xmu.edu.cn; lkhqz@xmu.edu.cn)

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

Jie Tian is with the Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China (e-mail: tian@ieee.org).

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 many very subtle and similar features with low sensitivity, which makes it 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, it 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]–[12]. 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, it in most time requires a large number of samples for training and consumes 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) [13] 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 data sets. Therefore, under the condition of limited training data sets, this paper develops an improved DF framework based on mixed feature fusion and distance transformation, i.e. 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 multi-grained 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 paper 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 multi-grained 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 feed to the corresponding classifiers at the next level. Cascading forests are constructed in this way until they converge. The outputs of the final layer is 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 paper are as follows:

1. A new deep forest framework (DTDF-HFF) based on hybrid-feature fusion and distance transformation is proposed for 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 multi-resolution image features to avoid global information loss under 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 distance transformation 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 convergency.

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 [14]–[17].

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. [18] 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. [19] 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 the strong representation learning ability via training multi-layer artificial neural networks [8], [10], [20]–[24]. Punn et al. [25] proposed a loss function approach for transfer learning on binary classification of posteroanterior CXR images. Rajpurkar et al. [26] developed a deep learning network termed as Chexnet based on ChestX-ray14 for pneumonia detection. The model achieved state of the art results on all 14 diseases in Chest X-rays. Nour et al. [27] proposed deep convolutional neural network (CNN) model which was trained from scratch to extract discriminative features on CXR images. Rubin et al. [28] proposed a novel DuaNet for processing both frontal and lateral CXR images. Their model improved performance in recognizing findings in CXR images. Lakhani et al. [29] built a model that integrated the AlexNet and GoogleNet for detecting tuberculosis in CXR. On account of the COVID-19 pandemic, more researchers have been applying deep learning to pneumonia diagnosis. Ozturk et al. [30] presented the DarkCovidNet for automatic COVID-19 detection in CXR images. The model achieved 87.02% of multi-type accuracy and 98.08% of two-type accuracy respectively. Chen et al. [31] designed a system based on CT for diagnosis of COVID-19 pneumonia. The model achieved 95.24% accuracy on each patient and 98.85% accuracy on each image. Xu et al. [32] established a deep learning models for the early screening of COVID-19 on CT datasets. Wang et al. [33] developed a deep convolutional neural network—COVID-Net for the detection of COVID-19 cases in CXR images. Yujin et al. [4] proposed a patch-based convolutional neural network approach with a relatively small number of trainable parameters for COVID-19 diagnosis. Wang et al. [34] utilized natural language processing method to mine disease image tags, demonstrating that these

common chest diseases can be detected by a unified weakly supervised multi-label image classification and disease localization framework. In Wang et al.'s work [33], the Covid-Net, an open source deep convolutional neural network platform, was proposed for detecting chest radiographs of COVID-19 cases. Oh et al. [4] proposed a patch-based convolutional neural network using relatively few parameters for COVID-19 diagnosis. Ozturk et al. [35] proposed a model for 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 [37] et al. 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 the COVID-19 patients from CXR images. Irfan et al. [39] proposed a hybrid deep neural network (HDNNs) to predict the risk of developing COVID-19 in patients using computed tomography (CT) and X-ray imaging.

Despite recent works in literatures [38]–[42] report a good 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 computing resources and training samples. Currently, collecting sufficient data samples with proper annotations is still one of the main challenges for effective training of deep neural network models.

B. Deep forest

Inspired by the layer to layer of DNNs which abstract information layer by layer, Zhou and Feng [13] 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 the diversity. Although the structures similar to DNNs, DF does not rely on back 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 multi-feature fusion for hyperspectral image feature fusion and classification [43], the Deep-Resp-Forest for the classification of the anti-cancer drug response [44], the Multi-Label based Deep Forest (MLDF) for the multi-label problems [45], the weighted Deep Forest for Schizophrenia data classification [46], and the adaptive feature selection guided Deep Forest for COVID-19 Classification with Chest CT [47]. These studies have demonstrated that DF can well handle a set of challenges in a variety of fields [48]–[51]. Given the limited medical image data sets, this study tries to develope 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 firstly preprocessed and segmented by FC-DenseNet [52] to generate the mask of lung regions, which is subsequently processed by two feature

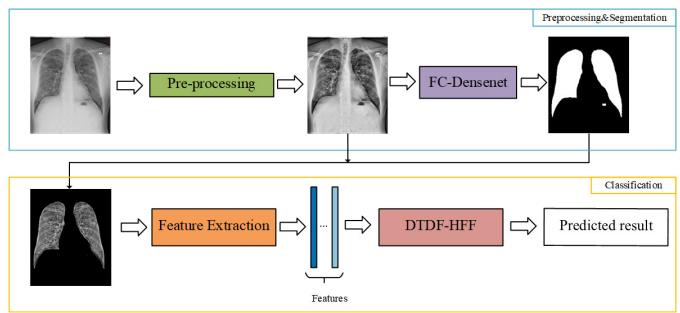


Fig. 1. Overall architecture of the proposed method. (1) The CXR images are preprocessed and segmented to generate the mask of lung regions. (2) Multiple types of features are extracted and input into our DTDF-HFF for classification.

extraction schemes: hand-crafted feature extraction and multi-grained scanning. These features are fed to the distance transformation DF with hybrid-feature fusion (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 collected from different datasets under various conditions, the data sets 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.

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 [48]. The common hand-crafted features obtained from discrete wavelet transform (DWT) and gray level difference method (GLDM), as well as multi-grained 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.

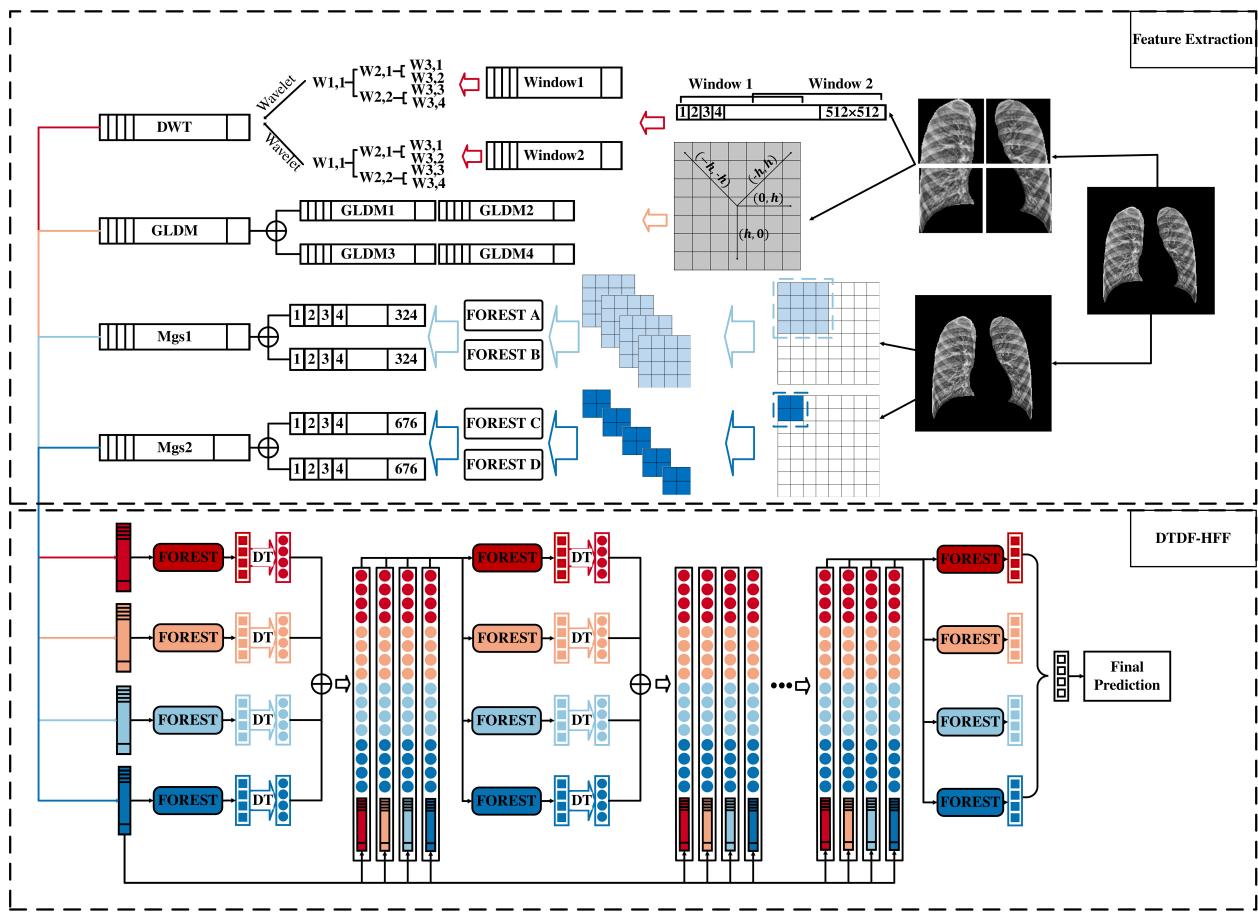


Fig. 2. The structure of our DTDF-HFF. (1) We extract hand-crafted features respectively for the upper and lower parts of the left and right lungs and generate multi-grained scanning features for the feature re-representation. (2) Different colors represent different input features. (3) Distance transformation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

The features of the low frequency components, horizontal high frequency components, vertical high frequency components and diagonal high frequency components are obtained by two-dimensional 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: $a = a_0^\alpha, a_0 > 0, \alpha \in Z; b = \beta b_0 a_0^\alpha, b \in R, n \in Z$. More specifically, for each pair of (α, β) , 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

$$\text{DWT}(\alpha, \beta) = \int_R f(r) \psi_{\alpha,\beta}(r) dr \quad (3)$$

The features based on the GLDM are block-based localized features, which are derived from the P function [53]. 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:

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

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

GLDM features are extracted from 4 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-dimension features are extracted from DWT and 224-dimensional features are extracted from GLDM, and they are concatenated to form a 672-dimension vector for each sample.

1
2 (2) *Multi-grained scanning features (Mgs)*

3 Besides the extracted hand-crafted features, the sliding
4 window scanning is deployed for feature re-representation. As
5 shown in the bottom half of Fig. 1, the images are process by
6 the sliding windows to input Extra-Trees (ETs). The feature
7 dimension G generated for each ET is:

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

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

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

14 *C. Distance transformation scheme*

15 The distance transformation (DT) scheme is designed to
16 enhance the discriminative ability of the DF framework, as
17 shown in Fig. 2.

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

26 After the preprocessing, there would still exist noise in the
27 image data, so that the probabilities corresponding to different
28 classes could still be biased towards the major classes. Fur-
29 thermore, the inconspicuous lesions as well as characterization
30 of different diseases may have similar appearance on some
31 regions of the samples. As a result, the probability values
32 of two classes in the prediction vector may not be of great
33 difference. When the output probabilities of different classes
34 of the model are close, the diseases represented by the images
35 may be misclassified.

36 To optimize the information from the probability vector,
37 the confusion matrix is deployed to measure the correlation
38 strength of various classes by a self-adaptive method. The
39 confusion matrix reflects the relationship between the ground-
40 truth labels and the prediction results. Assume that for the
41 classification task of C classes, the identification data set S
42 contains n samples. Let $cm_{i,j}$ represent the number of samples of
43 class i judged by the m_{th} ET as class j . Then, the confusion
44 matrix CM of ET m is a $C \times C$ matrix:

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

56 The main steps of the distance vector mapping are given as
57 follows:

$$V^n = \frac{1}{M} \sum_{m=1}^M (v_{m,l}^n) \quad (7)$$

58 **Algorithm 1** Distance transformation deep forest with hybrid-
59 feature fusion (DTDF-HFF)

```

60 1:  $V_l^{X_{Tr}} = \emptyset, V_l^{X_{Test}} = \emptyset$ 
61 2: //Training stage
62 3: for  $l = 1 : L$  do
63 4:   if  $i = 1$  then
64 5:     let  $X_{Tr} = S_{Tr}$ 
65 6:   else
66 7:     let  $X_{Tr} = concat(S_{Tr}, V_{l-1}^{X_{tr}})$ 
67 8:   end if
68 9:   for each  $m = 1 : M$  do
69 10:    train  $ET(m)$  using  $X_{Tr}$  to get the  $V_{m,l}^{X_{Tr}}$  and
70 11:     $CM(X_{Tr}, m)$ 
71 12:    calculate  $D_{m,l}^{X_{Tr}}$  by formula 9
72 13:   end for
73 14:    $V_l^{X_{tr}} = concat(D_{1,l}^{X_{Tr}}, \dots, D_{m,l}^{X_{Tr}})$ 
74 15: end for
75 16: //Testing stage
76 17: for  $l = 1 : L - 1$  do
77 18:   if  $i = 1$  then
78 19:     let  $X_{Test} = S_{Test}$ 
79 20:   else
80 21:     let  $X_{Test} = concat(S_{Test}, V_{l-1}^{X_{test}})$ 
81 22:   end if
82 23:   for each  $m = 1 : M$  do
83 24:     feed  $X_{Test}$  into  $ET(m)$  to get the  $V_l^{X_{Test}}$  and
84 25:      $CM(X_{Tr}, m)$ 
85 26:     calculate  $D_{m,l}^{X_{test}}$  by formula 9
86 27:   end for
87 28:    $V_l^{X_{test}} = concat(D_{1,l}^{X_{Test}}, \dots, D_{m,l}^{X_{Test}})$ 
88 29: end for
89 30: if  $l = L$  then
90 31:   for each  $m = 1 : M$  do
91 32:     feed  $X_{Test}$  into  $ET(m)$  to get  $v_{m,L}^{X_{Test}}$ 
92 33:   end for
93 34:   calculate  $V^{X_{Test}}$  by formula 7
94 35:    $Y_{Test} = argmax(V^{X_{Test}})$ 
95 36: end if

```

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:

$$v_{m,l}^n = \frac{1}{T} \sum_{t=1}^T p_{m,l}^{n,t} \quad (8)$$

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:

$$d_{m,l}^{n,c} = \sqrt{\sum_{k=1}^C (v_{m,l}^{n,k} - cm_{c,k})^2} \quad (9)$$

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 distance transformation 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:

$$X_{m,l}^n = (S_m^n, V_l^n) \quad (10)$$

, 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:

The data set 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 of the model is calculated to determine whether the cascade forest continues to grow. At the test phase, the outputs of the last layer are combined to form the output vectors, and an unknown sample 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 influenzae and virus pneumonia are the common causes of CAP. Even in the 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 [54] and SCR [55] 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 data set sources, such as NLM [56], CoronaHack [57], and Cohen et al. [58], which are fully open to any research community. These images have different dimensions and are resized to 1024

TABLE I
THE 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	—

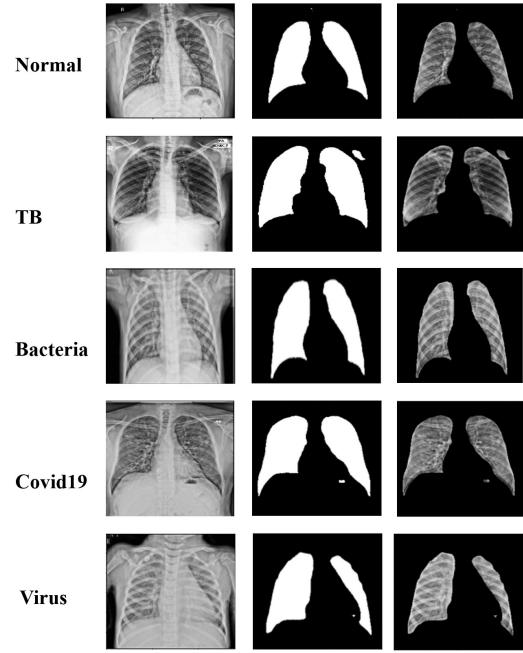


Fig. 3. An example of the preprocessed images (left), together with their corresponding segmentation results (middle), and the extracted lung areas (right) for different categorical classes of CXR images.

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 data set is divided into training set and test set, and test 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 data sets. The other data sets 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

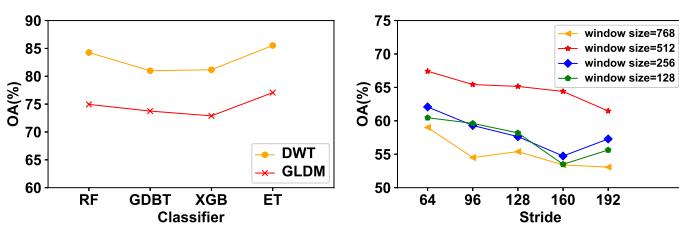


Fig. 4. (Left) The accuracy of GLDM and DWT on different base classifiers. (Right) The accuracy of different sliding window sizes on ET.

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 by follows:

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

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

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

$$4) F1score = 2(PrecisionRecall) / (Precision + Recall) \quad (14)$$

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

$$6) YoudenIndex = Recall + Specificity - 1 \quad (16)$$

The experimental environments in this paper are as follows: DTDF-HFF and other decision tree-based methods run on the 32 Intel(R) Xeon(R) CPU E5-2665 0 @ 2.40GHz, 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).

As for the multi-grained scanning, the window size may affect the generation of features, and the accuracy is observed by using ET to determine the relevant parameters. Two multi-grained 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 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

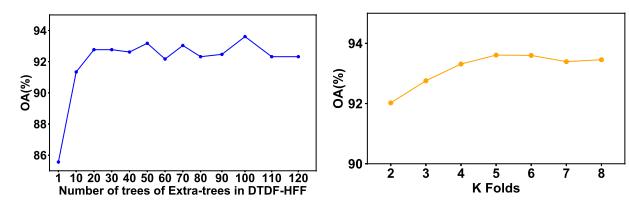


Fig. 5. (Left) The impact of the number of ETs on the performance of DTDF-HFF. (Right) K-value of the k fold cross validation of ET in DTDF-HFF.

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

types of features are set to 512, 256 respectively, and the stride sizes are both set as 64. The corresponding features dimensions generated by multi-granularity 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 sliding window is kind of memory consuming. However, it is not necessary to use very small stride size of sliding window for high resolution image, which has 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 increasing 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 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 five-fold 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

TABLE III
THE 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	73.05
Mgs2	61.73	60.63	59.89	61.57	64.66	65.25
DWT	84.26	80.99	81.17	85.53	86.48	86.91
GLDM	74.95	73.75	72.89	75.98	77.06	81.02
Mgs1+Mgs2 +DWT+GLDM	86.64	86.81	86.51	88.37	88.49	93.61

TABLE IV
CLASSIFICATION RESULTS OF DIFFERENT DECISION TREE-BASED
METHODS

Method	Accuracy	Precision	Recall	F1 score	Sensitivity	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	93.61	95.01	93.66	93.93	97.74	91.40

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 five-fold 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 5 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 multi-grained 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 accord 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 showed 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 improve its performance. To some extent, the

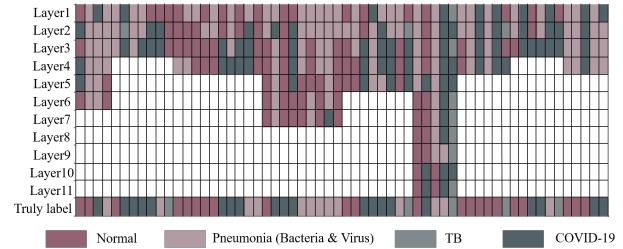


Fig. 6. The typical training results in different layer of DTDF-HFF without the distance transformation.(The colored area represents the predicted result of the corresponding layer sample, and the white area represents the stop of the model.)

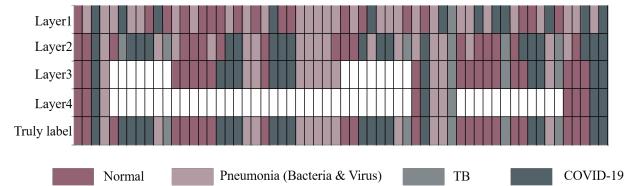


Fig. 7. The typical training results in different layers of DTDF-HFF with the distance transformation.(The colored area represents the predicted result of the corresponding layer sample, and the white area represents the stop of the model.)

diversity of these features maintain 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 distance transformation 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 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 distance transformation 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 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 perfor-

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	93.61	95.01	93.66	93.93	97.74	91.40

TABLE VI
THE DETAILED 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	100	96.52	96.75	96.75
	COVID-19	98	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	100	98.18	99.33	97.33
	COVID-19	96.39	94.36	94.82	98	92.364
Covid-classifier	Normal	88.89	80	84.21	85.71	60.13
	Bacteria &virus	80	80	80	95.23	71.04
	TB	100	71.4	83.33	71.42	95.34
	COVID-19	75	100	85.71	76.19	77.60
DTDF-HFF	Normal	96.89	85.82	90.57	98.58	84.41
	Bacteria &virus	91.27	93.09	91.91	95.94	89.0
	TB	99.09	97.14	97.85	99.84	96.98
	COVID-19	92.8	98.56	95.37	96.60	95.16

mances of the proposed method and the other SOTA approaches on the above data sets 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 [18], 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 method 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 detail 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-

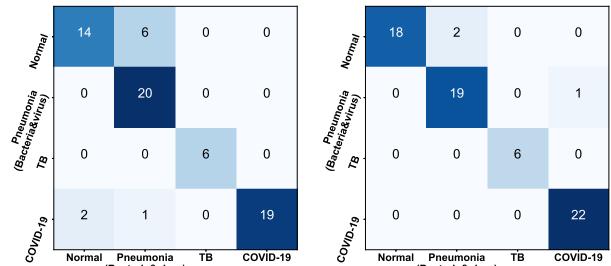


Fig. 8. Confusion matrices for CovidCXR (Left) and DTDF-HFF (Right).

TABLE VII
COMPARISON OF SENSITIVITY OF COVIDCXR AND DTDF-HFF

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

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 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 back propagation 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 data sets, 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 the small datasets.

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

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

G. Clinical application

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

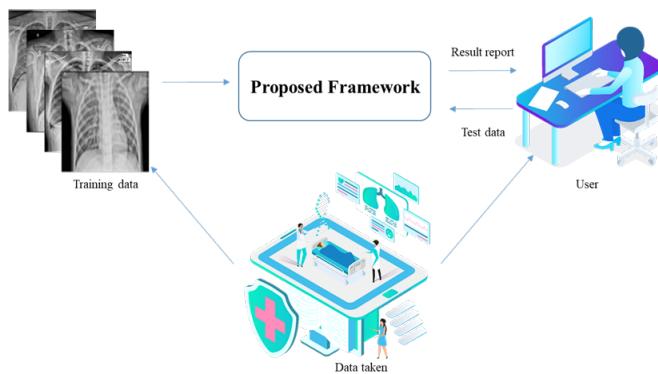


Fig. 9. A typical clinical application scenario of this study for the assessment and detection of pneumonia in clinical application.

V. CONCLUSION

In this paper, 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 data sets. 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 achieved good performance on CXR images classification, there is some research work remaining to be done in the future. First, the feature size deployed in our model is still relatively large, which may result in kind of memory consuming. Hence, the feature reduction techniques can be employed to reduce the feature size, especially the redundancy of the multi-grained scanning information. Secondly, the hierarchical structure of DTDF-HFF is inspired by the deep neural network, but there is no strategy to utilize the error for the modification of the generated structure, which may limit the performance of our model. Therefore, our future work will focus on the design of new strategies for feeding the errors to the previous layer or forests, aiming to improve the performance of the overall structure.

REFERENCES

- [1] WHO, "Pneumonia," <https://www.who.int/news-room/fact-sheets/detail/pneumonia>, 2019, accessed: 2021-07-19.

- [2] ———, "Who coronavirus disease (covid-19) dashboard," <https://covid19.who.int/>, 2022, accessed: 2021-04-08.
- [3] J. Li, G. Zhao, Y. Tao, P. Zhai, and T. Cai, "Multi-task contrastive learning for automatic ct and x-ray diagnosis of covid-19," *Pattern Recognition*, p. 107848, 2021.
- [4] Y. Oh, S. Park, and J. C. Ye, "Deep learning covid-19 features on cxr using limited training data sets," *Ieee Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2688–2700, 2020.
- [5] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. J. Soufi, "Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning," *Medical Image Analysis*, vol. 65, 2020.
- [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," *Scientific Reports*, vol. 10, no. 1, 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," 2020.
- [8] I. D. Apostopoulos and T. A. Mpesiana, "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 2, p. 635–640, Apr 2020.
- [9] E. E.-D. 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.
- [10] I. D. Apostopoulos, 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," *Journal of Medical and Biological Engineering*, vol. 40, no. 3, p. 462–469, May 2020.
- [11] M. Farooq and A. Hafeez, *COVID-ResNet: A Deep Learning Framework for Screening of COVID19 from Radiographs*, 2020.
- [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 Recognition Letters*, vol. 138, pp. 638–643, 2020.
- [13] Z.-H. Zhou and J. Feng, "Deep forest: Towards an alternative to deep neural networks.," *arXiv e-prints*, p. arXiv:1702.08835, Feb. 2017.
- [14] S. Korkmaz and H. Binol, "Analysis of molecular structure images by using ann, rf, lbp, hog, and size reduction methods for early stomach cancer detection," *Journal of Molecular Structure*, vol. 1156, 2017.
- [15] N. Bharanidharan, H. Rajaguru, and V. Geetha, "Performance analysis of knn classifier with and without glcm features in brain tumor detection," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, pp. 103–106, 2018.
- [16] K. K. M. D. Thiagarajan, and M. Shanmugam, "An efficient method for brain tumor detection using texture features and svm classifier in mr images," *Asian Pacific Journal of Cancer Prevention : APJCP*, vol. 19, pp. 2789–2794, 10 2018.
- [17] S. Wan, H. C. Lee, X. Huang, T. Xu, T. Xu, X. Zeng, Z. Zhang, Y. Sheikine, J. L. Connolly, and J. G. Fujimoto, "Integrated local binary pattern texture features for classification of breast tissue imaged by optical coherence microscopy," *other*, vol. 38, 2017.
- [18] A. Zargari, M. Heidari, and A. Shariati, *COVID-Classifier: An efficient machine learning model to assist in the diagnosis of COVID-19 infection in chest x-ray images*, 2020.
- [19] 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 Recognition Letters*, vol. 143, pp. 67–74, 2021.
- [20] Okeke, Stephen, Mangal, Sain, Uchenna, Joseph, Maduh, Do-Un, and Jeong, "An efficient deep learning approach to pneumonia classification in healthcare," *Journal of Healthcare Engineering*, vol. 2019 4180949, 2019.
- [21] M. Shoruzzaman and M. S. Hossain, "Metacovid: A siamese neural network framework with contrastive loss for n-shot diagnosis of covid-19 patients," *Pattern Recognition*, vol. 113, p. 107700, 2021.
- [22] 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," *Biomedical Signal Processing and Control*, vol. 68, p. 102588, 2021.
- [23] 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.
- [24] C.-C. Hsu, G.-L. Chen, and M.-H. Wu, "Visual transformer with statistical test for covid-19 classification," 2021.

- [25] N. S. Punn and S. Agarwal, "Automated diagnosis of covid-19 with limited posteroanterior chest x-ray images using fine-tuned deep neural networks," *Applied Intelligence*, vol. 51, no. 5, p. 2689–2702, Oct 2020.
- [26] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "CheXnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," 2017.
- [27] M. Nour, Z. Cömert, and K. Polat, "A novel medical diagnosis model for covid-19 infection detection based on deep features and bayesian optimization," *Applied Soft Computing*, vol. 97, p. 106580, 2020.
- [28] 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.
- [29] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, p. 162326, 2017.
- [30] T. O. A. M. T. B., E. A. Y. C., U. B. B. D., O. Y. E., and U. R. A. F. G. H., "Automated detection of covid-19 cases using deep neural networks with x-ray images," *Computers in Biology and Medicine*, vol. 121.
- [31] J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, Q. Chen, S. Huang, M. Yang, X. Yang, S. Hu, Y. Wang, X. Hu, B. Zheng, K. Zhang, H. Wu, Z. Dong, Y. Xu, Y. Zhu, and H. Yu, "Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography," *Scientific Reports*, vol. 10, 2020.
- [32] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, L. Yu, Q. Ni, Y. Chen, J. Su, G. Lang, Y. Li, H. Zhao, J. Liu, K. Xu, L. Ruan, J. Sheng, Y. Qiu, W. Wu, T. Liang, and L. Li, "A deep learning system to screen novel coronavirus disease 2019 pneumonia," *Engineering*, vol. 6, no. 10, pp. 1122–1129, 2020.
- [33] L. Wang and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," 2020.
- [34] 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," Los Alamitos, CA, USA, pp. 3462–3471, jul 2017.
- [35] T. Ozturk, M. Talo, A. Yildirim, U. Baloglu, Ö. yildirim, and U. R. Acharya, "Automated detection of covid-19 cases using deep neural networks with x-ray images," *Computers in Biology and Medicine*, vol. 121, 2020.
- [36] S. Jadon, "Covid-19 detection from scarce chest x-ray image data using few-shot deep learning approach," *Medical Imaging 2021: Imaging Informatics for Healthcare, Research, and Applications*, 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 Transactions on Neural Networks and Learning Systems*, pp. 1–9, 2021.
- [38] Y. E. Almalki, A. Qayyum, M. Irfan, N. Haider, A. Glowacz, F. M. Alshehri, S. K. Alduraibi, K. Alshamrani, M. A. Alkalik Basha, A. Alduraibi, M. K. Saeed, and S. Rahman, "A novel method for covid-19 diagnosis using artificial intelligence in chest x-ray images," *Healthcare*, vol. 9, no. 5, p. 522, 2021. [Online]. Available: <https://www.mdpi.com/2227-9032/9/5/522>
- [39] M. Irfan, M. A. Iftikhar, S. Yasin, U. Draz, T. Ali, S. Hussain, S. Bukhari, A. S. Alwadie, S. Rahman, A. Glowacz, and F. Althobiani, "Role of hybrid deep neural networks (hdnn), computed tomography, and chest x-rays for the detection of covid-19," *Int J Environ Res Public Health*, vol. 18, no. 6, 2021.
- [40] 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," *Biomedical Signal Processing and Control*, vol. 64, p. 102365, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1746809420304717>
- [41] H. Hosseinzadeh, "Deep multi-view feature learning for detecting covid-19 based on chest x-ray images," *Biomedical Signal Processing and Control*, vol. 75, p. 103595, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1746809422001173>
- [42] 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," *Computers in Biology and Medicine*, vol. 132, no. 2, p. 104319, 2021.
- [43] X. Cao, R. Li, L. Wen, J. Feng, and L. Jiao, "Deep multiple feature fusion for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 10, pp. 3880–3891, 2018.
- [44] 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, 2019, deep Learning in Bioinformatics.
- [45] Y. Liang, X.-Z. Wu, Y. Jiang, and Z.-H. Zhou, *Multi-Label Learning with Deep Forest*, 2019.
- [46] Y. Zhu, S. Fu, S. Yang, P. Liang, and Y. Tan, "Weighted deep forest for schizophrenia data classification," *IEEE Access*, vol. 8, pp. 62 698–62 705, 2020.
- [47] L. Sun, Z. Mo, F. Yan, L. Xia, F. Shan, Z. Ding, W. Shao, F. Shi, H. Yuan, and H. Jiang, "Adaptive feature selection guided deep forest for covid-19 classification with chest ct," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, no. 99, pp. 1–1, 2020.
- [48] O. Gueharia, A. Ouamane, F. Dornaika, and A. Taleb-Ahmed, "Feature fusion via deep random forest for facial age estimation," *Neural Networks*, vol. 130, 2020.
- [49] X. Zeng, S. Zhu, Y. Hou, P. Zhang, and F. Cheng, "Network-based prediction of drug-target interactions using an arbitrary-order proximity embedded deep forest," *Bioinformatics*, vol. 36, no. 9, 2020.
- [50] C. Ma, Z. Liu, Z. Cao, W. Song, and W. Zeng, "Cost-sensitive deep forest for price prediction," *Pattern Recognition*, vol. 107, 2020.
- [51] Z. Jia, Z. Liu, Y. Gan, C. M. Vong, and M. Pecht, "A deep forest based fault diagnosis scheme for electronics-rich analog circuit systems," *IEEE Transactions on Industrial Electronics*, vol. PP, no. 99, pp. 1–1, 2020.
- [52] S. Jegou, M. Drozdzal, D. Vazquez, A. Romero, Y. Bengio, and Ieee, *The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation*, ser. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 1175–1183.
- [53] J. K. Kim and H. W. Park, "Statistical textural features for detection of microcalcifications in digitized mammograms," *IEEE Trans Med Imaging*, vol. 18, no. 3, pp. 231–238, 1999.
- [54] J. Shiraiishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K.-i. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule," *American Journal of Roentgenology*, vol. 174, no. 1, pp. 71–74, 2000.
- [55] B. V. Ginneken, M. B. Stegmann, and M. Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database," *Medical Image Analysis*, vol. 10, no. 1, pp. 19–40, 2006.
- [56] S. Jaeger, S. Candemir, S. Antani, Y. Wáng, and G. Thoma, "Two public chest x-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative Imaging in Medicine and Surgery*, vol. 4, no. 6, pp. 475–477, 2014.
- [57] P. CoronaHack, "Coronahack -chest x-ray-dataset," <https://www.kaggle.com/praveengovi/coronahack-chest-xraydataset>, 2020, accessed: 2020-08-19. [Online]. Available: <https://www.kaggle.com/praveengovi/coronahack-chest-xraydataset>
- [58] J. P. Cohen, P. Morrison, and L. Dao, "Covid-19 image data collection," 2020.

Response Letter

Note: The reviewer's comments are in italic, while our reply in plain text.

We are very grateful to the editor and reviewers for their valuable comments. They helped us in enhancing our manuscript. All the comments and issues raised have been addressed.

- Responses to Reviewer 1's comments:

1. The paper should be interesting

Response: Thank you very much for your comments.

2. block diagram of research should be added. Fig. 1 is not sufficient.

Response: Thank you for your concern and suggestion. The figure is added to illustrate the application scenario of our study.

We are conducting a case study of deploying the proposed framework for the assessment and detection of pneumonia in clinical application. Figure 1 presents a typical clinical application scenario of this study. Firstly, the proposed framework is trained with well-curated data. Then, the trained model is 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.

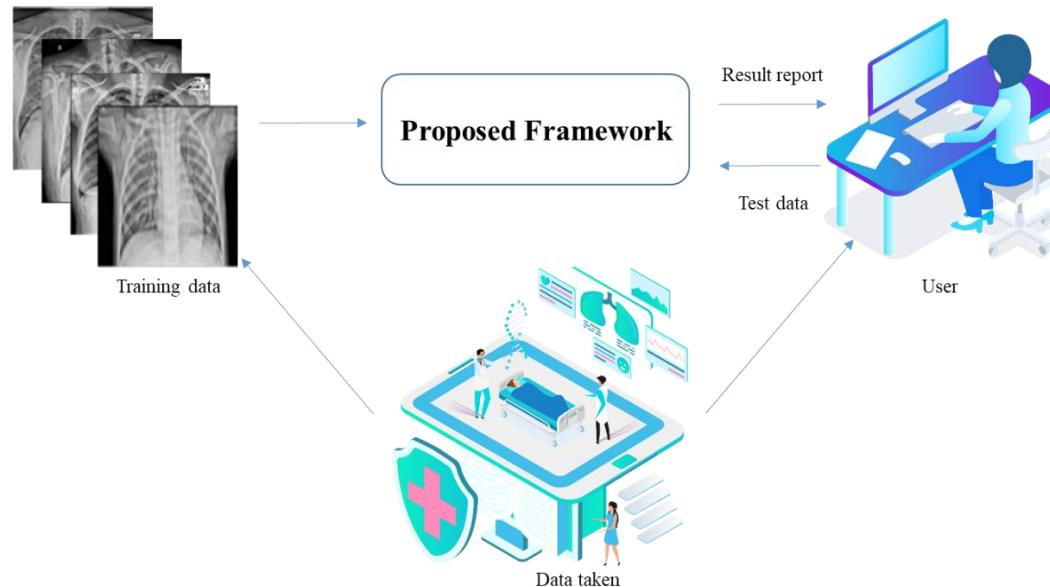


Figure 1: A typical clinical application scenario of this study for the assessment and detection of pneumonia in clinical application.

According to your suggestion, a new Subsection IV.G. concerning clinical application is added in

1
2
3
4
5
the revised manuscript.

6 3. photos of the application should be added, hospitals or something
7
8

9 **Response:** Thank you for your concern and suggestion.
10
11 As the response to the previous question, we are conducting a case study of deploying the proposed
12 framework for the assessment and detection of pneumonia in clinical application. The proposed
13 framework is on trial in the affiliated hospital of our university. As shown in the figure below, a
doctor is evaluating the diagnosis report obtained by our model.
14
15
16
17
18
19
20
21
22
23



24 Figure 2: A doctor is evaluating the diagnosis report obtained by our model.
25
26
27
28
29
30
31

32 4. references should be compared with other papers about COVID for example "Role of Hybrid
33 Deep Neural Networks (HDNNs), Computed Tomography, and Chest X-rays for the Detection of
34 COVID-19", "A Novel Method for COVID-19 Diagnosis Using Artificial Intelligence in Chest X-
35 ray Images".
36

37 **Response:** Thank you for your concern and suggestion.
38 We have cited those two papers and made a comparison in the Related Work.
39
40

Responses to Reviewer 2's comments:

41 No comments.
42
43

Responses to Reviewer 3's comments:

44 The author proposes a deep forest framework for classification of CXR images due to the small
45 sample size of the well-curated data. And positive effect appears when the CXR dataset is limited.
46 In response, most problems raised are revised well by the author. But some problems should be made
47 clear. Therefore, the recommendation of mine is reject and resubmission. My questions are listed as
48 follows:
49

50 1. In response 7, the author said when the stride size is smaller than 64, the computation is too much
51 for our machine, which will result in out of memory. However, the dataset is scale limited. Does it
52 mean the framework is memory consuming or CPU-consuming? Does the author have an estimation
53 about how the scale of dataset that the proposed work could be applied? The author should make
54 sure the limitation of the proposed work.
55

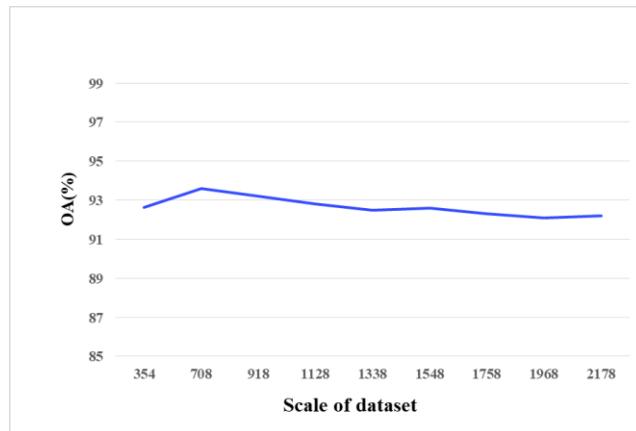
56 **Response:** Thank you for your concern and suggestion.
57
58 That's good point. This is an issue specifically associated with the computability of the computer
59 on which the experiment has been conducted. As our machine was bought 6 years ago, it is a bit too
60

1
2
3 old for this kind of task.
4
5 When the image resolution is high and the stride size is small, the operation of sliding window is
6 kind of memory consuming for our machine. However, in our experiment, we found that when the
7 image resolution is high (i.e. 1024*1024), it is not necessary to use very small stride size of sliding
8 window for feature extraction, which has little influence on the performance of the proposed
9 framework.
10
11
12 We've applied the proposed method to the dataset containing more than two thousand samples.
13 For very large dataset, the image can be subsampled to generate the sliding window feature, which
14 can greatly reduce memory consumption.
15 Of course, a direct solution to this issue would be to upgrade our machine. :)
16
17
18 In the revised manuscript, we've added this limitation of the proposed work at the end of Subsection
19 IV.C and Section V Conclusion.
20
21
22
23 2. Although, the author said "However, CT is not suitable for large-scale screening of pneumonia
24 due to the relatively high cost and high radiation dose" and "On the other hand, CXR examination
25 is simple to operate. It requires less imaging time and lower cost." By searching google scholar,
26 there are many studies conducted in CT. And some studies have several hundred CT scans. Millions
27 of patients were confirmed with COVID-19. However, the framework is proposed for limited dataset.
28 It seems that this work has limited practical value. Is it practical that the proposed work could be
29 used for predicted COVID-19? I wonder if the author could convince me that the study conducted
30 on limited X-ray is practical. I would like to see the authors comment on this.
31
32
33 **Response:** Thank you very much for your concern and suggestion.
34 As is commonly known, CT scan in general takes more than 15 minutes per patient in addition to
35 the time required for CT decontamination. In contrast, on average, CXR test takes about 15 seconds
36 per patient, which makes CXR one of the most time/cost effective assessment tools. Hence, as a
37 supplement to CT, CXR can be utilized to screen patients' initial diagnosis and confirm the priority
38 of patients' treatments, which is helpful for the saturated healthcare system in the pandemic situation.
39
40
41
42 Our method can achieve high accuracy in the case of small training samples. If another epidemic
43 pneumonia breaks out next time, it would be quite difficult to collect a large number of well-curated
44 data in a short time to train DNN-based models. However, our framework can be easily extended to
45 the new classification tasks in the case of limited training data.
46
47
48 CovidCXR [4] is a very popular and well-known method most relevant to our study. In terms of the
49 size of dataset used, the dataset used in our experiments is larger than that used in CovidCXR [4]
50 Given the popularity and practical applications of the method proposed in CovidCXR [4], the
51 research findings from our study are convincing and practical. In addition, as can be seen in the
52 response to Question 3, the proposed method has a robust performance when the scale of the dataset
53 changes.
54
55
56
57
58 We believe our study is practical. Some comments have been appended in Section V Conclusion.
59
60

1
2
3 *3. This work was proposed due to the small sample size of the well-curated data. In my opinion, if*
4 *the dataset is limited and small, the computation could be greatly reduced. Could the author give*
5 *any discussion about the influence of the scale of the dataset? Does the proposed work have a robust*
6 *performance when the scale of dataset changes?*

7
8 **Response:** Thank you for your concern and suggestion.

9
10 Yes, that's true. As shown in Table VIII, our method spends less than half the time required by
11 DNN based methods (i.e. CovidCXR) on the same dataset.
12
13
14 According to your suggestion, we enlarged the dataset and conducted an additional experiment to
15 assess how the robustness of our method may be affected by the scale of the dataset. As can be
16 seen from the figure below, the proposed method has a robust performance when the scale of the
17 dataset changes.
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33



34 Figure 3: The influence of the scale of the dataset
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60