

Adaptive Feature Selection Guided Deep Forest for COVID-19 Classification With Chest CT

Liang Sun , Zhanhao Mo , Fuhua Yan, Liming Xia, Fei Shan , Zhongxiang Ding, Bin Song, Wanchun Gao, Wei Shao , Feng Shi , Huan Yuan, Huiting Jiang, Dijia Wu , Ying Wei, Yaozong Gao, He Sui, Daoqiang Zhang , and Dinggang Shen , *Fellow, IEEE*

Abstract—Chest computed tomography (CT) becomes an effective tool to assist the diagnosis of coronavirus disease-19 (COVID-19). Due to the outbreak of COVID-19 worldwide, using the computed-aided diagnosis technique for COVID-19 classification based on CT images could largely alleviate the burden of clinicians. In this paper, we propose an Adaptive Feature Selection guided Deep Forest (AFS-DF) for COVID-19 classification based on chest CT images. Specifically, we first extract location-specific features from CT images. Then, in order to capture the high-level representation of these features with the relatively small-scale data, we leverage a deep forest model to learn high-level representation of the features. Moreover, we propose a feature selection method based on the trained deep forest model to reduce the redundancy of features, where the feature selection could be adaptively incorporated with the COVID-19 classification model. We evaluated our proposed AFS-DF on COVID-19 dataset with 1495 patients of COVID-19 and 1027 patients of community acquired pneumonia (CAP). The accuracy (ACC), sensitivity (SEN), specificity (SPE), AUC, precision and F1-score achieved by our method are 91.79%, 93.05%, 89.95%, 96.35%, 93.10% and 93.07%, respectively. Experimental results on the COVID-19 dataset suggest that the proposed AFS-DF achieves

superior performance in COVID-19 vs. CAP classification, compared with 4 widely used machine learning methods.

Index Terms—COVID-19 classification, deep forest, feature selection, chest CT.

I. INTRODUCTION

SINCE December 2019, the outbreak of coronavirus disease-19 (COVID-19) [1], [2] has infected more than 20 million people worldwide, and causing more than 800,000 deaths. World Health Organization (WHO) has declared the COVID-19 as a global health emergency on January 30, 2020 [3]. The chest computed tomography (CT) has shown to be useful to assist clinical diagnosis of COVID-19 [4]–[9]. However, the rapid growth of COVID-19 patients results in the shortage of the clinicians and radiologists. It is highly desired to develop automatic methods for computer-aided COVID-19 classification with chest CT images.

A few machine learning methods have been proposed for COVID-19 classification using chest CT images. For example, [10] employs a logistic regression method for COVID-19

Manuscript received May 7, 2020; revised July 13, 2020 and August 11, 2020; accepted August 22, 2020. Date of publication August 26, 2020; date of current version October 5, 2020. This work was supported in part by the National Key Research and Development Program of China under Grants 2018YFC2001600 and 2018YFC2001602, in part by the National Natural Science Foundation of China under Grants 61876082, 61861130366, 61732006, 61902183, and 81871337, in part by the Royal Society-Academy of Medical Sciences Newton Advanced Fellowship under Grant NAFR1180371, in part by China Postdoctoral Science Foundation funded project under Grant 2019M661831, in part by Wuhan Science and technology program under Grant 2018060401011326, in part by Hubei Provincial Novel Pneumonia Emergency Science and Technology Project under Grant 2020FCA021, in part by Huazhong University of Science and Technology Novel Coronavirus Pneumonia Emergency Science and Technology Project under Grant 2020kfyXGYJ014, and in part by the Novel Coronavirus Special Research Foundation of the Shanghai Municipal Science and Technology Commission under Grant 20441900600. (Liang Sun, Zhanhao Mo, Fuhua Yan, Liming Xia, Fei Shan, Zhongxiang Ding, Bin Song, Wanchun Gao, and Wei Shao contributed equally to this work.) (Corresponding authors: Daoqiang Zhang; Dinggang Shen.)

Liang Sun, Wei Shao, and Daoqiang Zhang are with the College of Computer Science and Technology, MIT Key Laboratory of Pattern Analysis and Machine Intelligence, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China (e-mail: sunl@nuaa.edu.cn; 527606857@qq.com; dqzhang@nuaa.edu.cn).

Zhanhao Mo and He Sui are with the Department of Radiology, China-Japan Union Hospital of Jilin University, Changchun 130021, China (e-mail: mozhanhao@jlu.edu.cn; suihe910402@126.com).

Digital Object Identifier 10.1109/JBHI.2020.3019505

Fuhua Yan is with the Department of Radiology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200240, China (e-mail: yfh11655@rjh.com.cn).

Liming Xia is with the Department of Radiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: xialiming2017@outlook.com).

Fei Shan is with the Department of Radiology, Shanghai Public Health Clinical Center, Fudan University, Shanghai 200433, China (e-mail: shanfei_2901@163.com).

Zhongxiang Ding is with the Department of Radiology, Hangzhou First Peoples Hospital, Zhejiang University School of Medicine, Hangzhou 310027, China (e-mail: hangzhoudzx73@126.com).

Bin Song is with the Department of Radiology, Sichuan University West China Hospital, Chengdu 610041, China (e-mail: anicesong@vip.sina.com).

Wanchun Gao is with the Department of Radiology, Qianjiang Central Hospital, Jishou University School of Medicine, Chongqing 409000, China (e-mail: 13908272019@163.com).

Feng Shi, Huan Yuan, Huiting Jiang, Dijia Wu, Ying Wei, Yaozong Gao, and Dinggang Shen are with the Department of Research and Development, Shanghai United Imaging Intelligence Co., Ltd., Shanghai 201399, China. He is also with Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, Republic of Korea (e-mail: feng.shi@united-imaging.com; huan.yuan@united-imaging.com; huiting.jiang@united-imaging.com; dijia.wu@united-imaging.com; ying.wei@united-imaging.com; yaozong.gao@united-imaging.com; dinggang.shen@gmail.com).

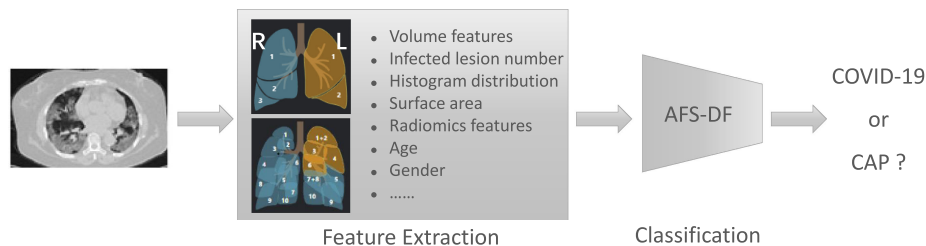


Fig. 1. Pipeline of the proposed adaptive feature selection guided deep forest for COVID-19 vs. CAP classification. We first extract the location-specific features from chest CT images. Then, the proposed AFS-DF is leveraged to train the classifier based on the location-specific features. Finally, based on the location-specific features, the trained AFS-DF is adopted for COVID-19 identification task.

classification by using clinical and laboratory features. [11], [12] use random forest model with the handcrafted features for COVID-19 classification. Moreover, some deep learning based methods are proposed for the diagnosis of COVID-19. For instance, [13] leverages a deep learning method to learn the feature representation of the chest CT image, and then uses the learned features for COVID-19 classification by combining decision tree and Adaboost algorithm. In addition, [14] employs an end-to-end network to map the CT images to label space for COVID-19 disease identification.

In summary, the existing machine learning methods for COVID-19 diagnosis are mainly based on the handcrafted features or the learned image representation by deep neural networks. However, simply adopting handcrafted features cannot fully utilize the high-level information for COVID-19 classification, while the features learned by neural networks require great effort for parameter tuning with a small amount of medical image data.

To this end, in this paper, we propose a novel adaptive feature selection guided deep forest method that takes advantage of the high-level deep features with small number of medical image data for the classification between COVID-19 and CAP. Specifically, as shown in Fig. 1, we first extract the location-specific features from the chest CT image. Then, a deep forest model [15] is introduced to learn the latent high-level representations of these features, which can effectively describe the high-level information within the extracted location-specific features by using a small-scale training data. Intuitively, the use of the feature selection could promote the performance of the classification task. Hence, in our study, we also introduce a task-driven feature selection method to adaptively reduce the redundancy of features. In particular, the feature selection operation will discard a portion of unimportant features based on the feature importance which is calculated from trained forests in each deep forest layer. Hence, the feature selection and classifier training are adaptively incorporated into a unified framework. Finally, the trained adaptive feature selection guided deep forest is used for COVID-19 prediction.

The major contributions of this paper are three-fold. *First*, a deep forest is leveraged to learn the high-level feature representation of the location-specific features from chest CT for the diagnosis of COVID-19. *Second*, we propose an adaptive feature selection method to adaptively select the discriminative features for the diagnosis of COVID-19. *Third*, the proposed method is evaluated on the collected COVID-19 dataset, which consists of 1495 COVID-19 patients and 1027 CAP patients. Experimental

TABLE I
DEMOGRAPHIC INFORMATION OF THE STUDIED 2522 CHEST CT SCANS FROM COVID-19 DATASET. M/F: MALE/FEMALE

| Category | Scan # | Gender (M/F) |
|----------|--------|--------------|
| COVID-19 | 1495 | 770/725 |
| CAP | 1027 | 488/539 |

results demonstrate that our method achieves superior classification performance than the comparison methods.

The rest of the paper is organized as follows. We first introduce the materials used in this study and the proposed adaptive feature selection guided deep forest in Section II. Then, in Section III, we present experimental settings and results. In Section IV, we study the influence of parameters in the proposed methods and present the limitations of the current study as well as possible future directions. We finally conclude this paper in Section V.

II. MATERIALS AND METHOD

In this section, we first introduce the dataset used in this study. Then, we present the feature extraction procedure for the location-specific features. Next, we describe the proposed adaptive feature selection guided deep forest (AFS-DF). Finally, we provide the implementation details for our proposed AFS-DF.

A. Materials

A total of 2522 chest CT images are used in our study, provided by China-Japan Union Hospital of Jilin University, Ruijin Hospital of Shanghai Jiao Tong University, Tongji Hospital of Huazhong University of Science and Technology, Shanghai Public Health Clinical Center of Fudan University, Hangzhou First People's Hospital of Zhejiang University, and Sichuan University West China Hospital. In this dataset, 1495 cases are from the confirmed COVID-19 cases diagnosed by positive nucleic acid testing. The other 1027 case are from CAP patients. COVID-19 images are acquired from Jan. 9, 2020 to Feb. 14, 2020, and CAP images are obtained from Jul. 30, 2018 to Feb. 22, 2020. The demographic information of these 2522 subjects is summarized in Table I.

All patients underwent chest CT scans with thin section. Specifically, CT scanners include uCT 780 from UIH, Optima CT520, Discovery CT750, LightSpeed 16 from GE, Aquilion ONE from Toshiba, SOMATOM Force from Siemens, and SCENARIA from Hitachi. CT protocol includes: 120 kV, reconstructed slice thickness ranging from 0.625 to 2 mm, with

breath hold at full inspiration. All images are de-identified before sending for analysis. The study are approved by the Institutional Review Board of participating institutes. Written informed consent is waived due to the retrospective nature of the study.

We evaluate the COVID-19 vs. CAP classification using 5-fold cross-validation on the collected chest CT images. Specifically, all subjects are randomly partitioned into 5 subsets (the first four subsets consist of 504 subjects, respectively. The last subset contains 506 subjects). Each subset is sequentially selected as the testing set (containing 504/506 subjects), while the remaining four subsets (containing 2018/2016 subjects) are treated as training set. The training set is further divided into five subsets for 5-fold cross-validation to choose the hyper-parameters in the proposed method and the competing methods. Then, we use all training data with chosen hyper-parameters to train the model. Finally, we test the trained model on the testing data (containing 504/506 subjects).

B. Feature Extraction

Similar to [12], we extract the location-specific features (i.e., infection locations and spreading patterns) to represent the chest CT images for diagnosis of COVID-19. Specifically, the chest CT images are first automatically segmented into infected lung regions and lung fields bilaterally by using VB-Net [16]. The infected lung regions are mainly related to mosaic sign, ground glass opacity (GGO), lesion-related signs (air bronchogram) and interlobular septal thickening. The lung fields include left lung, right lung, five lung lobes, and eighteen pulmonary segments. Then, we extract four kinds of location-specific handcrafted features, including volume, infected lesion number, histogram distribution and surface area from chest CT images. Meanwhile, we also extract the radiomics features for describing the CT images. More details are as follows.

- 1) **Volume features:** Based on the segmented infected lung regions, we extract the total volume of infected region, and then calculate the percentage of the infected region of the whole lung. Meanwhile, according to the lung field segmentation results, we further extract the volume and percentage in each lobe and each pulmonary segment, respectively. Since there are evidence that pneumonia caused by COVID-19 more likely occurs in both right and left lungs, we also calculate the infected lesion difference as well as the percentage difference between left and right lungs.
- 2) **Infected lesion number:** In comparison to CAP, most of the COVID-19 infections encompass bilateral lungs with multifocal involvement [6], [7], and COVID-19 generally has concentrated infection lesions while CAP shows small in volume and patchy in distribution [17]. Therefore, we calculate the features of the total number of infected regions in the bilateral lungs, lung lobes, and pulmonary segments, respectively.
- 3) **Histogram distribution:** The predominant chest CT findings show that bilateral and peripheral GGO and consolidation are a radiologic hallmark of COVID-19 [18], [19]. GGO is a pattern of hazy increased lung opacity with

preservation of bronchial and vascular margins, whereas consolidation is characterized by a homogeneous increase in lung parenchymal attenuation that obscures the margins of vessels and airway walls on CT images [17]. To extract the intensity distribution of the infected regions in chest CT images, we calculated the histogram features of the infected regions.

- 4) **Surface area:** In previous study [20], it has been found that COVID-19 had a predominate distribution in the posterior and peripheral lung, and the abnormalities of lung parenchyma eventually spread to the central area and bilateral upper lobes [6]. Therefore, we constructed the infection surface as well as the lung boundary surface. We further calculated the distance of each infection surface vertex to the nearest lung boundary surface, and categorized them into 5 ranges, i.e., 3, 6, 9, 12 and 15 voxels. For each feature, the number of infection surface vertices within each range of distances to the lung wall is calculated. Furthermore, the percentage of infection vertex number against the number of whole infection surface vertices in each range is also considered.
- 5) **Radiomics features:** Radiomics features extracted from infected lesions, including intensity features (e.g., average gray level intensity, range of gray values) and texture features (e.g., gray level co-occurrence matrix, gray-level run-length matrix, gray-level size-zone matrix, and neighborhood gray-tone difference matrix) are used in our study.

Besides, we also adopt the age and gender into the location-specific features for the diagnosis of COVID-19. In summary, a total of 239 dimensions features are used in our study. More details of these features are presented in [12].

C. Adaptive Feature Selection Guided Deep Forest

As mentioned in Section II-B, we extract location-specific features from the chest CT images. But, simply using these features cannot adequately describe the high-level information for COVID-19 classification. In this work, we propose an adaptive feature selection guided deep forest to learn the latent high-level representation of the extracted location-specific features with adaptive task-driven feature selection process for diagnosis of COVID-19. The architecture of proposed adaptive feature selection guided deep forest is shown in Fig. 2.

As shown in Fig. 2, each layer of proposed adaptive feature selection guided deep forest consists of N independent random forests and a feature selection unit. Here, each random forest produces a probability distribution of the COVID-19 and CAP (the yellow rectangle in Fig. 2). Then, the N probability distribution vectors of the COVID-19 and CAP are concatenated with the input feature vector. To reduce the redundancy of the features, we further perform an adaptive feature selection operation. In particular, for each trained random forest, we calculate the feature importance for each feature within the input feature vector. Thus, we calculate the overall feature importance c_i for i -th feature as follows,

$$c_i = \frac{1}{N} \sum_{n=1}^N c_{i,n} \quad (1)$$

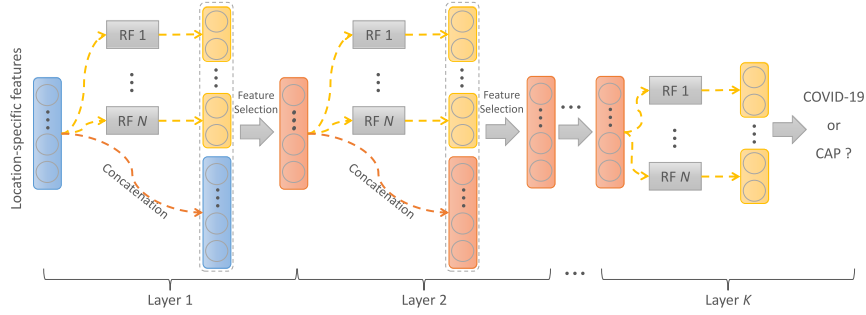


Fig. 2. Overview of the proposed adaptive feature selection guided deep forest. Each layer of the proposed adaptive feature selection guided deep forest consists of N random forests and an adaptive feature selection unit.

where $c_{i,n}$ is the feature importance for i -th feature in n -th random forest. Herein, we discard the features with low feature importance by a specific ratio based on the calculated feature importance. Hence, the feature selection and classifier training are adaptively incorporated into a unified framework. Thus, the selected feature vector as the input of the next layer. Finally, we cascade multiply layers to learn the deep discriminative feature representation for COVID-19 classification task.

D. Implementation

As shown in Fig. 2, in the proposed adaptive feature selection guided deep forest, we employ a Xgboost [21] with 20 trees, a random forest [22] with 20 trees, and two extremely randomized trees [23] with 20 and 50 trees, respectively. We empirically set the feature discard ratio as 0.2 in our study. In the *training* stage, we feed the extracted location-specific features to the adaptive feature selection guided deep forest. The training set is further divided to 5 subsets for 5-fold cross-validation. Thus, the numbers of cascade layers and selected features are automatically determined by using the cross-validation strategy. Hence, the adaptive feature selection guided deep forest can adaptively train the feature selection and classification model in a task-driven manner. Notably, the AFS-DF is trained in a layer-to-layer manner.

In the *testing* stage, we also feed the location-specific features of test subject to the trained adaptive feature selection guided deep forest model. In the last layer, each forest will produce a probability distribution p for the identification of COVID-19. For each subject, we use the following equation to ensemble the predicted value for diagnose of COVID-19,

$$p(l = c) = \frac{1}{N} \sum_{n=1}^N p(l = c|n) \quad (2)$$

where $p(l = c|n)$ is the probability of subject belongs to category c (i.e., COVID-19 or CAP) that is provided by the n -th forest in last layer. Finally, we use the MAP criterion to obtain the label for each subject, i.e., $\operatorname{argmax}_c p(l = c)$

III. EXPERIMENT

In this section, we first illustrate the competing methods and experimental settings in our study. Then, we present experimental results achieved by different methods on the chest CT images with 1495 patients of COVID-19 and 1027 patients of CAP.

A. Competing Methods

In our experiments, we compare our proposed AFS-DF with the following four widely adopted machine learning methods.

- 1) **Logistic Regression (LR)**: A Logistic regression method is employed for COVID-19 classification by using the extracted location-specific features.
- 2) **Support Vector Machine (SVM)**: The extracted location-specific features are fed into the SVM [24] classifier by using radial basis function kernel, and $C = 1.0$ via cross-validation.
- 3) **Random Forests (RF)**: The random forest classifier is applied on the location-specific features for COVID-19 classification, and the number of trees in random forest is set as 500 via cross-validation.
- 4) **Neural Networks (NN)**: In this method, a fully connected neural network is employed for COVID-19 classification. Specifically, we empirically set the mini-batch size as 64, the number of epochs as 100, and the learning rate as 0.001.

B. Experimental Settings

For all extracted location-specific features from chest CT images, we first perform the normalization with center 0 and deviation 1 for each feature. In order to measure the classification performance of different methods, six evaluation metrics are adopted, including classification accuracy (ACC), sensitivity (SEN), specificity (SPE), the Area Under the receiver operating characteristic Curve (AUC) [25], precision and F1-score. Here, the ACC, SEN, SPE, precision and F1-score are defined as,

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3)$$

$$\text{SEN} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

TABLE II

PERFORMANCE OF COVID-19 vs. CAP CLASSIFICATION ACHIEVED BY LR, SVM, RF, NN AND AFS-DF. THE TERMS a AND b IN " $a \pm b$ " DENOTE THE MEAN AND STANDARD DEVIATION, RESPECTIVELY

| Method | ACC (%) | SEN (%) | SPE (%) | AUC (%) | Precision (%) | F1-score(%) |
|--------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| LR | 89.81 \pm 1.63 | 91.64 \pm 1.69 | 87.10 \pm 2.14 | 95.52 \pm 1.31 | 91.22 \pm 1.27 | 91.42 \pm 1.37 |
| SVM | 89.97 \pm 1.37 | 91.51 \pm 2.04 | 87.70 \pm 0.89 | 95.62 \pm 0.91 | 91.38 \pm 0.60 | 91.43 \pm 1.21 |
| RF | 89.41 \pm 1.67 | 90.51 \pm 0.87 | 87.85 \pm 1.01 | 95.40 \pm 1.21 | 91.21 \pm 1.10 | 90.85 \pm 1.26 |
| NN | 89.96 \pm 1.58 | 92.66 \pm 1.57 | 86.04 \pm 1.97 | 95.73 \pm 1.00 | 90.65 \pm 1.16 | 91.64 \pm 1.22 |
| AFS-DF | 91.79 \pm 1.04 | 93.05 \pm 1.67 | 89.95 \pm 1.28 | 96.35 \pm 1.13 | 93.10 \pm 0.75 | 93.07 \pm 0.91 |

$$\text{SPE} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (5)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6)$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{SEN}}{\text{Precision} + \text{SEN}} \quad (7)$$

where TP, TN, FP and FN in Eqs.3–5 represent True Positive, True Negative, False Positive, and False Negative, respectively.

C. Classification Performance

We evaluate the COVID-19 vs. CAP classification on the collected chest CT images dataset. Table 2 shows the quantitative results (i.e., ACC, SEN, SPE, AUC, Precision and F1-score) achieved by different methods.

From Table 2, we can observe that our AFS-DF archives the best results in the terms of ACC, SEN, SPE, AUC, Precision and F1-score. In particular, our AFS-DF method achieves the highest classification accuracy (i.e., 91.79%), which is better than the LR (i.e., 89.81%), SVM (i.e., 89.97%), RF (i.e., 89.41%) and NN (i.e., 89.96%). In general, the proposed AFS-DF achieves 1.98%, 1.82%, 2.38% and 1.83% improvements in terms of ACC over the LR, SVM, RF and NN, respectively. We also perform the t -test between AFS-DF and the competing methods. AFS-DF shows significant improvement ($p < 0.05$) over LR (p -value = 0.0034), SVM (p -value = 0.0009), RF (p -value=0.0023) and NN (p -value = 0.0275). The possible reason for improvements is that our AFS-DF not only leverages the high-level feature representation by using the deep forest to improve the performance of COVID-19 vs. CAP, but also employs the discriminative features by using the adaptive feature selection process. The deep methods (i.e., NN and AFS-DF) show better classification accuracy in task of diagnosis of COVID-19 when compared with the conventional classifiers (i.e., LR, SVM and RF). The possible reason is that, the deep models can learn the high-level feature representative, which can boost the classification performance. It is worth noting that COVID-19 is the highly contagious disease, the higher SEN should have practically meaningful advantage for timely COVID-19 diagnosis to prevent the spread of the COVID-19. The SEN achieved by our AFS-DF for COVID-19 vs. CAP is 93.05%, which is better than other baseline methods. These results imply that using the adaptive feature selection guided deep forest model can improve the identification ability of COVID-19. We show the misclassification cases in Fig. 3. As shown in Fig. 3, the false negative cases are mostly the patients with small abnormality regions. In contrast, the false positive cases are mostly the patients with large abnormality regions.

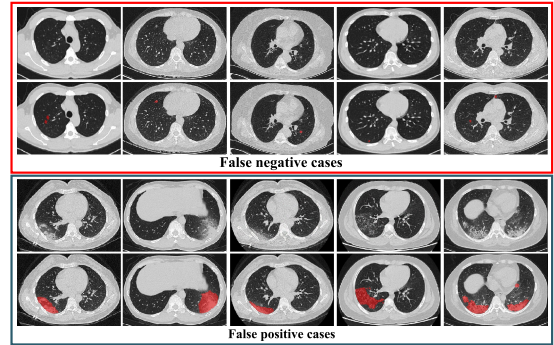


Fig. 3. Illustration of false negative cases and false positive cases.

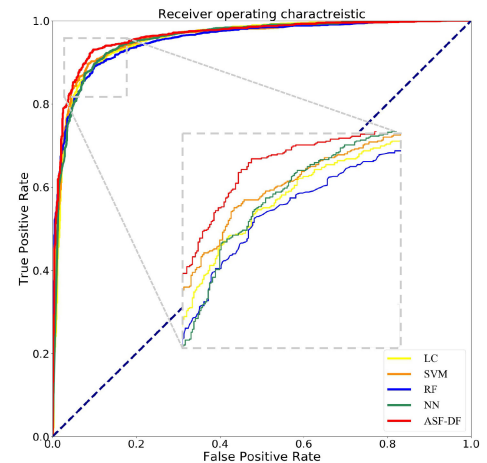


Fig. 4. ROC curves achieved by LR, SVM, RF, NN and AFS-DF in COVID-19 vs. CAP classification.

with small abnormality regions. In contrast, the false positive cases are mostly the patients with large abnormality regions.

As shown in Fig 4, our proposed AFS-DF method produces the best classification performance when compared with the baseline methods. These results further validate that using the high-level representation and adaptive feature selection strategy could improve the performance for COVID-19 vs. CAP classification. The top 30 important features in the extracted location-specific features are shown in Fig. 5. We calculate the average of normalized feature importance of location-specific features over the last layer in 5 trained AFS-DF model (as shown in Fig. 5(a)) and normalized feature importance of location-specific features in each trained AFS-DF model (as shown in Fig. 5(b)-Fig. 5(f)). Fig. 5 shows that the surface area features are more important on each fold for COVID-19 classification.

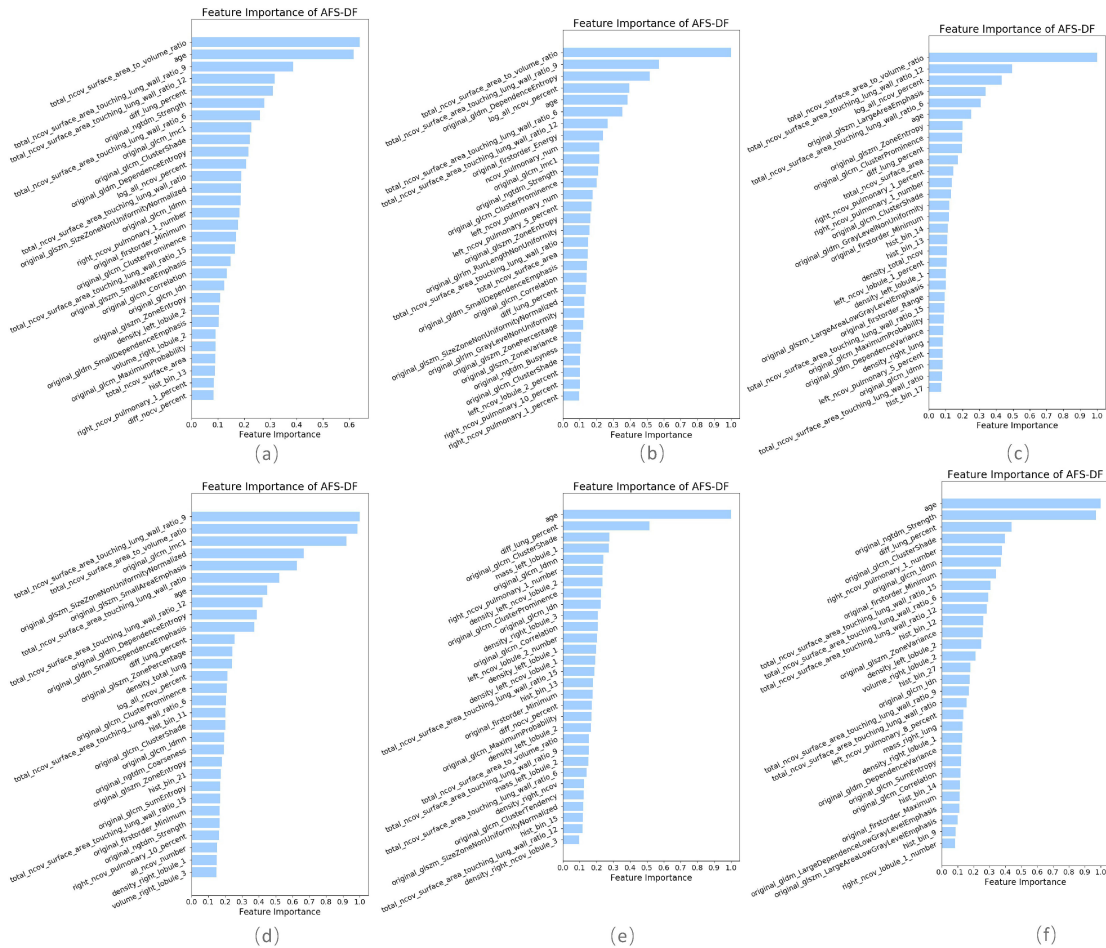


Fig. 5. The top 30 important features of location-specific features in AFS-DF. (a) average importance on all trained AFS-DF and (b) – (f) the importance on each trained AFS-DF.

IV. DISCUSSION

In this section, we first compare our proposed adaptive feature selection guided deep forest with several state-of-the-art methods for COVID-19 classification. Then, we study the influence of the adaptive feature selection strategy and the selected deep features that are learned by adaptive feature selection guided deep forest. Finally, we present the limitations of this work as well as possible future research directions.

A. Comparison With State-of-the-Art Methods

Since several attempts have been made for COVID-19 classification, we now compare our proposed AFS-DF with state-of-the-art methods. [13] employs a convolutional neural networks (CNN) to extract the features of chest CT images, and then combines the decision tree and Adaboost to produce the classification result of COVID-19 vs. typical viral pneumonia (using a total of 670 CT scans). [14] leverages a ResNet [27] to predict the COVID-19, Influenza-A viral pneumonia, and healthy cases (using a total of 618 CT scans). [10] extracts clinical and laboratory features and uses a logistic regression model for non-severe and severe patient classification (using a total of 196 CT scans). [11] extracts quantitative features, and

TABLE III
COMPARISON WITH STATE-OF-THE-ART METHODS FOR COVID-19 CLASSIFICATION

| Method | ACC (%) | SEN (%) | SPE (%) | AUC (%) |
|------------------|---------|---------|---------|---------|
| Wang et al. [13] | 73.1 | 67.0 | 74.0 | 78.0 |
| Xu et al. [14] | 86.7 | — | — | — |
| Shi et al. [10] | 89.0 | 82.2 | 82.8 | 89.0 |
| Tang et al. [11] | 87.5 | 93.3 | 74.5 | 91.0 |
| Shi et al. [12] | 87.9 | 90.7 | 83.3 | 94.2 |
| Mei et al. [26] | 83.5 | 84.3 | 82.8 | 92 |
| AFS-DF | 91.79 | 93.05 | 89.95 | 96.35 |

introduces a random forest method to assess the severity of the COVID-19 patient (using a total of 176 CT scans). [12] extracts location-specific handcrafted features, and proposed an infection size-adaptive random forest for COVID-19 classification (using a total of 2685 CT scans). [26] integrates the CNN and traditional machine learning method for COVID-19 vs. SARS-Cov-2 classification (using a total of 1324 CT scans).

The results are reported in Table 3. One can observe from Table 3 that the proposed AFS-DF shows competitive classification performance for COVID-19 patient identification. The underlying reason could be that our AFS-DF can utilize the

TABLE IV
PERFORMANCE OF COVID-19 vs. CAP CLASSIFICATION BY USING LASSO, ELASTICNET, DF AND AFS-DF

| Method | ACC (%) | SEN (%) | SPE (%) | Precision (%) | F1-score (%) |
|------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| Lasso | 90.33 ± 0.87 | 92.51 ± 1.25 | 87.14 ± 0.53 | 91.28 ± 0.60 | 91.88 ± 0.81 |
| ElasticNet | 90.25 ± 1.26 | 92.71 ± 1.41 | 86.68 ± 1.35 | 90.99 ± 1.23 | 91.84 ± 1.15 |
| DF | 91.16 ± 1.34 | 92.85 ± 1.54 | 88.65 ± 1.59 | 92.30 ± 0.71 | 92.57 ± 1.08 |
| AFS-DF | 91.79 ± 1.04 | 93.05 ± 1.67 | 89.95 ± 1.28 | 93.10 ± 0.75 | 93.07 ± 0.91 |

high-level discriminative representation of the extracted features. It is worth noting that the deep forest-based method could handle the small-scale data, and just spend less computational resource than the deep neural network-based methods.

B. Influence of Adaptive Feature Selection

To study the effectiveness of the proposed adaptive feature selection, we compare it to two feature selection methods (i.e., Lasso and ElasticNet [28]) and its variant (i.e., Deep Forest (DF)). Lasso is used to select a discriminative subset of features from the feature vector by using l_1 -norm sparsity constraint. The parameter for the sparsity constraint in this method is set as 0.001 by using cross-validation. In ElasticNet, a l_2 -norm is leveraged to reduce dimension of extracted features. Also, a l_2 -norm is further introduced into the classification model to ensure the smoothness of the linear model. The parameters for the sparsity constraint and smoothness constraint in ElasticNet are set as 0.001 and 0.1 by using cross-validation, respectively. DF is a variant of the proposed AFS-DF, which employs the same initial architecture of our AFS-DF model without feature selection block. Hence, the predictions of each layer are concatenated with all original features, which will be fed into next layer for prediction. Note that this variant method has same parameter setting with AFS-DF for training and test in our study. The experimental results are reported in Table 4.

As shown in Table 4, the proposed AFS-DF achieves the best classification performance, when compared with Lasso, ElasticNet and DF. In particular, compared with the feature selection methods (i.e., Lasso and ElasticNet), the proposed AFS-DF achieves better performance by using the high-level feature representation. Meanwhile, by using the adaptive feature selection operation, AFS-DF achieves better performance when compared with DF. These results imply the effectiveness of the proposed AFS-DF. Besides, as can be seen from Table 2, Table 3 and Table 4, the feature selection methods (i.e., Lasso, ElasticNet and proposed AFS-DF) show the competitive results. The possible reason is that the extracted features by using the deep learning methods or handcraft features may degrade classification performance due to heterogeneity between extracted features and subsequent traditional classification algorithms (i.e., decision tree, Adaboost, etc.). While the feature selection methods could select the features that are more relevant to the COVID-19 classification for subsequent classification task.

C. Influence of Features

To evaluate the effectiveness of the selected deep features (e.g. the features used in the last layer of AFS-DF) for COVID-19 vs. CAP classification, we further develop three methods

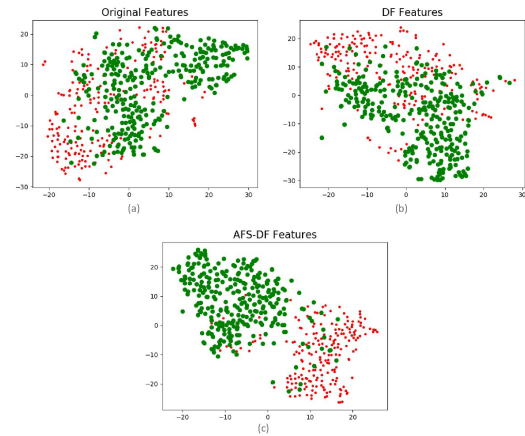


Fig. 6. Visual illustration of original location-specific features (a), the features of the last layer in DF (b), and the features of the last layer in AFS-DF (c), which uses the t-SNE to do the feature dimensionality reduction.

based on LR, SVM and RF by using the selected deep features (i.e., AFSDF-LR, AFSDF-SVM and AFSDF-RF). Meanwhile, we also compare these methods with LR, SVM and RF based on the features used in deep forest (i.e., DF-LR, DF-SVM and DF-RF). We evaluate these nine methods for COVID-19 vs. CAP classification, with the results reported in Table 5.

As can be seen from Table 5, the proposed AFSDF-LR, AFSDF-SVM and AFSDF-RF outperform their counterparts (i.e., LR, SVM and RF) in most of evaluation metrics. Of note, the proposed methods consistently achieve better results in terms of ACC and SEN. For example, AFSDF-LR, AFSDF-SVM and AFSDF-RF achieve 1.38%, 1.15% and 1.11% improvement over LR, SVM and RF in terms of ACC for COVID-19 vs. CAP classification, respectively. Compared with LC, SVM and RF, the AFSDF-LR, AFSDF-SVM and AFSDF-RF also show improvement in terms of SEN for COVID-19 vs. CAP classification, respectively. The possible reason is that, with the selected deep features by using AFS-DF, the features include the high-level and discriminative information. Hence, the conventional machine learning methods (i.e., LR, SVM and RF) can use these selected deep features to improve the performance of COVID-19 classification task. In addition, the AFSDF-based methods achieve better classification results over the DF-based methods (i.e., DF-LR, DF-SVM and DF-RF). Besides, In Fig 6, we plot the original location-specific features, the features of the last layer in DF, and the features of the last layer in AFS-DF, which performs dimensionality reduction by using t-SNE [29]. As shown in Fig 6, our AFS-DF produces more discriminative features for COVID-19 vs. CAP classification. Although the

TABLE V

PERFORMANCE OF COVID-19 vs. CAP CLASSIFICATION BY USING LR, DF-LR, AFSDF-LR, SVM, DF-SVM, AFSDF-SVM, RF DF-RF AND AFSDF-RF

| Method | ACC (%) | SEN (%) | SPE (%) | Precision (%) | F1-score (%) |
|-----------|--------------|--------------|--------------|---------------|--------------|
| LR | 89.81 ± 1.63 | 91.64 ± 1.69 | 87.10 ± 2.14 | 91.22 ± 1.27 | 91.42 ± 1.37 |
| DF-LR | 90.33 ± 1.08 | 91.50 ± 1.50 | 88.62 ± 0.68 | 92.10 ± 1.10 | 91.80 ± 1.00 |
| AFSDF-LR | 91.19 ± 1.11 | 92.52 ± 1.51 | 89.08 ± 2.00 | 92.52 ± 1.20 | 92.51 ± 0.94 |
| SVM | 89.97 ± 1.37 | 91.51 ± 2.04 | 87.70 ± 0.89 | 91.38 ± 0.60 | 91.43 ± 1.21 |
| DF-SVM | 90.24 ± 1.08 | 91.44 ± 1.69 | 88.47 ± 1.00 | 92.05 ± 0.30 | 91.74 ± 0.95 |
| AFSDF-SVM | 91.12 ± 1.50 | 92.31 ± 1.82 | 89.38 ± 1.52 | 92.67 ± 1.11 | 92.49 ± 1.33 |
| RF | 89.41 ± 1.67 | 90.51 ± 0.87 | 87.85 ± 1.01 | 91.21 ± 1.10 | 90.85 ± 1.26 |
| DF-RF | 89.65 ± 1.64 | 90.95 ± 2.04 | 87.70 ± 1.51 | 91.51 ± 1.00 | 91.23 ± 1.48 |
| AFSDF-RF | 90.52 ± 1.21 | 92.51 ± 1.27 | 87.57 ± 2.50 | 91.62 ± 1.29 | 92.05 ± 0.94 |

deep forest could learn the high-level features, the learned features are still difficult to classify.

D. Limitations and Future Work

There are still several limitations in the current study. *First*, the adaptive feature selection guided deep forest is only validated on COVID-19 vs. CAP classification task. In the future, we plan to collect more data with multiple diseases, and perform our proposed method on other COVID-19 classification tasks (e.g., COVID-19 vs. Influenza-A viral pneumonia and CAP, severe patients vs. non-severe patients, etc.). *Second*, we extract the handcraft features by using prior knowledge in current work, in future, the features learned by deep learning method are expected to leverage our proposed method for further performance improvement.

V. CONCLUSION

In this paper, we propose an adaptive feature selection guided deep forest for COVID-19 vs. CAP classification by using the chest CT images. Specifically, the AFS-DF uses the deep forest to learn the high-level representation based on the location-specific features. Meanwhile, an adaptive feature selection operation is employed to reduce the redundancy of features based on the trained forest. Experimental results on the collected COVID-19 dataset with 1495 COVID-19 cases and 1027 CAP cases show that our proposed AFS-DF approach can achieve superior performance on COVID-19 classification with chest CT images in comparison with several existing methods.

REFERENCES

- [1] Z. Y. Zu *et al.*, "Coronavirus disease 2019 (COVID-19): A perspective from China," *Radiology*, vol. 296, no. 2, pp. E15–E25, 2020, Art. no. 200490.
- [2] E. K. Jeong *et al.*, "Coronavirus disease-19: The first 7,755 cases in the Republic of Korea," *Osong Public Health and Research Perspectives*, vol. 11, no. 2, 2020, doi: [10.24171/j.phrp.2020.11.2.05](https://doi.org/10.24171/j.phrp.2020.11.2.05).
- [3] C. Sohrabi *et al.*, "World Health Organization declares global emergency: A review of the 2019 novel coronavirus (covid-19)," *Int. J. Surg.*, vol. 76, pp. 71–76, 2020.
- [4] J. Wu *et al.*, "Chest CT findings in patients with corona virus disease 2019 and its relationship with clinical features," *Invest Radiol.*, vol. 55, no. 5, pp. 257–261, 2020.
- [5] Y. Fang *et al.*, "Sensitivity of chest CT for COVID-19: Comparison to RT-PCR," *Radiology*, vol. 296, no. 2, pp. E115–E117, 2020.
- [6] Y. Li and L. Xia, "Coronavirus disease 2019 (COVID-19): Role of chest CT in diagnosis and management," *Amer. J. Roentgenol.*, vol. 214, no. 6, pp. 1280–1286, 2020.
- [7] M. Chung *et al.*, "CT imaging features of 2019 novel coronavirus (2019-nCoV)," *Radiology*, vol. 295, no. 1, pp. 202–207, 2020.
- [8] M. Li *et al.*, "Coronavirus disease (covid-19): Spectrum of CT findings and temporal progression of the disease," *Academic Radiol.*, vol. 27, no. 5, pp. 603–608, 2020.
- [9] C. Long *et al.*, "Diagnosis of the coronavirus disease (covid-19): rRT-PCR or CT?" *Eur. J. Radiol.*, vol. 162, 2020, Art. no. 108961.
- [10] W. Shi *et al.*, "Deep learning-based quantitative computed tomography model in predicting the severity of COVID-19: A retrospective study in 196 patients," 2020, doi: [10.2139/ssrn.3546089](https://doi.org/10.2139/ssrn.3546089).
- [11] Z. Tang *et al.*, "Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest CT images," 2020, *arXiv:2003.11988*.
- [12] F. Shi *et al.*, "Large-scale screening of COVID-19 from community acquired pneumonia using infection size-aware classification," 2020, *arXiv:2003.09860*.
- [13] S. Wang *et al.*, "A deep learning algorithm using CT images to screen for Corona Virus Disease (covid-19)," 2020, *MedRxiv*.
- [14] X. Xu *et al.*, "Deep learning system to screen coronavirus disease 2019 pneumonia," 2020, *arXiv:2002.09334*.
- [15] Z.-H. Zhou and J. Feng, "Deep forest: Towards an alternative to deep neural networks," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, 2017, pp. 3553–3559.
- [16] F. Shan *et al.*, "Lung infection quantification of COVID-19 in CT images with deep learning," 2020, *arXiv:2003.04655*.
- [17] D. M. Hansell, A. A. Bankier, H. MacMahon, T. C. McLoud, N. L. Muller, and J. Remy, "Fleischner Society: Glossary of terms for thoracic imaging," *Radiology*, vol. 246, no. 3, pp. 697–722, 2008.
- [18] X. Li, X. Zeng, B. Liu, and Y. Yu, "Covid-19 infection presenting with CT Halo Sign," *Radiol.: Cardiothor. Imag.*, vol. 2, no. 1, 2020, Paper e200026.
- [19] A. Bernheim *et al.*, "Chest CT findings in coronavirus disease-19 (COVID-19): Relationship to duration of infection," *Radiology*, vol. 295, no. 3, 2020, Art. no. 200463.
- [20] F. Song *et al.*, "Emerging 2019 novel coronavirus (2019-nCoV) pneumonia," *Radiology*, vol. 295, no. 1, pp. 210–217, 2020, Art. no. 200274.
- [21] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 785–794.
- [22] A. Liaw *et al.*, "Classification and regression by randomForest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
- [23] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, 2006.
- [24] Y. Fan *et al.*, "Multivariate examination of brain abnormality using both structural and functional mri," *Neuroimage*, vol. 36, no. 4, pp. 1189–1199, 2007.
- [25] G. S. Fletcher, *Clinical Epidemiology: The Essentials*. Baltimore, MD, USA: Williams & Wilkins, vol. 26, 1224–1228, 2019.
- [26] X. Mei *et al.*, "Artificial intelligence-enabled rapid diagnosis of patients with covid-19," *Nat. Med.*, pp. 1–5, 2020.
- [27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [28] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Royal Stat. Soc.: Series B (Stat. Methodol.)*, vol. 67, no. 2, pp. 301–320, 2005.
- [29] L. V. D. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.