Pulmonary Lesion Subtypes Recognition of COVID-19 From Radiomics Data With Three Dimensional Texture Characterization in CT Images

Wei Li
Northeastern University
Yangyong Cao (1971582@stu.neu.edu.cn)
Northeastern University
Kun Yu
Northeastern University
Yibo Cai
Northeastern University
Feng Huang
Neusoft Medical System Co., Ltd.
Minglei Yang
Neusoft Medical System Co., Ltd.
Weidong Xie
Northeastern University

Research

Keywords: COVID-19, Lesion subtypes, 3D texture feature, Random forest, Hybrid adaptive feature selection, Radiomics

Posted Date: May 24th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-532131/v1

RESEARCH

Pulmonary Lesion Subtypes Recognition of COVID-19 from Radiomics Data with Three Dimensional Texture Characterization in CT Images

Wei Li[1], Yangyong Cao[2*†], Kun Yu[3], Yibo Cai[2], Feng Huang[4], Minglei Yang[4] and Weidong Xie[2†]

*Correspondence:
1971582@stu.neu.edu.cn
[2]School of Computer Science and Engineering, Northeastern University, China
Full list of author information is available at the end of the article
†Co-Corresponding author

Abstract

Background: The COVID-19 disease is putting unprecedented pressure on the global healthcare system. The CT examination as a auxiliary confirmed diagnostic method can help clinicians quickly detect lesions locations of COVID-19 once screening by PCR test. Furthermore, the lesion subtypes classification plays a critical role in the consequent treatment decision. Identifying the subtypes of lesions accurately can help doctors discover changes in lesions in time and better assess the severity of COVID-19.

Method: The most four typical lesion subtypes of COVID-19 are discussed in this paper, which are ground-glass opacity (GGO), cord, solid and subsolid. A computer aided diagnosis approach of lesion subtype is proposed in this paper. The radiomics data of lesions are segmented from COVID-19 patients CT images with diagnosis and lesions annotations by radiologists. Then the three dimensional texture descriptors are applied on the volume data of lesions as well as shape and First order features. The massive feature data are selected by hybrid adaptive selection algorithm and a classification model is trained at the same time. The classifier is used to predict lesion subtypes as side decision information for radiologists.

Results: There are 3734 lesions extracted from the dataset with 319 patients collection and then 189 radiomics features are obtained finally. The random forest classifier is trained with data augmentation that the number of different subtypes of lesions is imbalanced in initial dataset. The experimental results show that the accuracy of the four subtypes of lesions is (0.9306, 0.9684, 0.9958, and 0.9430), the recall is (0.9552, 0.9158, 0.9580 and 0.8075) and the f-score is (0.93.84, 0.92.37, 0.95.47, and 84.42).

Conclusion: The method is evaluated in multiple sufficient experiments. The results show that the 3D radiomics features chosen by hybrid adaptive selection algorithm can better express the advanced information of the lesion data. The classification model obtains a good performance and is compared the models of COVID-19 in the stat of art, which can help clinicians more accurately identify the subtypes of COVID-19 lesions and provide help for further research.

Keywords: COVID-19; Lesion subtypes; 3D texture feature; Random forest; Hybrid adaptive feature selection; Radiomics

Background

With the rapid growth of patients of the 2019 Coronavirus Disease (COVID-19), the shortage of clinicians is increasingly severe. Currently, clinicians mainly use RT-PCR technology to detect RNA in sputum or nasopharyngeal swabs to detect COVID-19 pneumonia. But this method has a certain false-negative rate[1]. Therefore, clinicians will also use chest CT images as a additional diagnostic method to improve the accuracy of COVID-19 detection for confirmed diagnosis. Moreover, the imaging pattern can change rapidly in a short period of time within the treatment process[2]. The research work of lesion segmentation have achieved good results in the diagnosis of COVID-19 through machine learning or deep learning methods.

In terms of machine learning, Shi et al. use medical imaging features and clinical features as input, and logistic regression as a classifier to distinguish COVID-19 [3]. Barstugan et al. have made improvements in 2D feature extraction. GLCM, LDP, GLRLM, GLSZM, and DWT were used to obtain the second-order statistical features for classification of COVID-19 [4]. Ozkaya et al. also proposed a new method that fuses and ranks deep features for early detection in SVM [5]. Elaziz et al. used the new Fractional Multichannel Exponent Moments (FrMEMs) to extract features from chest X-ray images. Then an improved Manta-ray search optimization based on differential evolution is used to select the most important features and a K-Nearest Neighbor classifier is used to distinguish COVID-19 [6]. Tuncer et al. proposed a feature generation method called Residual Exemplar Local Binary Pattern (ResExLBP) and used a novel iterative ReliefF (IRF) for feature selection. In their work, the SVM classifier achieved 100.0% classification accuracy by using 10-fold cross-validation [7].

In addition, some scholars have proposed some deep learning methods for the diagnosis of COVID-19. For example, Zhou et al. segment COVID-19 lesions from CT by using the U-Net segmentation network with a spatial and multi-channel attention mechanism to assist in diagnosis COVID-19 [8]. Khan et al. proposed a deep convolutional neural network called Coro-Net based on Xception architecture, which can detect COVID-19 infection from chest X-ray images [9]. Afshar et al. pointed out that CNN is easy to lose the spatial information between image instances, so an alternative framework based on the capsule network is proposed, which can handle small data sets [10]. Khalifa et al. fine-tuned deep transfer learning for limited data sets to detect pneumonia chest X-ray based on generative confrontation network [11]. Minaee et al. trained four popular convolutional neural networks, including ResNet18, ResNet50, SqueezeNet, and DenseNet-121, to identify COVID-19 disease in the analyzed chest X-ray images [12]. He et al. propose a synergistic learning framework for automated severity assessment of COVID-19 in 3D CT images, by jointly performing lung lobe segmentation and multi-instance classification [13]. Xu et al. use a 3D deep learning model to segment candidate infection areas from lung CT images, then score these areas, and finally uses noise or Bayes function to calculate the final confidence score to classify patients as COVID-19, Influenza-A viral pneumonia (IAVP), and not infected [14].

In summary, the above existing work mainly focuses on lesion detection of COVID-19 or its severity assessment. Few studies are paid attention to the classification of lesion subtypes, which ignores the important role of lesion subtypes in the diagnosis

of COVID-19 disease. The subtypes identification of lesions in a timely manner can enable clinicians to better assess the patient's condition and prescribe precise medicines in personality. Zhao et al. pointed out the severity and the symptoms of COVID-19 pneumonia are different from common pneumonia [15], and lesions and characteristics of it are also different from common pneumonia [16]. So the lesions caused by COVID-19 pneumonia are more worthy of further study. At the same time, if the patient lesion type can be determined more accurately, the doctor can more accurately determine the COVID-19 patient's condition by referring to the lesion type.

Therefore, it is necessary to study the computer aided diagnosis of COVID-19 lesion subtypes recognition, at the same time, the study can reduce the image reading burdens of radiologists in vast data. Moreover, the machine learning methods of classification for COVID-19 are mainly based on features extracting from the 2D CT images in the exist papers. 3D radiomics features can make full use of the advanced information of lesions which are not discussed in present work to the best of our knowledge in the state of art.

The contributions in this paper can be discussed by three aspects.

1 The pilot research work on lesions subtypes of COVID-19 is discussed in this paper, which has never been seen in previous studies so far and may greatly assist doctor diagnosis and evaluating severity of COVID-19 patients more effectively.
2 The 3D texture radiomics analysis method is applied on COVID-19 lesions diagnosis which is better to explore more hidden inner characters within the lesions to help experts better understand the pathological features of COVID-19.
3 Extensive experiments on clinical real-world datasets demonstrate the effectiveness of the proposed model of hybrid adaptive feature selection method. Moreover, we show the capability of the proposed model for the high dimension feature data with serious imbalance problem.

The rest of the paper is organized as follows. This paper first introduce the method with the composition of the data set, the characteristics of 3D features, and the feature selection strategy in Section Methodology. The experimental results on the prepared database are discussed in Section Results. We finally conclude this paper and look forward to the future work in Section Conclusion.

Results

In this section, the classifier evaluation criteria is illustrated firstly. Then, we present experimental results achieved by different methods on the evaluation dataset. Finally, the comparative experiments are conducted to prove the influence of data augmentation, 3D features and HAFS.

Experimental settings

Since COVID-19 is a new disease, there is few public data set of CT images with annotations suitable for this study. Therefore, we extracted lesions from 319 cases of COVID-19 pneumonia patients provided by Neusoft Medical to construct a dataset. All patients received a thin-slice CT scan of the chest by Neusoft 256 slice CT.

The final subsets of features are evaluated by RF classifications associated with 10-fold cross validations. Precision, Recall , Accuracy and F-measure are used to compare the estimated and known labels according to the following expressions:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

where TP, TN, FP and FN in Eqs. 7-10 represent True Positive, True Negative, False Positive, and False Negative, respectively.

In our experiments, we compare our model with the following widely adopted machine learning methods.

1 LogisticRegression(LR)
2 Support Vector Machine(SVM) (Use a radial basis function kernel with default parameters.)
3 KNeighborsClassifier(KNN) (Set $k = 20$ through cross-validation)
4 DecisionTreeClassifier(DT)
5 GaussianNB
6 Quadratic Discriminant Analysis(QDA)

Experimental evaluation of HAFS with RF model

We evaluate HAFS with RF (HAFS-RF) model on the collected chest CT images dataset. Table 2 shows the quantitative results achieved by different methods.

From Table 2, we can clearly observe that HAFS-RF achieved an accuracy of (93.06, 96.84, 99.58 and 94.3) for Label 1, 2, 3 and 4 respectively under the condition of $\alpha = 0.5$. Followed by DT (91.49, 95.53, 98.89 and 91.79), next is SVM (82.3, 92.31, 98.04 and 91.75). The accuracy of the remaining models such as KNN, LR, GaussianNB, and QDA is much lower than theirs. Obviously, HAFS-RF achieved the best performance. For each method, especially the accuracy of QDA is (63.72, 66.82, 99.97 and 79.07), the accuracy of Label 3 is always the highest, some of them even close to 100. The possible reason for this phenomenon is that although we have enhanced the data in the experiment, the number of the four types of lesions tends to be balanced. However, the number of lesions on Label 3 is still the least. On the contrary, the precision value of GaussianNB is (88.57, 42.62, 14.64, and 37.82), and

Table 1: Performance of COVID-19 classification achieved by SVM, KNN, LR, GaussianNB, QDA, HAFS-RF ($\alpha = 0.5$).

Method	Label	Precision	Recall	Accuracy	F-measure
SVM	1	76.34	99.42	82.3	86.37
	2	99.48	62.07	92.31	76.45
	3	100.0	57.75	98.04	73.21
	4	96.84	58.2	91.75	72.71
KNN	1	88.04	86.32	85.66	87.17
	2	83.09	83.23	93.19	83.16
	3	78.05	86.49	98.17	82.05
	4	65.98	67.96	87.58	66.96
LR	1	83.46	88.4	83.8	85.86
	2	76.93	75.3	89.83	76.11
	3	66.67	52.11	96.58	58.5
	4	55.86	50.27	83.7	52.92
GaussianNB	1	88.57	59.6	72.88	71.25
	2	42.62	62.72	75.52	50.75
	3	14.64	86.62	76.01	25.05
	4	37.82	10.19	79.89	16.05
QDA	1	95.14	36.67	63.72	52.94
	2	39.1	97.27	66.82	55.78
	3	100.0	99.3	99.97	99.65
	4	43.38	48.66	79.07	45.87
DT	1	91.85	92.57	91.49	92.21
	2	91.53	87.21	95.53	89.32
	3	86.75	90.34	98.89	88.51
	4	78.36	79.93	91.79	79.14
HAFS-RF	1	92.21	95.52	93.06	93.84
	2	93.17	91.58	96.84	92.37
	3	95.14	95.8	99.58	95.47
	4	88.43	80.75	94.3	84.42

the fitting ability is seriously insufficient. The reason may be that GaussianNB is prone to under-fitting for a small number of samples.

Fig.4 shows the ROC curves of different models. It is also obvious that our method has the highest ROC curve area. These results all show that HAFS-RF can improve the performance and efficiency of COVID-19 classification.

Influence of data augmentation

To evaluate the effectiveness of the data augmentation, we compare it to without data augmentation, with the results reported in Table 3.

Table 2: Performance of COVID-19 classification achieved with data augmentation.

Augmentation	Label	Number	Precision	Recall	Accuracy	F-measure
with	1	2637	93.17	96.85	92.95	94.97
	2	519	89.84	86.02	96.88	87.89
	3	103	89.47	77.27	99.16	82.93
	4	475	82.94	73.25	93.55	77.79
without	1	2637	92.21	95.52	93.06	93.84
	2	1098	93.17	91.58	96.84	92.37
	3	386	95.14	95.8	99.58	95.47
	4	976	88.43	80.75	94.3	84.42

As can be seen from Table 3, after data augmentation, the number of the four types of lesions changed from (2637, 519, 103 and 475) to (2637, 1098, 386 and 976). The data augmentation consistently achieve better results in Label 2, 3 and 4, and worse in Label 1. For example, data augmentation achieve (93.84, 92.37, 95.47 and 84.42) in terms of F-measure, None achieve (93.84, 92.37, 95.47 and 84.42).

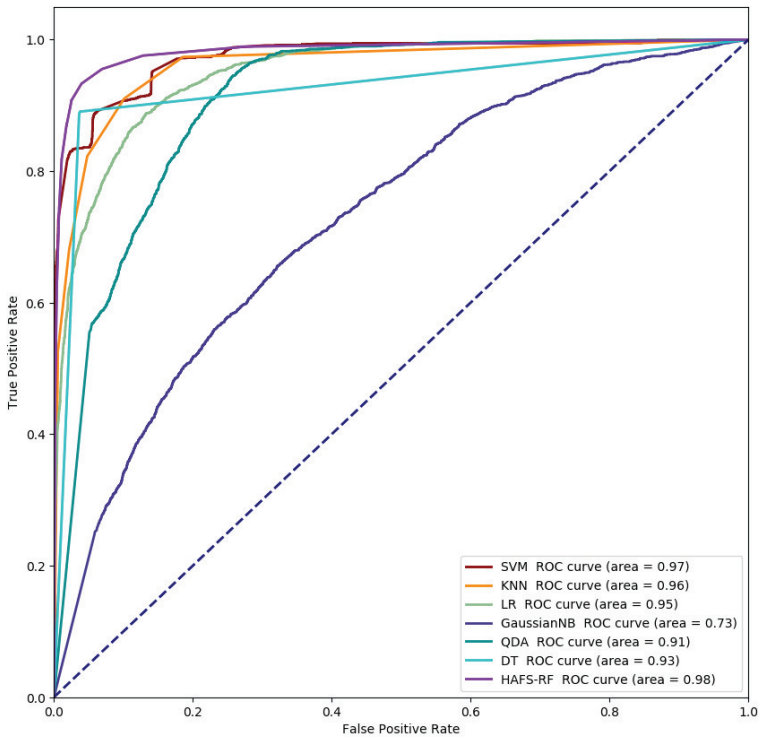

Figure 1: ROC curves achieved by different models.

The possible reason is that the excessive number of samples of Label 1 leads to the over-fitting of the model. On the contrary, the insufficient number of other types leads to insufficient fitting ability. After using data augmentation, the four sample sizes are relatively balanced, thus avoiding overfitting to Label 1, so the score of Label 1 decreases, but the overall score increases.

Influence of 3D features

As shown in Fig.5, each feature will have its own score in HAFS. The green features are discarded in the first stage of HAFS, the orange features are discarded in the second stage of HAFS, and the blue features are selected after HAFS. Tabel 4 shows the details of blue features of Fig.5. It is obvious that after feature selection, a total of 48 features out of 189 features were retained. Among them, there are 18(6×3) 2D features and 30 3D features. More 3D features are retained than 2D features. So 3D features may be more effective than 2D features.

To study the influence of 3D features, a comparison was done with and without the 3D features for our model. Results of the evaluated criteria are given for the 189 features in Table 5. As shown in Table 5, HAFS-RF achieves the better classification accuracy when we use 2D and 3D features (93.06, 96.84, 99.58 and 94.3) than when we use 2D features (89.37, 93.12, 98.47 and 91.2). The possible reason for improvements is that the 3D features we use have high-level feature representation, thereby improving the typing performance.

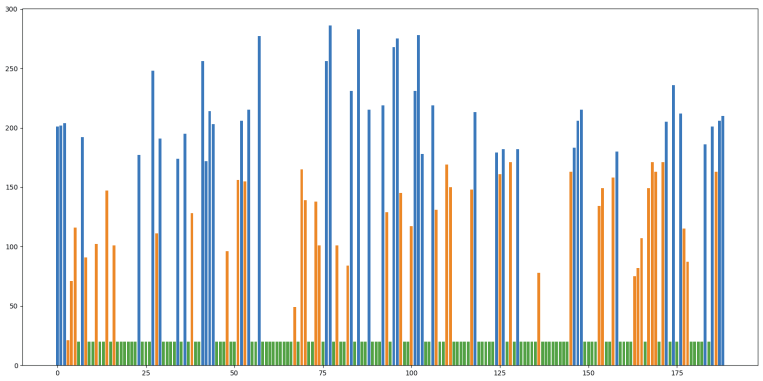

Figure 2: Scores of 189 features.

Table 3: Features after HAFS.

Dimension of features	Kind of features	Characteristics
2D	Firstorder	Length,Mean,Max,Var,ASM,Energy
3D	Firstorder	RobustMeanAbsoluteDeviation, Mean, RootMeanSquared, Range, InterquartileRange, Skewness
	Glszm	GrayLevelVariance, HighGrayLevelZoneEmphasis, ZonePercentage, SmallAreaLowGrayLevelEmphasis
	Glrlm	LongRunHighGrayLevelEmphasis, DifferenceVariance, GrayLevelNonUniformityNormalized, RunPercentage
	Glcm	SumSquares, Id, JointAverage
	Gldm	DependenceNonUniformityNormalized, DependenceEntropy, DependenceEntropy
	Shape	MajorAxisLength

Table 4: Performance of COVID-19 classification achieved by using 2D features and using 2D and 3D features.

Feature	Label	Precision	Recall	Accuracy	F-measure
2D	1	88.38	93.56	89.37	90.89
	2	84.18	81.9	93.12	83.02
	3	85.94	79.14	98.47	82.4
	4	79.84	69.29	91.2	74.19
2D and 3D	1	92.21	95.52	93.06	93.84
	2	93.17	91.58	96.84	92.37
	3	95.14	95.8	99.58	95.47
	4	88.43	80.75	94.3	84.42

Influence of HAFS

To study the effectiveness of the HAFS selection. Firstly, We compare HAFS with state-of-the-art feature selection methods (F-test, MIC, REF, and Lasso). Since they can't determine the optimal number of features, we select the same number of features of them as HAFS-RF($\alpha = 0.5$) for comparison experiments. The results are reported in Table 6.

One can observe from Table 6 that compared to the other four methods, HAFS gets the highest accuracy (93.06, 96.84, 99.58, and 94.3). This proves from the side that the features selected by HAFS are more representative.

Secondly, we further develop four methods based on SVM, KNN, GaussianNB, and QDA by using HAFS (*i.e.*, HAFS-SVM, HAFS-KNN, HAFS-GaussianNB, and

Table 5: Performance of different feature selection algorithm achieved by F-test, MIC, RFE, Lasso, HAFS($\alpha = 0.5$) Using Random Forest.

Method	Label	Precision	Recall	Accuracy	F-measure
F-test	1	86.59	91.66	87.32	89.05
	2	78.32	71.99	90.51	75.02
	3	74.62	64.67	97.2	69.29
	4	71.43	67.52	88.66	69.42
MIC	1	87.4	93.09	88.69	90.16
	2	76.75	75.04	90.91	75.89
	3	88.0	70.06	97.98	78.01
	4	73.42	65.59	88.27	69.28
RFE	1	84.82	93.32	86.99	88.87
	2	81.29	77.26	92.28	79.23
	3	88.07	61.15	97.59	72.18
	4	75.43	63.97	88.53	69.23
Lasso	1	87.69	93.44	89.05	90.47
	2	77.84	73.85	91.0	75.79
	3	80.45	68.15	97.52	73.79
	4	74.69	67.69	88.85	71.02
HAFS	1	92.21	95.52	93.06	93.84
	2	93.17	91.58	96.84	92.37
	3	95.14	95.8	99.58	95.47
	4	88.43	80.75	94.3	84.42

HAFS-QDA). We evaluate these eight methods, with the results reported in Table 7.

Table 6: Performance of HAFS achieved by SVM, KNN, GaussianNB and QDA by using and not using HAFS.

Method	α	Label	Precision	Recall	Accuracy	F-measure
SVM		1	76.34	99.42	82.3	86.37
		2	99.48	62.07	92.31	76.45
		3	100.0	57.75	98.04	73.21
		4	96.84	58.2	91.75	72.71
HAFS-SVM	0.1	1	90.08	95.03	91.3	92.49
		2	91.64	87.03	95.8	89.28
		3	99.1	77.46	98.92	86.96
		4	84.98	80.14	93.58	82.49
KNN		1	88.04	86.32	85.66	87.17
		2	83.09	83.23	93.19	83.16
		3	78.05	86.49	98.17	82.05
		4	65.98	67.96	87.58	66.96
HAFS-KNN	0.5	1	89.01	87.01	86.6	88.0
		2	83.17	84.52	93.42	83.84
		3	80.77	85.14	98.31	82.89
		4	66.89	69.37	87.97	68.11
GaussianNB		1	88.57	59.6	72.88	71.25
		2	42.62	62.72	75.52	50.75
		3	14.64	86.62	76.01	25.05
		4	37.82	10.19	79.89	16.05
HAFS-GaussianNB	0.1	1	81.77	77.8	77.71	79.74
		2	46.21	51.38	78.19	48.66
		3	34.96	55.63	93.16	42.93
		4	46.67	41.11	80.02	43.71
QDA		1	95.14	36.67	63.72	52.94
		2	39.1	97.27	66.82	55.78
		3	100.0	99.3	99.97	99.65
		4	43.38	48.66	79.07	45.87
HAFS-QDA	0.3	1	86.07	82.19	82.69	84.09
		2	60.85	82.88	84.84	70.17
		3	58.93	92.96	96.68	72.13
		4	56.51	31.84	83.12	40.73

As shown in Table 7, the accuracy of QDA is increased from (63.72, 66.82, 99.97, 79.07) to (82.69, 84.84, 96.68, 83.12). We can see that HAFS is effective and can

improve the performance of the method for different methods. The reason may be that HAFS can select a small number of irrelevant features from a large number of features, thereby avoiding the phenomenon of over-fitting. And we can see that for different models, the best α is different, for example, for SVM, the best α is 0.1, KNN is 0.5, GaussianNB is 0.1, and QDA is 0.3. The possible reason for the difference is that the principle of the classifier is not the same.

Conclusion

The most four typical lesion subtypes of COVID-19 are discussed and a computer aided diagnosis approach of lesion subtype is proposed in this paper. Then the three dimensional texture descriptors are applied on the volume data of lesions as well as shape and first order features. The massive feature data are selected by hybrid adaptive selection algorithm and a classification model is trained at the same time. Extensive experiments on clinical real-world datasets demonstrate the effectiveness of the proposed model of hybrid adaptive feature selection method. Moreover, we show the capability of the proposed model for the high dimension feature data with serious imbalance problem. The results show that the 3D radiomics features chosen by hybrid adaptive selection algorithm can better express the advanced information of the lesion data. The classification model obtains a good performance and is compared the models of COVID-19 in the stat of art, which can help clinicians more accurately identify the subtypes of COVID-19 lesions and provide help for further research.

Methodology

In this study, four lesion subtypes are studied, namely ground-glass opacity (GGO, referred as label 1), cord (referred as label 2), solid (referred as label 3), and subsolid (referred as label 4) [17]. The CT images of the prepared dataset in this paper are shown in Fig.1, which are presented in transverse, sagittal and coronal plane respectively, and the lesions are annotated by two radiologists with ITK-SNAP software.

The numbers of the four types of lesions are 2637, 519, 103 and 475 respectively and the total is 3734 in original dataset. Table 1 shows the summary of the prepared dataset that are maximum, minimum and standard variance values of the sizes of three directions and volumes for each lesions subtypes. The subtype of ground-glass opacity hosts the majority of COVID-19 which shows imbalance data problem. The subsolid subtype is the largest in size of lesions while the cord subtype is the smallest one.

Based on this data set, the method process is shown in Fig.2 which includes four steps. We firstly introduce the algorithm of lesions extraction and data augmentation used in this study. Then, the feature extraction process for the 2D and 3D features are discussed. The implementation details is presented subsequently. Finally, we describe the random forest model in the forth step.

Lesions data extraction and augmentation

VB-Net is to predict and segment the image of the unknown lesion location. It combines V-Net and bottleneck layers to reduce and combine redundant information

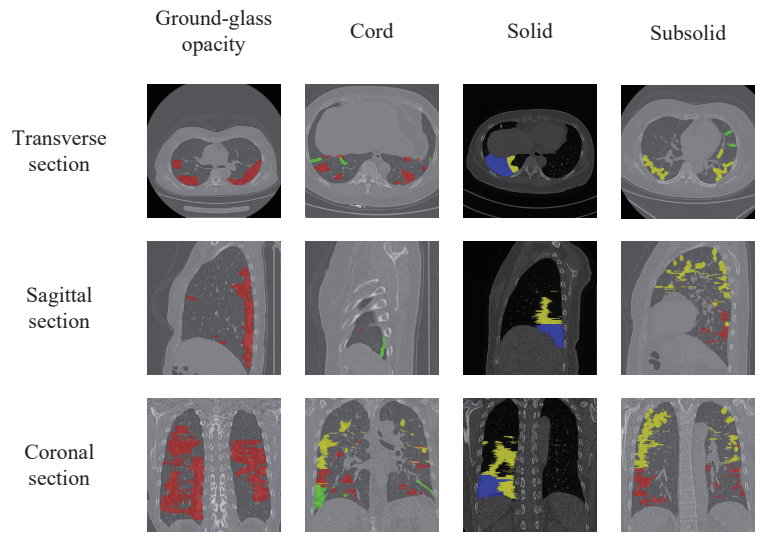

Figure 3: Typical four lesion subtypes in CT images of COVID-19. The labels and regions are given by medical experts. The red area represents ground-glass opacity, the green area represents cord, the blue area represents solid and the yellow area represents subsolid.

Table 7: Samples of lesion from prepared dataset COVID-19.

Label	Num	Statistics	X-length	Y-length	Z-length	Volume
1	2637	Max	247.0	312.0	399.0	20148480.0
		Mean	46.78	49.2	24.84	395826.08
		Std	42.2	50.68	39.63	1596465.23
2	519	Max	186.0	205.0	310.0	5142630.0
		Mean	43.38	41.28	24.86	124041.76
		Std	31.4	27.92	28.17	409901.04
3	103	Max	217.0	301.0	223.0	5878530.0
		Mean	40.46	37.9	21.53	210307.5
		Std	46.06	45.99	26.49	752977.45
4	475	Max	204.0	283.0	378.0	16873920.0
		Mean	61.9	63.26	38.79	722428.14
		Std	50.65	57.84	53.43	1979445.47

[18]. However, all the 3D volume data used in this experiment has the mask position of the corresponding lesion from annotations by the radiologists. Therefore, the VB-Net or V-Net will cause an error in extracting lesions. So we firstly use a breadth-first traversal (BFS) based lesion extraction algorithm to extract lesions which is mostly used in graph structure data. It should be noted that the lesions locations and subtypes are labeled by two radiologists that ensure the accuracy of evaluation data.

The BFS-based lesion extraction algorithm is shown as Algorithm 1. We traverse the mask array, which is created by radiologists, in the entire 3D volume data. When traversing the pixels with a lesion mark, this paper uses the BFS to extract the lesion range from the case with that mask [19]. The input is generally a point in the graph, then use the point to initialize a queue. The main idea of the algorithm is to take out a point from it each time, and then for this point, all nearby points

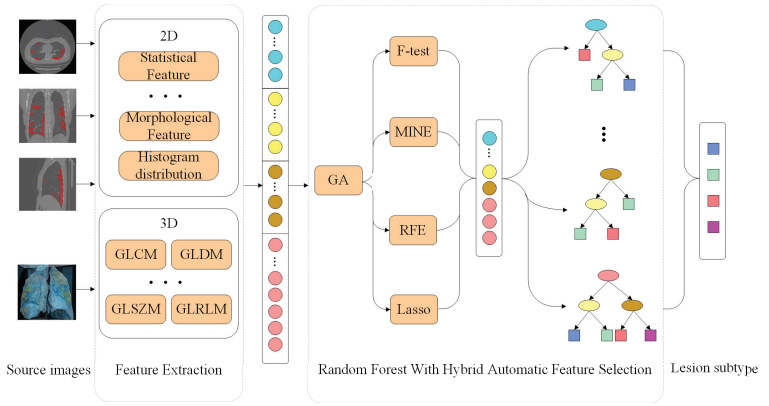

Figure 4: Overview of the COVID-19 Classification Using Random Forest Based on Hybrid Adaptive Feature Selection.

that meet the requirements are enqueued, and then the above process is repeated until the queue is empty.

Input: 3D coordinates of a point,A; Mask of the lesion, $Mask$; Array of mark, $mark$; Length of the search range,L.
Output: Range
$Q = \{A\}$;
$Range = \{+\infty, -\infty, +\infty, -\infty, +\infty, -\infty\}$;
while $Q \neq \emptyset$ **do**
  $Cur = Q.get()$;
  Update Range with Cur
  for N(x,y,z) $\in$ Mask **do**
    if $dist(N, Cur) \leq 5$ **then**
      if mark[N] = false **then**
        if Mask[N] = true **then**
          $Q.put(N)$;
        end
      end
    else
      $mark[N] = ture$;
  end
end

Algorithm 1: The lesion extraction algorithm.

Because we set the search range L in the algorithm, the algorithm can well avoid the discontinuity of the lesion data in a certain dimension caused by inaccurate data labeling. And we set a global mark in the algorithm, all traversed points will be recorded. There are two advantages to this:

1 In the current BFS process, the marked points will not be enqueued, which can avoid double calculation;

2 In the global traversal of the mask data, the marked points will not call the BFS algorithm repeatedly. This can ensure that each time the BFS algorithm only generates one lesion and returns its 3D ranges.

As we all know, the data imbalance problem will lead to a decrease in the accuracy of multi-classification. The four different subtypes of lesions number are 2637, 519, 103 and 475. Obviously they are unbalanced, and the number of Label 1 lesions is far greater than Label 3, which can leads to insufficient fitting ability of the classifier to Label 3 samples. The unbalance characteristic data can shift the decision boundary of the classifier, and affect the final classification effect. Therefore, we have adopted a data enhancement method based on ADASYN proposed by He [20] to reduce its impact on classification accuracy. This paper adopts a data enhancement method based ADASYN to increase the number of Label 2, 3 and 4. The method is briefly introduced in Algorithm 2.

Input: $Dataset_{train}$
Output: The synthetic data x_{syn}
The number of majority class is defined as: n_l;
The number of minority class is defined as: n_s;
Calculate the degree of class imbalance: $Degree_{imbalance} = n_s/n_l$;
if $Degree_{imbalance} < Degree_{threshold}$ **then**
  Calculate the number Δ_{syn} of synthetic data to be generated:
  $\Delta_{syn} = (n_l - n_s) \times \alpha \quad \alpha \in [0, 1]$;
  for $x_i \in$ minority class **do**
    Calculate the ratio r_i defined as $r_i = N_{KNeighbor}/K$ where
    $N_{KNeighbor}$ is the number of majority class examples in the K
    nearest neighbors of x_i;
    Normalize r_i $\hat{r_i} = r_i / \sum_{i=1}^{n_s} r_i$;
    Calculate the number g_i of synthetic data defined as $g_i = \hat{r_i} \times \Delta_{syn}$;
    for $i = 1$ to g_i **do**
      for Random choose $x_{mi} \in$ K nearest neighbors of x_i **do**
        $x_{syn} = x_i + (x_{mi} - x_i) \times \beta \quad \beta \in [0, 1]$;
      end
    end
  end
end

Algorithm 2: The algorithm of data enhancement based on ADASYN in lesion subtypes unbalance improvement process.

The algorithm increases the number of four subtype lesions from (2637, 519, 103 and 475) to (2637, 1098, 386 and 976) which is evaluated that the data augmentation can effectively improve the classification performance evaluated by the experiments in Section Results.

Three dimentional feature extraction

The most of the current medical imaging research on COVID-19 are mainly based on X-rays images or ignoring the characteristics of CT planar images. Furthermore, 3D features are also seldom considered in the research. Therefore, this paper extracted

the 2D features of some certain layers and more 3D features of the CT data to better characterize the lesion information.

The existing methods are mainly based on extracting 2D features from CT images. This ignores that the COVID-19 lesion is a kind of volume data. Therefore, in the process of feature extraction, the connection between layers is ignored, and some hidden features are lost.

In order to improve the accuracy of classification, we extracted multiple types of features of the lesion, that include 2D and 3D features and are shown in details as following.

1 **Infected lesions number**: The stage of the covid-19 affects the number of lesions and also affects the distribution of different types.Therefore, we add the total number of lesions in the same patient to the feature.

2 **Shape features**: Some cord-type lesions are significantly different from other lesions in shape, so we extracted three-dimensional shape features from the lesions in order to improve the accuracy of multi-classification.

3 **First order features**: The first order feature provide information related to the gray-level distribution of the image. We first obtain the middle layer and the layer with the largest lesion area of CT images in three directions, and extract 14 two-dimensional manual features from them, including mean, var, max, skewness, kurtosis, area, compact, rough, contrast, dissimilarity, homogeneity, energy, correlation, ASM. Then we also extract the three-dimensional features and hybrid them into the total features.

4 **Second-order features**: Second-order features give more information about the relative positions of the various gray levels within the lesion image.

   (a) **The gray-level co-occurrence matrix (GLCM)**[21]: The GLCM is a statistical method of analyzing texture that considers the spatial relationship of pixels. The $(i,j)^{th}$ element of GLCM $P(i,j|\delta,\theta)$ describes the number of times the combination of levels i and j occur in two pixels in the image, that are separated by a distance of δ pixels along angle θ. The 3D-GLCM considers 13 directions and need to be calculated separately and finally averaged.

   (b) **The Gray Level Run Length Matrix (GLRLM)**[22]: The GLRLM quantifies gray level runs, which are defined as the length in number of pixels of the same gray level value. The $(i,j)^{th}$ element of the GLRLM $P(i,j|\delta,\theta)$ represents the number of runs with gray level i and length j occur in the image along angle θ. Similar to 3D-GLCM, 3D-GLRLM also needs to be calculated separately for 13 directions.

   (c) **The Gray Level Size Zone Matrix(GLSZM)**[23]: In the GLSZM $P(i,j)$, the $(i,j)^{th}$ element represents the number of zones with gray level i and size j appear in the image. A zone is defined as the number of connected voxels that share the same gray level intensity. Contrary to GLCM and GLRLM, the 3D-GLSZM is rotation independent, with only one matrix calculated for all directions.

   (d) **Gray Level Dependence Matrix(GLDM)**[24]: The GLDM quantifies gray level dependencies in the image. The $(i,j)^{th}$ element in GLDM $P(i,j|\alpha)$ equals the number of times a voxel with gray level i with j

dependent voxels in its neighbourhood appears in an image. A neighbouring voxel with gray level j is considered dependent on center voxel with gray level i if $|i-j| \leq \alpha$. Similar to 3D-GLRLM, the 3D-GLDM is also rotation independent.

We introduced some 3D features from GLCM, GLSZM, GLRLM, GLDM, and the numbers are 22, 16, 16, 14 respectively.The detailed information of features can be found in the Pyradiomics[25] document on https://pyradiomics.readthedocs.io. These features correspond to the GLCM(1-15 16-21 23-24), GLRLM(1-16), GLSZM(1-16), and GLDM(1-14).

In addtion, we also extracted 4 features that are width, height, length and volume of the 3D lesion from the bounding box of COVID-19 lesion. In summary, a total of 189 features are used in our study.

Hybrid adaptive feature selection method

As described in Section Three dimentional feature extraction, we extract specific features from 3D lesion data. However, too much irrelevant information in features can easily cause over-fitting of the model, which will reduce the accuracy of the test set. Therefore, before training the model, it is necessary to perform feature selection processing to reduce the influence of the redundant feature.

There are three kinds of algorithms in the existing feature selection process, namely filters, wrappers, and embedded. Although these methods have their advantages, they all have one obvious disadvantage that the number of features in the subset after feature selection cannot be determined. So at this stage, we proposed a Hybrid Adaptive Feature Selection (HAFS) algorithm. It can not only solve the problem of the uncertain number of features in the subset, but also integrate the advantages of various traditional methods.

HAFS method is divided into two stages. In the first stage, the feature set $F = \{f_1, f_2 ... f_{189}\}$ is used as input, and the genetic algorithm (GA) is used to accurately select features. First, a population of q chromosomes will be initialized, and each chromosome is a binary set of length n $U = \{U_1, U_2 ... U_{189}\}$, each value of U_i is 1 or 0, if $U_i = 1$, it means f_i is selected , otherwise, f_i is not selected. Next, p iteration will be sperformed. Before each iteration, individual fitness is evaluated for each chromosome. Then, according to the fitness, the chromosomes in the population are calculated with different probabilities of three genetic operators that are selection, crossover, and mutation. At the end of the iteration, the feature subset $F_{GA} = \{f_{GA1}, f_{GA2} ... f_{GAm}\}$ is determined according to the value of each bit according to the chromosome U_{max} with the highest fitness. A schematic illustration of the first stage is shown in Fig.3.

In the second stage, in order to integrate the advantages of multiple feature selection methods, we use F_{GA} as input, and take two filters method that are F-test [26] and Maximal Information Coefficient (MIC) [27], one wrappers method that is Recursive Feature Elimination (RFE) [28], and one embedded method that is L1 regularized linear regression model(Lasso) [29] into data preprocessing methods to score features and sort them in ascending order, the results are presented as F_1, F_2, F_3 and F_4:

$$F_1 = F - test(F_{GA}) = \{f_{F1}, f_{F2} ... f_{Fm}\} \tag{5}$$

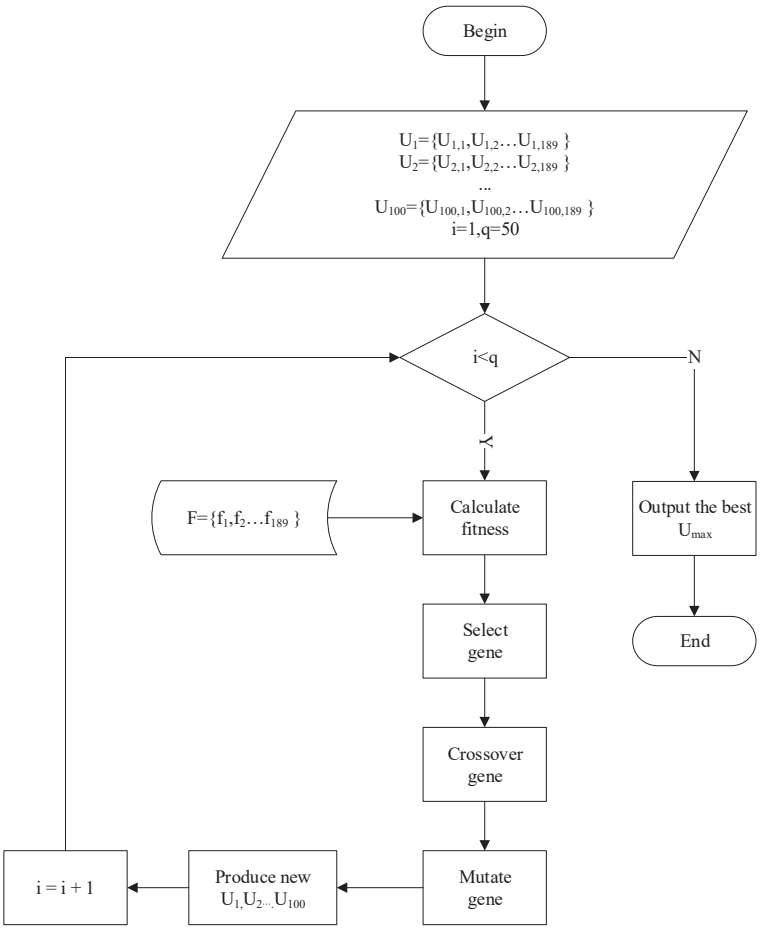

Figure 5: A schematic illustration of first stage of HAFS.

$$F_2 = MIC(F_{GA}) = \{f_{M1}, f_{M2}...f_{Mm}\} \tag{6}$$

$$F_3 = RFE(F_{GA}) = \{f_{R1}, f_{R2}...f_{Rm}\} \tag{7}$$

$$F_4 = Lasso(F_{GA}) = \{f_{L1}, f_{L2}...f_{Lm}\} \tag{8}$$

So we score each f_{GA} in F_{GA} according to F_1, F_2, F_3 and F_4. According to the $S(f_{GA})$ score, F_{GA} is sorted in ascending order. In this way we can decide which features to keep. The scoring standard, $S(f_{GA})$, is:

$$S(f_{GA}) = \frac{\sum_{i=1}^{4} index(F_i, f_{GA})}{4} \tag{9}$$

On the other hand, in order to further prevent feature over fitting after GA selection, We decided to select some features of F_{GA} and we set a parameter α as the ratio of the number of selected features. The final result of feature selection F_{HAFS} is :

$$F_{HAFS} = \alpha F_{GA} \tag{10}$$

Random forest based classification model

In the classification stage, the classifier used in this paper is random forest (RF) model. The main idea is: Select a subset S_b from all sample set S through randomly selecting sample features and sampling. Then a classification and regression tree is established for S_b. The classification and regression tree uses the Gini coefficient as the criterion [30]. In this paper, the above process was repeated 1000 times to construct 1000 CART trees to construct a random forest. The random forest based classification model is shown in Algorithm 3.

Input: Samples S; Features of the lesion F; Number of iterations m.
Output: Type of lesion predicted, P.
j=1;
m=1000;
while j $\leq$ m **do**
  Take S_b samples with Bootstrap from S.
  Randomly select F_s features from F_{HAFS}.
  Build a CART tree by S_b samples and F_s features.
  j=j+1;
end
For each new sample. Generate m results through the CART trees.
Determine the type of lesion by the principle of minority obeying the majority P.

Algorithm 3: Random Forest algorithm for COVID-19 Classification.

Acknowledgments
This work was supported by National Key R&D Program of China (2018YFC0830701), Fundamental Research Funds for the Central Universities (N2016006), Shenyang Medical Imaging Processing Engineering Technology Research Center (17-134-8-00) and the National Natural Science Foundation of China (No. U1708261).

Conflict of interest
There are no conflicts of interest declared.

Ethical Approval
This article does not contain any studies with human participants or animals performed by any of the authors.

Data Availability Statement

The patient population data used to support the findings of this study have not been made available because the data are supplied by Neusoft under license and so cannot be made freely available. Requests for access to these data should be made to the corresponding author.

Author details

[1]Key Laboratory of Intelligent Computing in Medical Image (MIIC), Northeastern University, Ministry of Education, China. [2]School of Computer Science and Engineering, Northeastern University, China. [3]Biomedical and Information Engineering School, Northeastern University, China. [4]Neusoft Medical System Co., Ltd., Liaoning, China.

References

1. Woloshin, S., Patel, N., Kesselheim, A.S.: False negative tests for sars-cov-2 infection—challenges and implications. New England Journal of Medicine (2020)
2. Li, X., Zeng, W., Li, X., Chen, H., Shi, L., Li, X., Xiang, H., Cao, Y., Chen, H., Liu, C., *et al.*: Ct imaging changes of corona virus disease 2019 (covid-19): a multi-center study in southwest china. Journal of translational medicine **18**, 1–8 (2020)
3. Shi, W., Peng, X., Liu, T., Cheng, Z., Lu, H., Yang, S., Zhang, J., Li, F., Wang, M., Zhang, X., et al.: : A retrospective study in 196 patients. http://dx.doi.org/10.2139/ssrn.3546089 (2020)
4. Barstugan, M., Ozkaya, U., Ozturk, S.: Coronavirus (covid-19) classification using ct images by machine learning methods. arXiv preprint arXiv:2003.09424 (2020)
5. Ozkaya, U., Ozturk, S., Barstugan, M.: Coronavirus (covid-19) classification using deep features fusion and ranking technique. arXiv preprint arXiv:2004.03698 (2020)
6. Elaziz, M.A., Hosny, K.M., Salah, A., Darwish, M.M., Lu, S., Sahlol, A.T.: New machine learning method for image-based diagnosis of covid-19. Plos one **15**(6), 0235187 (2020)
7. Tuncer, T., Dogan, S., Ozyurt, F.: An automated residual exemplar local binary pattern and iterative relieff based corona detection method using lung x-ray image. Chemometrics and Intelligent Laboratory Systems, 104054 (2020)
8. Zhou, T., Canu, S., Ruan, S.: An automatic covid-19 ct segmentation network using spatial and channel attention mechanism. arXiv preprint arXiv:2004.06673 (2020)
9. Khan, A.I., Shah, J.L., Bhat, M.M.: Coronet: A deep neural network for detection and diagnosis of covid-19 from chest x-ray images. Computer Methods and Programs in Biomedicine, 105581 (2020)
10. Afshar, P., Heidarian, S., Naderkhani, F., Oikonomou, A., Plataniotis, K.N., Mohammadi, A.: Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images. arXiv preprint arXiv:2004.02696 (2020)
11. Khalifa, N.E.M., Taha, M.H.N., Hassanien, A.E., Elghamrawy, S.: 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. arXiv preprint arXiv:2004.01184 (2020)
12. Minaee, S., Kafieh, R., Sonka, M., Yazdani, S., Soufi, G.J.: Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning. arXiv preprint arXiv:2004.09363 (2020)
13. He, K., Zhao, W., Xie, X., Ji, W., Liu, M., Tang, Z., Shi, F., Gao, Y., Liu, J., Zhang, J., et al.: Synergistic learning of lung lobe segmentation and hierarchical multi-instance classification for automated severity assessment of covid-19 in ct images. arXiv preprint arXiv:2005.03832 (2020)
14. Xu, X., Jiang, X., Ma, C., Du, P., Li, X., Lv, S., Yu, L., Ni, Q., Chen, Y., Su, J., et al.: A deep learning system to screen novel coronavirus disease 2019 pneumonia. Engineering (2020)
15. Zhao, D., Yao, F., Wang, L., Zheng, L., Gao, Y., Ye, J., Guo, F., Zhao, H., Gao, R.: A comparative study on the clinical features of covid-19 pneumonia to other pneumonias. Clinical Infectious Diseases (2020)
16. Zhao, W., Zhong, Z., Xie, X., Yu, Q., Liu, J.: Relation between chest ct findings and clinical conditions of coronavirus disease (covid-19) pneumonia: a multicenter study. American Journal of Roentgenology **214**(5), 1072–1077 (2020)
17. Kim, H., Park, C.M., Koh, J.M., Lee, S.M., Goo, J.M.: Pulmonary subsolid nodules: what radiologists need to know about the imaging features and management strategy. Diagnostic and Interventional Radiology **20**(1), 47 (2014)
18. Milletari, F., Navab, N., Ahmadi, S.-A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571 (2016). IEEE
19. Beamer, S., Asanovic, K., Patterson, D.: Direction-optimizing breadth-first search. In: SC'12: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, pp. 1–10 (2012). IEEE
20. He, H., Bai, Y., Garcia, E.A., Li, S.: Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), pp. 1322–1328 (2008). IEEE
21. Haralick, R.M., Shanmugam, K., Dinstein, I.H.: Textural features for image classification. IEEE Transactions on systems, man, and cybernetics (6), 610–621 (1973)
22. Mohanty, A.K., Beberta, S., Lenka, S.K.: Classifying benign and malignant mass using glcm and glrlm based texture features from mammogram. International Journal of Engineering Research and Applications **1**(3), 687–693 (2011)
23. Thibault, G., Fertil, B., Navarro, C., Pereira, S., Cau, P., Levy, N., Sequeira, J., Mari, J.-L.: Shape and texture indexes application to cell nuclei classification. International Journal of Pattern Recognition and Artificial Intelligence **27**(01), 1357002 (2013)
24. Sun, C., Wee, W.G.: Neighboring gray level dependence matrix for texture classification. computer vision, graphics, and image processing **23**(3), 341–352 (1983)
25. Van Griethuysen, J.J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R.G., Fillion-Robin, J.-C., Pieper, S., Aerts, H.J.: Computational radiomics system to decode the radiographic

phenotype. Cancer research **77**(21), 104–107 (2017)
26. Elssied, N.O.F., Ibrahim, O., Osman, A.H.: A novel feature selection based on one-way anova f-test for e-mail spam classification. Research Journal of Applied Sciences, Engineering and Technology **7**(3), 625–638 (2014)
27. Lin, C., Miller, T., Dligach, D., Plenge, R., Karlson, E., Savova, G.: Maximal information coefficient for feature selection for clinical document classification. ICML Workshop on Machine Learning for Clinical Data. Edingburgh, UK (2012)
28. Yan, K., Zhang, D.: Feature selection and analysis on correlated gas sensor data with recursive feature elimination. Sensors and Actuators B: Chemical **212**, 353–363 (2015)
29. Fonti, V., Belitser, E.: Feature selection using lasso. VU Amsterdam Research Paper in Business Analytics **30**, 1–25 (2017)
30. Menze, B.H., Kelm, B.M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., Hamprecht, F.A.: A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC bioinformatics **10**(1), 213 (2009)

Figures

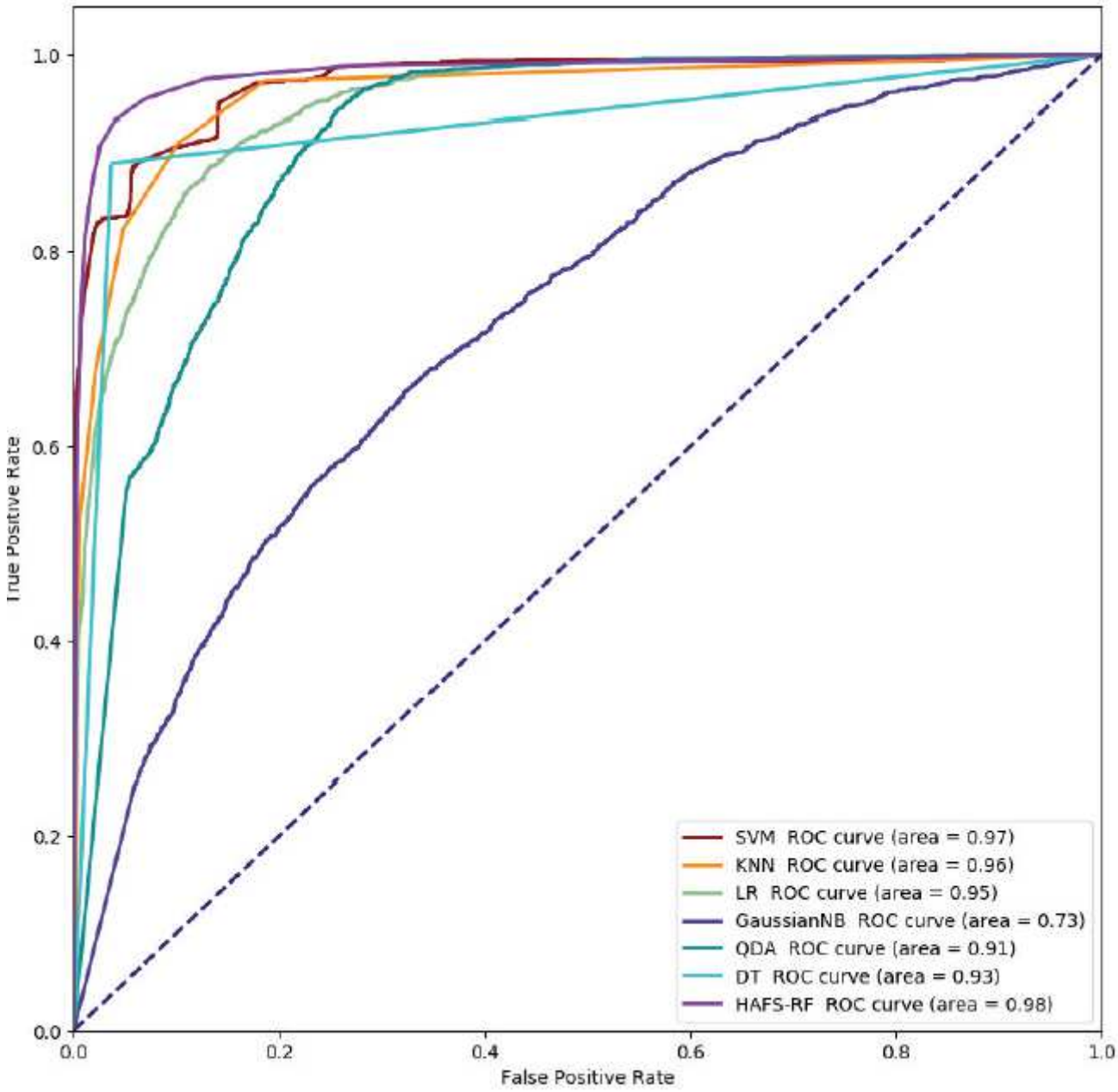

Figure 1

ROC curves achieved by different models.

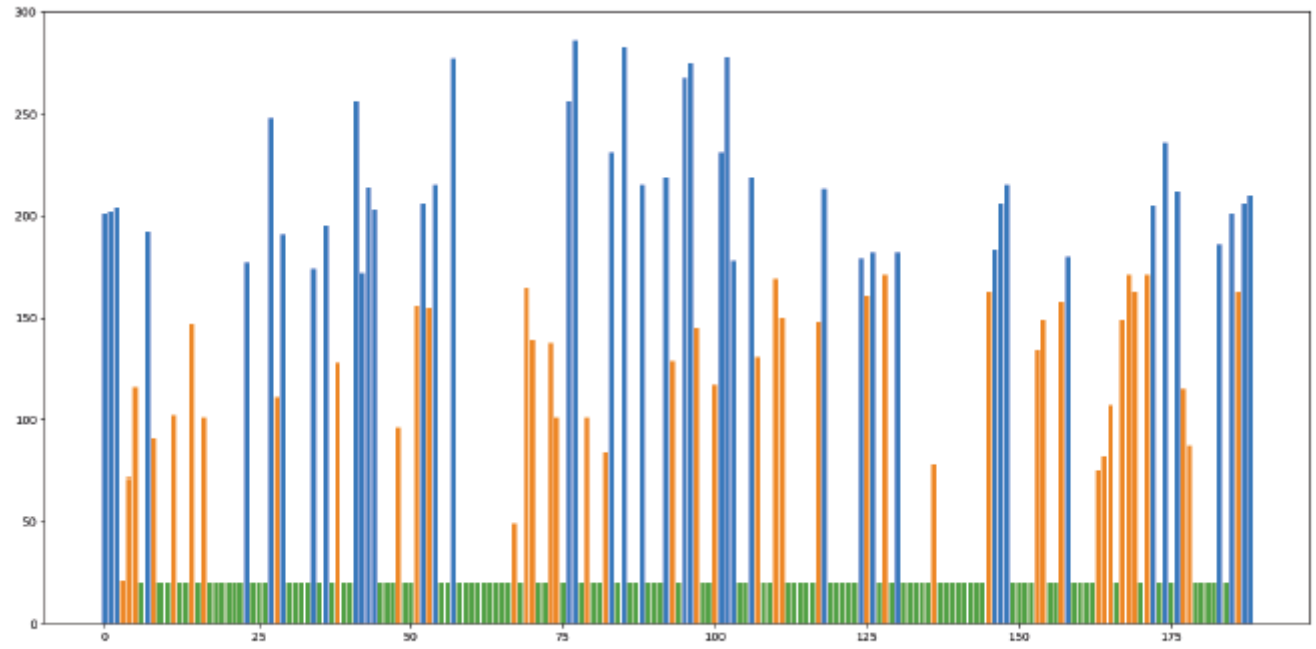

Figure 2

Scores of 189 features.

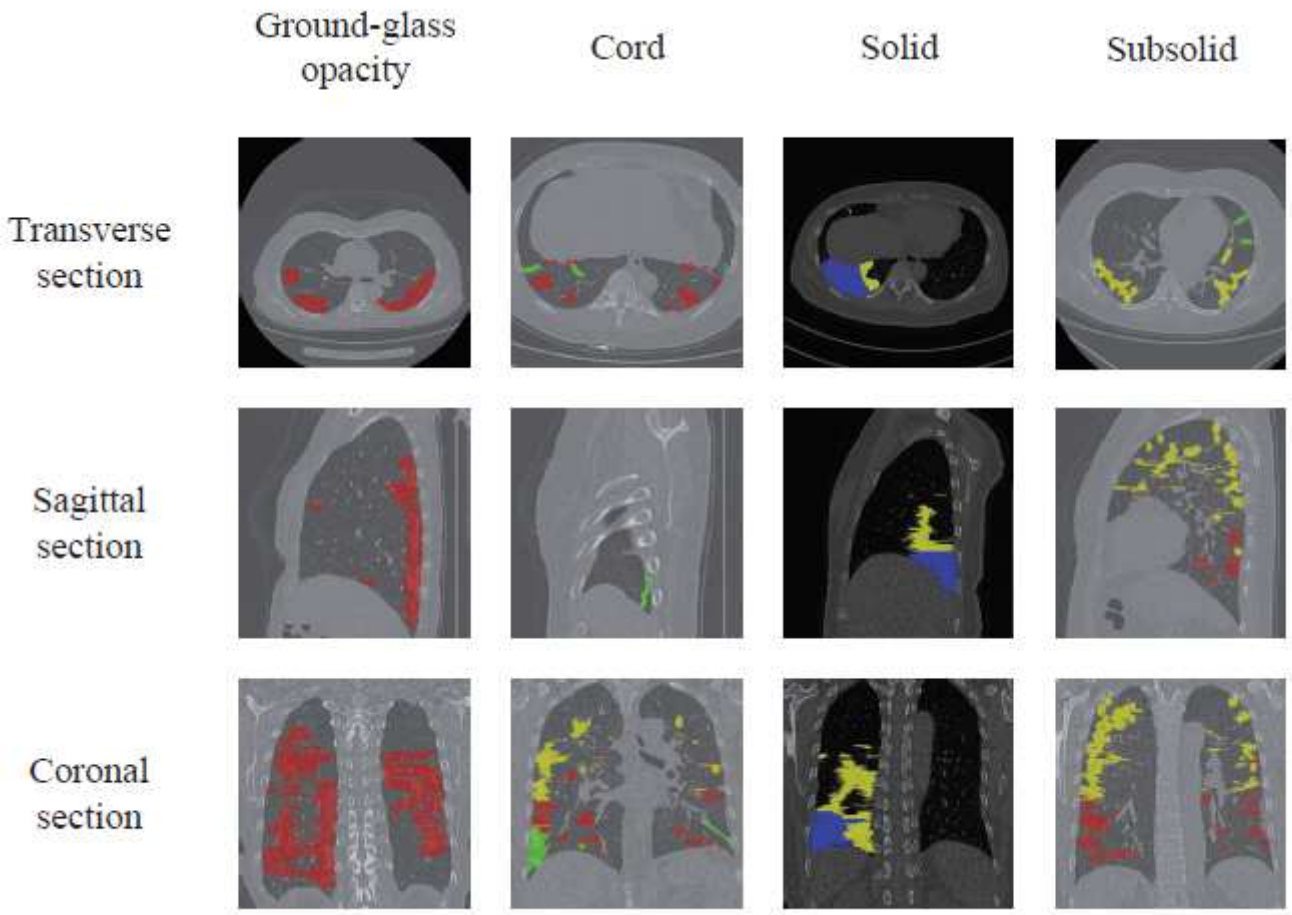

Figure 3

Typical four lesion subtypes in CT images of COVID-19. The labels and regions are given by medical experts. The red area represents ground-glass opacity, the green area represents cord, the blue area represents solid and the yellow area represents subsolid.

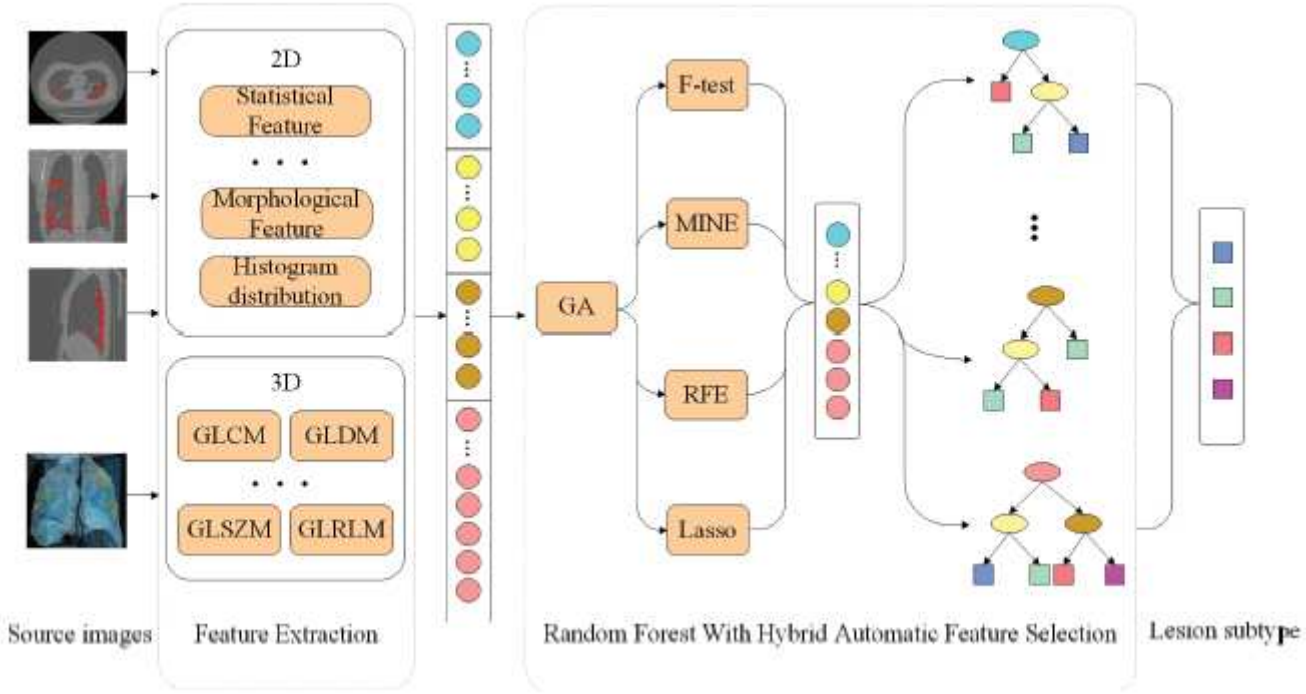

Figure 4

Overview of the COVID-19 Classication Using Random Forest Based on Hybrid Adaptive Feature Selection.

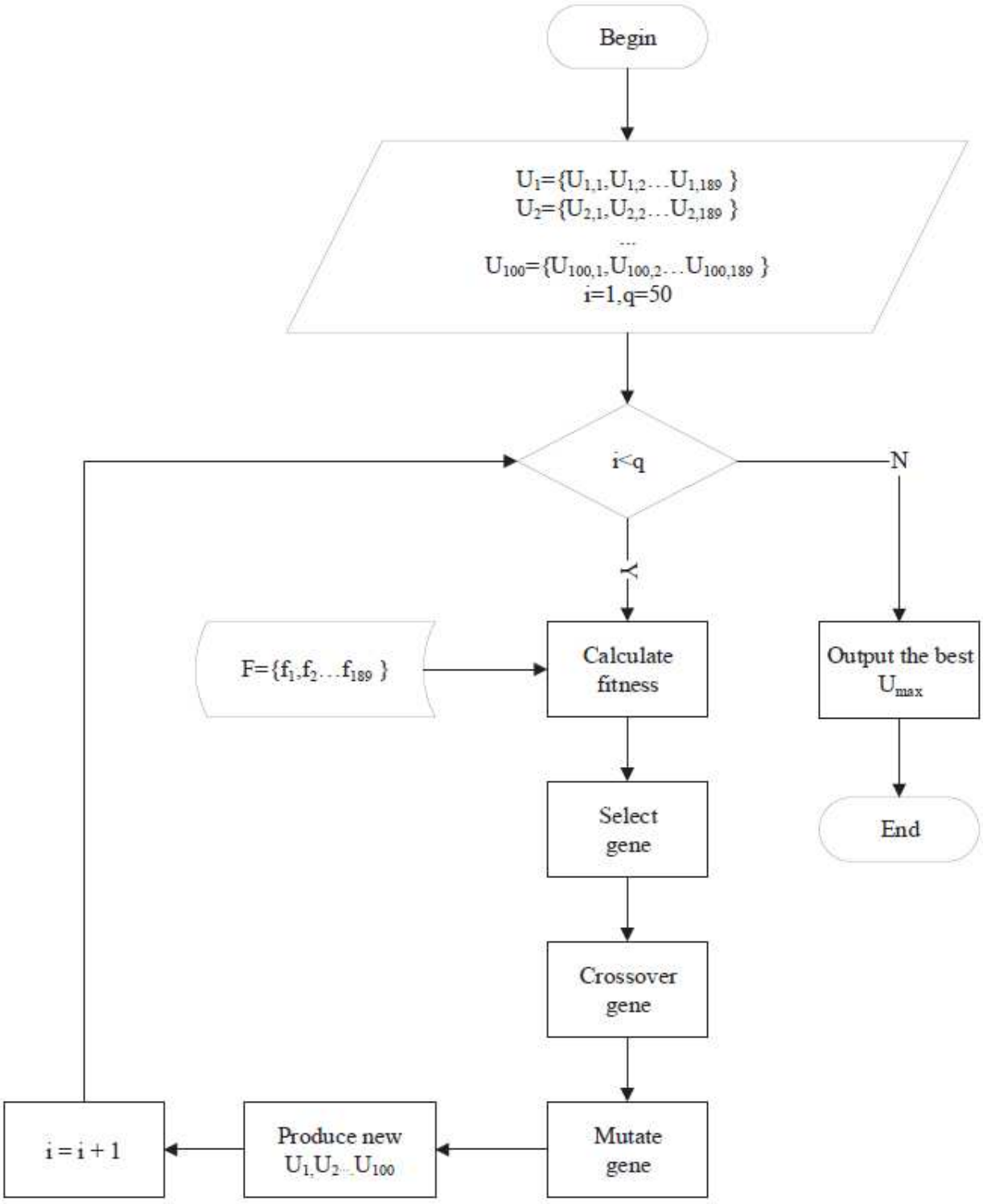

Figure 5

A schematic illustration of rst stage of HAFS.