Adaptive Distribution Learning with Statistical Hypothesis Testing for COVID-19 CT Scan Classification

[1]Guan-Lin Chen, [2]Chih-Chung Hsu, and [3]Mei-Hsuan Wu
Institute of Data Science, National Cheng Kung University
No.1, University Rd., Tainan City, Taiwan ROC.
[1]alright1117@gmail.com and {[2]cchsu,[3]re6091054}@gs.ncku.edu.tw

Abstract

With the massive damage in the world caused by Coronavirus Disease 2019 SARS-CoV-2 (COVID-19), many related research topics have been proposed in the past two years. The Chest Computed Tomography (CT) scan is the most valuable materials to diagnose the COVID-19 symptoms. However, most schemes for COVID-19 classification of Chest CT scan are based on single slice-level schemes, implying that the most critical CT slice should be selected from the original CT volume manually. In this paper, a statistical hypothesis test is adopted to the deep neural network to learn the implicit representation of CT slices. Specifically, we propose an Adaptive Distribution Learning with Statistical hypothesis Testing (ADLeaST) for COVID-19 CT scan classification can be used to judge the importance of each slice in CT scan and followed by adopting the non-parametric statistics method, Wilcoxon signed-rank test, to make predicted result explainable and stable. In this way, the impact of out-of-distribution (OOD) samples can be significantly reduced. Meanwhile, a self-attention mechanism without statistical analysis is also introduced into the backbone network to learn the importance of the slices explicitly. The extensive experiments show that both the proposed schemes are stable and superior. Our experiments also demonstrated that the proposed ADLeaST significantly outperforms the state-of-the-art methods.

1. Introduction

With the rapid growth of the deep learning approach recently, the performance of many research fields has been boosted with deep learning. One essential application among them is medical image analysis based on deep learning. The chest Computed Tomography (CT) scan is an effective way to trace the symptoms of Coronavirus Disease 2019 SARS-CoV-2 (COVID-19). However, both the analysis and diagnosis of CT scan series require an experienced doctor or expert in the related field. Unfortunately, it is hard to meet this requirement in remote areas. An automatic computer-aid diagnosis system for CT scan for COVID-19 is thus highly desired. In this paper, a explainable and effective model based on statistical test will be proposed for COVID-19 CT scan classification, as an example in Fig.1.

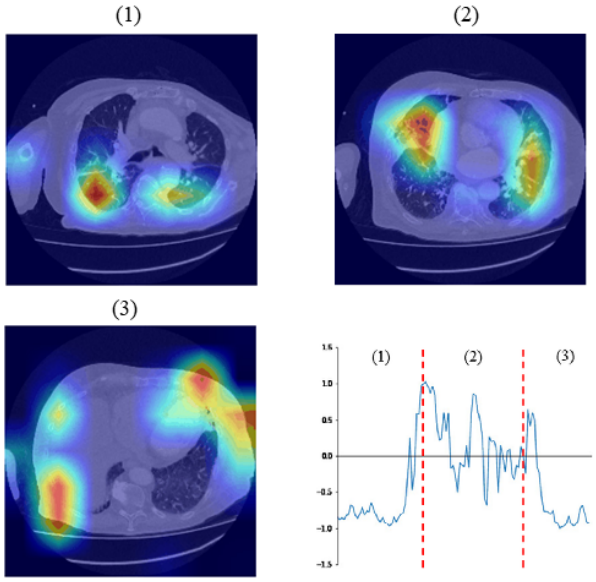


Figure 1: The visualized heatmaps based on Eigen-CAM[20] of the proposed adaptive distribution learning with statistical hypothesis test for a COVID-19 CT scan, where (1) indicates the front slice, (2) the middle part of the CT scan, and (3) the bottom slice.

The COVID-19 classification for CT scan is usually treated as a particular case of the image/video recognition tasks. In the past decade, deep learning has achieved state-of-the-art image recognition tasks compared to conventional machine learning and computer vision techniques. Similarly, deep learning-related schemes were widely adopted in the medical image field. However, the CT

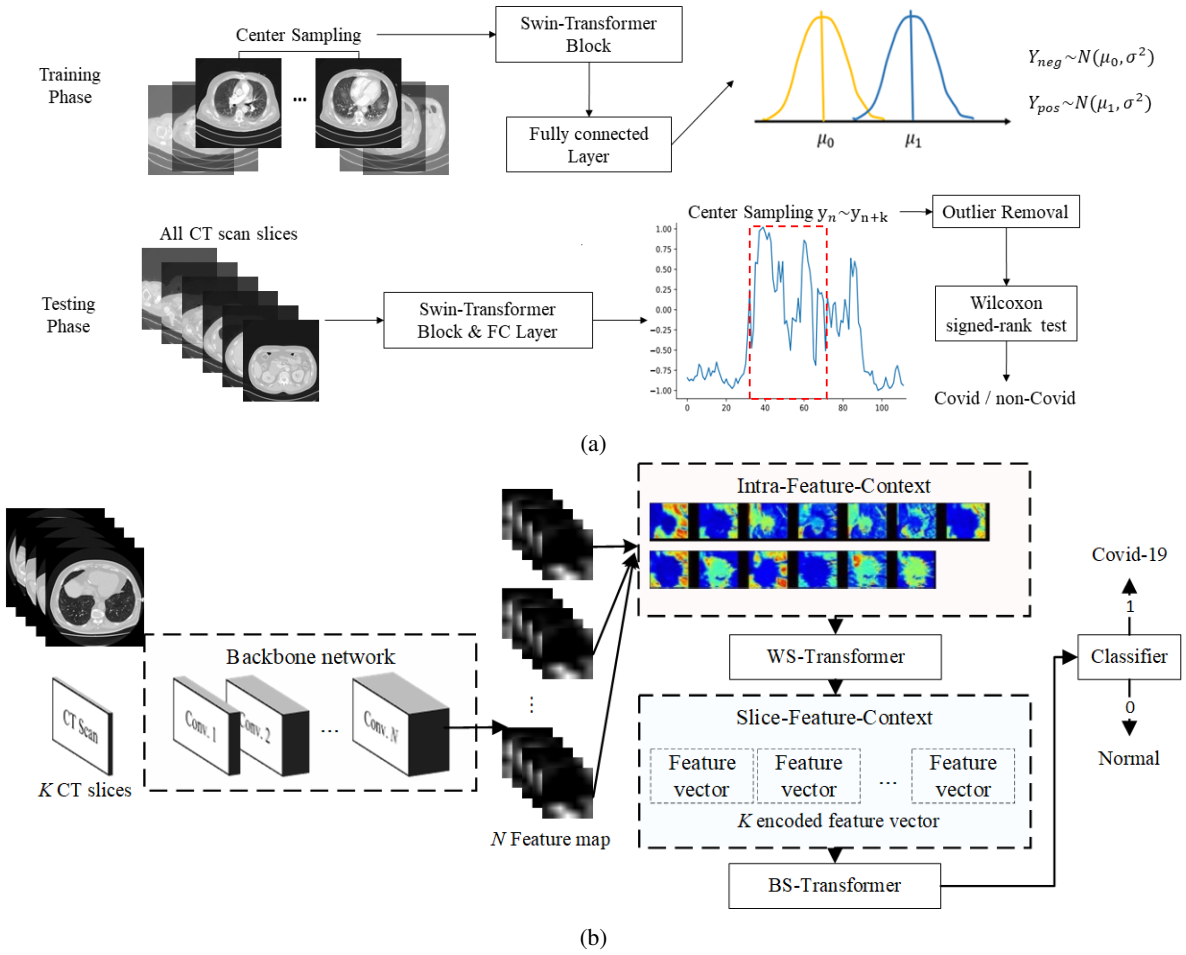


Figure 2: Flowchart of the proposed (a) ADLeaST using implicitly statistical inference for deep learning and (b) the explicit learning based on our CCAT for between- and within-slice-context context mining.

scan series is usually treated as a three-dimensional (3-D) data volume, where the traditional convolutional neural network (CNN) can only perform the two-dimensional (2-D) convolutional operation on the image. 3-D convolution was then adopted to tickle these issues [11][22][13][19]. However, the space and computational complexity of 3-D convolution is significantly higher than the 2-D one, and therefore high-end hardware is necessary. Another critical issue is that the large-scale dataset is the power source of deep neural networks. The insufficient training samples lead to a fatal overfitting issue in training a deep neural network, especially in 3-D convolutional neural networks.

Compared to the 2-D information-only approach (single-slice-based), the symptoms of COVID-19 might present at different depths (slice) for different patients, as suggested in [14]. Fortunately, a large-scale 3-D-shaped CT scan series, termed COV19-CT-DB, has been released in [9], including $5,000$ 3-D CT scans with more than $1,000$ patients. Compared to the well-known traditional 2-D CT dataset for COVID-19 classification [26], the COV19-CT-DB contains 3-D information providing more semantic features to help the diagnosis of COVID-19 symptoms. The conventional COVID-19 classification approach is usually based on a single slice-level approach [15][7][4][3][6][25][1]. It is hard to extend those 2-D models to deal with the 3-D (i.e., CT scan series) information without significant revision. The classification of COVID-19 based on a single slice can be treated as a conventional image recognition issue, as the suggestion in [7]. However, the performance usually relies on the large-scale training set, whereas medical data such as CT scans are relatively hard to collect. Therefore, several recent schemes are focused on dealing with small-scale issues. In [4], for instance, self-supervised learning was proposed to improve classification performance for a small-scale COVID-19 CT scan dataset. Other than the issue caused by the small-scale training set, the pixel-level annotation is very tedious and time-consuming. Therefore, recent research schemes were moved to focus on developing

a weakly-supervised learning approach to predict the pixel annotation (e.g., semantic segmentation) based on image-level annotation. In [6], weakly-supervised learning was introduced to tickle insufficient training samples and reduce the annotation requirements. However, the performance of those single slice-level models is restricted since the critical slice of a CT scan should be extracted by experienced experts.

In [10][12], the proposed approach can alleviate the catastrophic forgetting problem when another type of dataset related to the target disease has been used as the new training set. However, the 2-D and 3-D information are significantly different, implying that the performance for 3-D data with the model learned from 2-D data might not be promising. The CT scan-level information provided in COV19-CT-DB makes it hard to perform these existing single slice-level schemes to 3-D CT scan-level directly without significant revision. Recently, several schemes were focused on CT scan-level (3-D information) directly. In [22] adopted the multiple image-level CNNs to aggregate the predicted result of each CT image in a 3-D CT scan, while the 3D-CNN is directly adopted in [13] to extract the cube-like feature from a whole 3-D CT scan.s Both methods need considerable computational and space complexity to meet the model ensemble and 3-D convolution requirements. In [23], weakly-supervised learning was adopted, as a suggestion in [6], to achieve lesion localization with classification annotation only based on an improved 3-D U-shape Network (U-Net). In [19] and [9], the recurrent neural network (RNN) and long short-term memory (LSTM) network were proposed to integrate the cross-slice information to make the 3-D CT scan classification possible. However, it is well-known that the RNN and LSTM are hard to parallelize, leading to both training and inference time being hard to accelerate. Furthermore, 3-D convolution is significantly high computational complexity compared to 2-D convolution, leading to the fact that both training and inference costs are expensive. In addition, the number of the slices of different CT scan might be varied, implying that the input shape might be changed in the data pipeline. However, the number of the slices should be fixed in a 3-D convolutional neural network, implying that dealing with the changing number of the slices is impossible to tickle in a 3-D model.

Consider the computational complexity and effectiveness of CT scan-level classification of COVID-19, a maximum-likelihood estimation approach of Swin-Transformer [17] with statistical analysis for single slice-level classification is proposed to tickle this issue. A Convolutional CT scan-Aware Transformer (CCAT) for CT scan-level classification is also proposed to explore the slices' importance in this paper in an explicit way. Finally, comprehensive experiments are conducted to verify the effectiveness of the proposed two models.

The primary contribution of this paper is fourfold:

- To the best of our knowledge, the proposed Adaptive Distribution Learning with Statistic Test, termed ADLeaST, is the first approach to integrate statistical analysis and visual transformer. Our approach is stable, explainable, and effective for COVID-19 CT classification.
- Our ADLeaST is lightweight and without 3-D data processing, making the computational and space complexity relatively low.
- The proposed statistical hypothesis test can provide explainable prediction, beneficial for outlier removal and important slices selection, as shown in Fig.1.
- We also propose an auxiliary model, CCAT, to fully explore the context of slices and pixels by visual transformer, where the importance of each slice can be learned without statistical analysis.

The rest of this paper is organized as follows. Section II presents the proposed Adaptive Distribution Learning (ADL) and CCAT. In Section III, the superiority of the proposed method over peer methods is demonstrated. Finally, conclusions are drawn in Section IV.

2. The Proposed Method

2.1. Overview

In this paper, two schemes are proposed based on the 2-D (i.e., single slice-level) and 3-D (i.e., CT scan-level) respectively. In the first model, we propose to integrate CNNs with adaptive distribution learning for COVID-19 CT scan classification with a statistical test, termed ADLeaST, to explore the importance of each slice in implicitly statistical analysis. To verify the effectiveness of our statistical analysis strategy, we also explore the full 3-D information of a CT scan series based on context feature learning, termed CCAT. In our experiments, we will discuss the stability of the proposed ADLeaST and CCAT for the COVID-19 classification.

The flowchart of our ADLeaST model is illustrated in Fig. 2 (a). During clinical diagnosis, radiologists determine whether the chest CT image is positive (i.e., COVID-19) by the ground-glass opacities symptoms [14]. Since COV19-CT-DB [9] provides the raw slices for each CT scan, it is essential to indicate which slice is beneficial for training and testing phases. We proposed a novel adaptive distribution learning for slices in CT scan to deal with this issue. Initially, the feature representation of the positive and negative (i.e., COVID-19 and non-COVID-19) slices of the CT scan extracted using deep neural networks are mapped to different distributions. In this paper, Swin-Transformer [17] is

adopted as our backbone network since it is one of the state-of-the-art models for image recognition tasks. In this way, the important slices can converge to the mean of the target distribution since these slices account for the majority as well as their gradients show a similar direction. Moreover, the out-of-distribution (OOD) slices can be far from the target distribution mean because the target loss function is formed by randomly sampling from the target distribution. In this way, the impact of the performance of OOD samples and outliers will be minimized. Finally, slice-wise center crop and Wilcoxon signed-rank test [21] are adopted to generate the samples for training and testing phases, in which Wilcoxon signed-rank test can give meaning and explainable to the predicted results by statistical inference.

An auxiliary model is also proposed to learn the importance of slices from CT scans. The entire flowchart of our CCAT is referred to Fig. 2 (b). The key components of the proposed CCAT are the Within-Slice-Transformer (WST) and Between-Slice-Transformer (BST). The details of the proposed WST and BST will be described in the late subsection. In this WST, the training and testing CT scans $\mathbf{X}_t$ and $\mathbf{X}_v$ will be resized to a fixed size in the spatial domain as well as the L_s slices will be sampled from the original CT scan series $\mathbf{X_{ct}}$. Afterward, a conventional CNN (ResNet-50 [5] is used in this paper) is adopted to extract the feature maps $\mathbf{f} \in R^{c \times w_f \times h_f}$, where w_f and h_f indicate the width and height of c feature maps. The global averaging pooling (GAP) is discarded to preserve the spatial information of $\mathbf{f}$. BST is adopted to explore how to aggregate the feature maps to context-encoded feature vector f_{bst} based on the self-attention mechanism. While the L_s context-encoded features are aggregated $\mathbf{f}_t = [\mathbf{f}^0_{bst}, \mathbf{f}^1_{bst}, .., \mathbf{f}^{L_s}_{bst}]$, the proposed WST is then used to mine the context features between features of slices $\mathbf{f}_{wst}$. Finally, a three-layer perceptron with LeakyReLU activation is designed as the classifier.

2.2. Adaptive Distribution Learning

The proposed ADLeaST aims to tickle the slice importance selection using statistical analysis under the single-slice-level framework. First, Swin-Transformer [17] is adopted as the backbone network to have better feature representational power. In Swin-Transformer, a hierarchical feature representation was proposed starting from small-sized patches and gradually merging neighboring patches in deeper Transformer layers [17]. With these hierarchical feature maps, the Swin-Transformer model can conveniently leverage advanced techniques for dense prediction. In this paper, Swin-Transformer is adopted to generate the embedding feature vector for positive and negative CT scan slices, and followed by using a fully-connected layer to map their discrete features to different distribution functions. Given $i-th$ input CT scan slice $\mathbf{X_i} \in R^{c \times w \times h}$, Swin-Transformer Block f, and fully-connected layer $\mathbf{W}$, the generated random variable Y_i is:

$$Y_i = \mathbf{W} \cdot f(\mathbf{X_i}, \Theta) \tag{1}$$

where $Y_i \sim \mathcal{N}(\mu, \sigma^2)$. The Swin-Transformer Block and fully-connected layer will be trained jointly based on gradient descent. Our ADLeaST aims to map the positive and negative samples to different normal distributions to have sufficient discriminant.

2.2.1 Normal Likelihood Function

Given a training set $D = \{(\mathbf{X}_i, \mu_i)\}_{i=1}^n$ and the corresponding output samples $y_1, y_2, ..., y_n$ by (1). The probability density function of y_i is defined as:

$$f(y_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(-\frac{1}{2}\left(\frac{y_i - \mu_i}{\sigma}\right)^2\right)} \tag{2}$$

Then the likelihood function of the generated samples is as follows:

$$L(y_1, ..., y_n; \mu_i, \sigma^2) = (\frac{1}{\sigma\sqrt{2\pi}})^n e^{\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \mu_i)^2\right)} \tag{3}$$

Maximizing the log-likelihood function is equivalent to minimize the distance between the generated samples y_i and the normal distribution for positive or negative settings. This fashion can ensure that the target output sample y_i can converge to a normal distribution. To this end, we define the log-likelihood function as follows:

$$\begin{aligned} L_{log} &= -\frac{n}{2} log 2\pi - \frac{n}{2} log \sigma^2 \\ &\quad - \left(\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \mu_i)^2\right) \propto -\frac{1}{n}\sum_{i=1}^n (y_i - \mu_i)^2 \end{aligned} \tag{4}$$

The negative-side log-likelihood function L_{log} is proportional to MSE (mean square error) loss in the normal distribution assumption. In order to map the feature representation of the input CT scan slices to be asymptotically normal, MSE loss is adopted in our ADLeaST. In this case, we set the μ of negative sample is -1 while that of the positive sample is $+1$, and the σ^2 is empirically determined to be 0.2. Finally, the total loss of our ADLeaST will be given:

$$\begin{aligned} L_t &= \frac{1}{n_{neg}} \sum_{i=1}^{n_{neg}} \left(y_i - y^{(i)}_{neg}\right)^2 \\ &\quad + \frac{1}{n_{pos}} \sum_{j=1}^{n_{pos}} \left(y_j - y^{(j)}_{pos}\right)^2 \end{aligned} \tag{5}$$

where n_{neg} and n_{pos} are the batch size sampled from negative and positive CT scan, $y^{(i)}_{neg}$ and $y^{(j)}_{pos}$ are sampled from $\mathcal{N}(-1, 0.2)$ and $\mathcal{N}(1, 0.2)$ to prevent from overfitting the distribution mean.

2.2.2 Statistical Hypothesis Testing

In the inference phase, a slice-wise center cropping is proposed to sample the middle part of the CT scan. With the sampled slices, it is easy to generate the corresponding predicted values. The OOD $(\mu - 2\sigma, \mu + 2\sigma)$ can be treated as the outlier and will be removed. Afterward, Wilcoxon signed-rank test [21] is used on the remaining samples to determine whether the CT scan is COVID-19. The null hypothesis and alternative hypothesis are denoted as follows:

$$\begin{aligned} H_0 &: M_d \leq 0\,(Negative) \\ H_1 &: M_d > 0\,(Positive) \end{aligned} \tag{6}$$

We reject the null hypothesis when $p - value$ is less than the significance level α, implying that there is significance to verify the CT scan is with COVID-19 symptoms. The best sample size is decided empirically, and $\alpha = 0.05$ in our experiments.

2.3. Convolutional CT-Aware Transformer

The critical slices and outlier removal are well-addressed in our ADLeaST scheme. However, there might exist a slice selection strategy without statistical analysis. That is, the *substantial slices can be determined from end-to-end architecture*.

In this Section, a 3-D volume-based CNN is proposed, termed CCAT, to tickle both the slice selection and COVID-19 classification tasks. The slice importance can be learned via the proposed Within-Slice-Transformer (WST) and Between-Slice-Transformer (BST), as described in the following subsections.

2.3.1 Within-Slice-Transformer

Conventionally, the incoming of the Softmax classifier is based on simply averaging the feature maps in spatial dimension (i.e., Global averaging pooling), in which the spatial information is greatly reduced, leading to the fact that the context information in spatial dimension is missing.

Assumed that the training sample $\mathbf{X}_{ij} \in R^{c \times w \times h}$ is sampled from the original CT scan series $\mathbf{X_i}$. First, the single slice-level feature map is extracted based on the CNN backbone network (ResNet-50 [5] in this paper). Then, a visual transformer is adopted to fully discover the contextual information of the extracted feature map, as the middle part depicted in Fig. 2(b). Let the extracted feature map of $j-th$ slice of $i-th$ CT scan be $\mathbf{z}_{ij} \in R^{c \times w_z \times h_z}$, we re-arrange the feature map to $\mathbf{z}_{ijk}^T \in R^{s \times c}$, where $s = w_z/p_z \times h_z/p_z$, p_z is the size of $k-th$ patch. As suggested in [2], the positional encoding matrix $\mathbf{P}_k^w \in R^{s \times c}$ is used to embed the ordered information to the feature vector $\mathbf{z}_{ijk}^T$ such that

$$\mathbf{z}_{ijk}^{PE} = \mathbf{z}_{ijk}^T + \mathbf{P}_k^w \forall k. \tag{7}$$

Then, the within-slice-context features will be obtained via the multi-head attention (MSA) with the residual connection.

$$\mathbf{z}_{ijk}^{MSA} = MSA(LN(\mathbf{z}_{ijk-1}^{PE}) + \mathbf{z}_{ijk-1}^{PE}), \forall k, \tag{8}$$

where LN stands for layer normalization. Since the key of our CCAT is to explore the context of within and between slices, spatial gated multi-layer perceptron (gMLP) [16] is also adopted to capture the within-slice-context features, as follows:

$$\mathbf{z}_{ijk}^{gMLP} = gMLP(CN(\mathbf{z}_{ijk-1}^{PE}), \forall k, \tag{9}$$

where CN is the channel-wise normalization.

2.3.2 Between-Slice-Transformer

Since the fixed length of the sampled slices of the CT scan will significantly restrict the performance during the training phase, it is hard to know which slices are the most important. We randomly sample a set of slices from a CT scan to learn their between-slice-context feature to solve this issue. Let the extracted feature be $\mathbf{z}_i$ of $i-th$ slice of CT scan, we collect a set of feature vectors as $\mathbf{q}^i = [\mathbf{z}_0, \mathbf{z}_1, ..., \mathbf{z}_{L_s}]$, where L_s denotes the number of sampled slices. The BST will perform on the set of the extracted feature vectors as follows:

$$\mathbf{q}_{ik}^{MSA} = MSA(LN(\mathbf{q}_{ik-1}^{PE}) + \mathbf{q}_{ik-1}^{PE}), \forall k, \tag{10}$$

where $\mathbf{q}_{ik}^{PE} = \mathbf{q}_{ik} + \mathbf{P}_k^b$ and $k = 1, ..., L_s$. The gMLP variant will be given as follows:

$$\mathbf{q}_{ik}^{gMLP} = gMLP(CN(\mathbf{q}_{ik-1}^{PE}) + \mathbf{q}_{ik-1}^{PE}), \forall k, \tag{11}$$

Finally, the extracted between-slice-context feature $\mathbf{q}^{gMLP}$ will concatenate to a three-layer MLPs to classify the input CT sub-volume (a set of slices). In this way, the importance of each slice can be learned since our CCAT fully discover the context features in both the within-slice and between-slice dimensions.

3. Experimental Results

In this paper, the dataset used to evaluate the performance of the proposed approach is COV19-CT-DB [9]. In COV19-CT-DB, the training and validation set are partitioned by [9], where the number of training and validation CT scans are $1,560$ and 374, respectively. Since the annotation of the test set provided in [9] is unavailable, the validation set provided in [9] is used to evaluate the performance of the proposed methods.

In the training phase of our ADLeaST, the optimizer is AdamW[18], the learning rate and weight decays are $1e-5$ and 0.01, and the maximum epochs is 150. For training CCAT, the optimizer used in this paper is Adam [8], the initial learning rate is $1e-4$ and the learning decay is step scheduler with step size 20. The total epochs is 100.

3.1. Data pre-processing

3.1.1 ADLeaST

Due to some slices of CT scan might be useless for recognizing the COVID-19 (e.g., top/bottom slices might not contain chest information), we treat this case as the OOD samples. To reduce the influence of the outlier slices, selecting the CT scan in the training phase is essential, as well as in the evaluation phase. In the training phase, 40% slices in the center of the CT scan are sampled, then augmentation and normalization will be performed on these selected slices. To empirically determine the best fraction of the slices selection, we conduct a performance based on different sampling sizes, as illustrated in Fig. 3. As a result, $30\% \sim 60\%$ sample sizes in the center of CT scan make the best performance and therefore is suggested in our experiments. Despite 20% sampling size achieves the best area under the ROC curve (AUC), as shown in Fig. 4, a sufficient number of the sampled slices is suggested in this paper to ensure that the crucial slices can be preserved.

3.1.2 CCAT

Since the number of the slices in each CT scan is significantly different from each other, the number of the input slices should be fixed to meet the requirements of our CCAT. In this paper, the number of slice $L_s = 16$, and the sampling interval $L_{freq} = 2$. We randomly sample 16 slices in a CT scan to be input data of the proposed CCAT. Meanwhile, the common data augmentation schemes, including blurring, noise, random contrast and brightness, and optical distortion, are adopted in the training phase. Note that the random rotation and cropping are not performed on each slice separately since it will lead to unstable of the context of slices of a CT scan. Therefore, the random cropping and rotation are performed on the sampled 3-D volume instead of each slice. Each pixel value will be normalized to be ranged $[0, 1]$.

3.2. Performance evaluation

The evaluation metrics used in this paper are accuracy, sensitivity (SE), specificity (SP), macro-precision (P), macro-recall (R), and macro F1-score (F1). The performance comparison between the proposed and other peer methods is conducted in Table 1. It is clear that the proposed ADLeaST and CCAT significantly outperform the baseline model [9] and other state-of-the-art model [6]. Compare with the proposed ADLeaST and CCAT methods, the CCAT slightly outperforms ADLeaST on the validation set because of the WS- and BS-transformer in CCAT effectively extracting the context features from CT scan. However, the performance of CCAT declines in testing data since we did not carry out slices selected on CT scan. In contrast, the proposed ADLeaST tackles the OOD issues based on the proposed adaptive distribution learning, substantial slices selected, and Wilcoxon signed-rank test to make the predicted result more stable and reliable. It is noteworthy that the proposed ADLeaST has a very high specificity on the validation set because the type-I error is controlled in 0.05 (α), there is a lower probability of misclassifying the non-COVID-19 patient as COVID-19 since it needs to be statistically significant in probability to reject the null hypothesis in statistical hypothesis testing. In order to explain the stability of the ADLeaST, we make some examples of sequence samples generated by ADLeaST in a CT scan. As shown in Fig. 1 and Fig. 5, the slices in the center of a CT scan contains the precise structure of the chest, making the model could capture the ground-glass opacities or other symptoms of a COVID-19. Meanwhile, the statistic test is beneficial to determine whether most CT scan slices have this symptom by its probability, implying the predicted result is more stable and reliable. Finally, a model ensemble based on majority voting policy is adopted to fuse the predicted results of ADLeaST and CCAT to further improve the performance, as shown in the last row at Table 1.

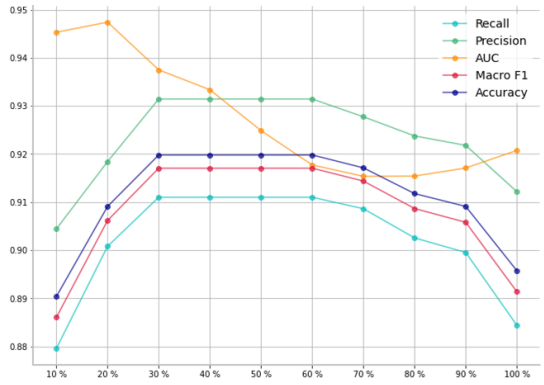


Figure 3: Performance comparison of the proposed ADLeaST with different sampling sizes.

3.3. Visualization and Discussion

As we claimed that the proposed ADLeaST should be able to robust to OOD cases, the visualization of the feature responses on slices is essential. Here, Eigen-CAM[20] is utilized to visualize the feature response of COVID-19 CT scan for our ADLeaST, as shown in Fig. 6 in different perspectives. The left and right sides of each sub-figure indicate the feature responses of negative and positive slices, respective. A higher response (more closer to the red one) higher confidence the model will be. In Fig. 6(a), it can be observed that the highest responses on the non-COVID-19 (left side) and COVID-19 with its corresponding symptoms

	Validation set						Testing set[3]		
	Acc.	SE	SP	P	R	F1	F1-C[1]	F1-NC[2]	F1
baseline [9]	0.724	0.388	0.952	0.731	0.688	0.700	0.5438	0.7962	0.6700
DenseNet201 [6]	0.732	0.455	0.947	0.714	0.703	0.708	–	–	–
Proposed ADLeaST	0.919	0.836	**0.986**	0.931	0.911	0.917	**0.8057**	**0.9675**	**0.8865**
Proposed CCAT	**0.933**	**0.897**	0.962	**0.935**	**0.929**	**0.932**	0.7080	0.9440	0.8260
Model Ensemble	**0.941**	0.885	**0.986**	**0.947**	**0.935**	**0.939**	**0.8063**	**0.9684**	**0.8874**

[1] COVID; [2] NON-COVID; [3] Evaluated by the official benchmark in [9].

Table 1: Performance evaluation of validation and testing set of the proposed methods and other peer methods in terms accuracy, sensitivity, specificity, macro-precision, macro-recall, macro-F1-score (results in blue indicates the implemented ourselves).

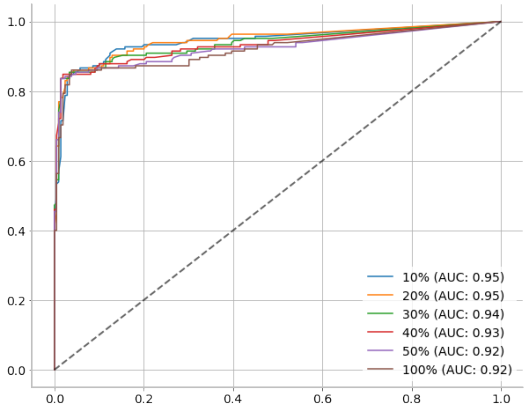


Figure 4: ROC curve measured on the validation set of the proposed ADLeaST with different testing sample sizes, where the threshold is the significant level α.

ADL	Train-S	Test-S	WSR-test	F1
				0.885
	✓			0.889
		✓		0.901
	✓	✓		0.896
✓				0.875
✓		✓		0.882
✓	✓			0.889
✓			✓	0.893
✓		✓	✓	0.893
✓	✓		✓	0.891
✓	✓	✓		0.912
✓	✓	✓	✓	**0.917**

Table 2: Ablation study of the proposed ADLeaST.

(right side) are correctly located at the lung tissue, implying that the proposed ADLeaST can learn the meaningful and explainable feature representation from slices of CT scan. Still, a few cases might be failures, as an example in Fig. 6(c), since the serious data OOD issue. When the number of OOD samples is increased, The predicted results could be disturbed, as shown in Fig. 6(b). However, the outlier could be removed based on the proposed statistical hypothesis test in the inference phase, as an example in Fig. 1. Furthermore, we compare the ordinary and failure cases in different anatomical plane of chest CT scan based on 3-D volumetric reconstruction in [24], as shown in Fig. 6(d)(e)(f). As the result, our ADLeaST model could successfully capture the meticulous ground-glass opacities symptoms of a COVID-19 in both the Coronal and Sagittal planes.

3.4. Ablation study

In this subsection, the ablation study is conducted to explore the influence of each part of the proposed ADLeaST model. The key components in the ADLeaST are listed as follows:

- The proposed adaptive distribution learning (ADL), or linear classifier with cross-entropy loss [17].
- Sampling 40% slices in the center of CT scan during the training phase.
- Sampling 40% slices in the center of CT scan during the testing phase.
- With/Without Wilcoxon signed-rank test for inference. The simple averaging strategy is adopted for the model without Wilcoxon signed-rank test to obtain the predicted result.

The result is shown in Table 2, our proposed adaptive distribution learning with Wilcoxon signed-rank test for inference achieves the best result. The testing sampling is also beneficial for all cases, showing that the inconsistent problem of the data distribution still remains. When the out-of-distribution data is included in the training phase, the perfor-

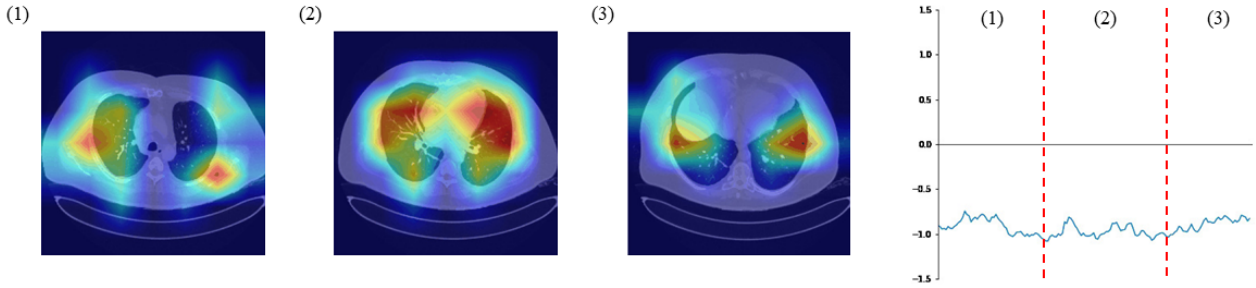
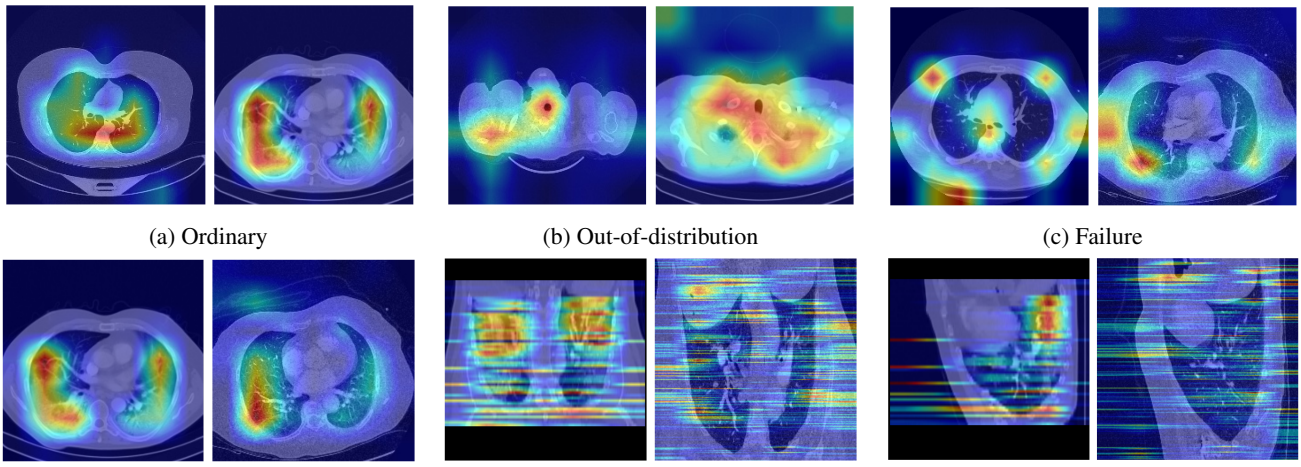


Figure 5: Time series charts of generated samples by the proposed ADLeaST and visualized feature response for the last layer of Swin-Transformer based on Eigen-CAM[20] for negative CT scan.



(a) Ordinary (b) Out-of-distribution (c) Failure

(d) Horizontal plane (e) Coronal plane (f) Sagittal plane

Figure 6: Visualized feature response for the proposed ADLeaST based on Eigen-CAM [20] for COVID-19 CT slices in (a) ordinary case, (b) OOD slices (no lung tissue presented), and (c) failure cases, where the left and right sides represent non-COVID and COVID-19 CT scan slices. (d) to (f) are examples of different anatomical plane CT scans, where the left and the right sides are ordinary and failure cases, respectively.

mance of the vanilla linear classifier with cross-entropy loss and our adaptive distribution learning is similar because it might not be appropriate to map the positive and negative samples to the target distributions due to higher variants in the training samples. However, it has an improvement for Macro-F1 with adaptive distribution learning when both training and testing sampling are taken. We can even use the Wilcoxon signed-rank test to give meaning and explainable to the predicted results.

4. Conclusion

This paper has proposed two deep neural networks for two-dimensional (2-D) and three-dimensional (3-D) CT scans for COVID-19 classification tasks. First, Adaptive distribution learning with statistical hypothesis testing for COVID-19 has been proposed to tickle the learning issue for out-of-distribution slices carefully. A nonparametric statistics with deep learning to make the predicted result more stable and explainable, finding a series of slices with the most significant symptoms in CT scan. Second, the auxiliary 3-D visual transformer has also been proposed in this paper based on the between- and within-slice-contexts, termed as CCAT (Convolutional CT scan-Aware Transformer), to automatically explore the intrinsic features in both slice and spatial dimensions. The visualization of the CT scan of the proposed models also verified that the critical insights of the symptoms caused by COVID-19 should be able to localize, as only CT scan-level annotation has been given. The extensive experiments have demonstrated that the proposed ADLeaST and CCAT significantly outperform the state-of-the-art methods for COVID-19 classification of CT scan. Our experiments also verified that the model with a statistical hypothesis test could significantly improve the stability and performance.

References

[1] Amine Amyar, Romain Modzelewski, Hua Li, and Su Ruan. Multi-task deep learning based ct imaging analysis for covid-19 pneumonia: Classification and segmentation. *Computers in Biology and Medicine*, 126:104037, 2020.

[2] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[3] El-Sayed M El-Kenawy, Abdelhameed Ibrahim, Seyedali Mirjalili, Marwa Metwally Eid, and Sherif E Hussein. Novel feature selection and voting classifier algorithms for covid-19 classification in ct images. *IEEE Access*, 8:179317–179335, 2020.

[4] Nicolas Ewen and Naimul Khan. Targeted self supervision for classification on a small covid-19 ct scan dataset. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1481–1485. IEEE, 2021.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[6] Shaoping Hu, Yuan Gao, Zhangming Niu, Yinghui Jiang, Lao Li, Xianglu Xiao, Minhao Wang, Evandro Fei Fang, Wade Menpes-Smith, Jun Xia, et al. Weakly supervised deep learning for covid-19 infection detection and classification from ct images. *IEEE Access*, 8:118869–118883, 2020.

[7] Aayush Jaiswal, Neha Gianchandani, Dilbag Singh, Vijay Kumar, and Manjit Kaur. Classification of the covid-19 infected patients using densenet201 based deep transfer learning. *Journal of Biomolecular Structure and Dynamics*, pages 1–8, 2020.

[8] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[9] Dimitrios Kollias, Anastasios Arsenos, Levon Soukissian, and Stefanos Kollias. Mia-cov19d: Covid-19 detection through 3-d chest ct image analysis. *arXiv preprint arXiv:2106.07524*, 2021.

[10] Dimitrios Kollias, N Bouas, Y Vlaxos, V Brillakis, M Seferis, Ilianna Kollia, Levon Sukissian, James Wingate, and S Kollias. Deep transparent prediction through latent representation analysis. *arXiv preprint arXiv:2009.07044*, 2020.

[11] Dimitrios Kollias, Athanasios Tagaris, Andreas Stafylopatis, Stefanos Kollias, and Georgios Tagaris. Deep neural architectures for prediction in healthcare. *Complex & Intelligent Systems*, 4(2):119–131, 2018.

[12] Dimitris Kollias, Y Vlaxos, M Seferis, Ilianna Kollia, Levon Sukissian, James Wingate, and Stefanos D Kollias. Transparent adaptation in deep medical image diagnosis. In *TAILOR*, pages 251–267, 2020.

[13] Debanjan Konar, Bijaya K Panigrahi, Siddhartha Bhattacharyya, Nilanjan Dey, and Richard Jiang. Auto-diagnosis of covid-19 using lung ct images with semi-supervised shallow learning network. *IEEE Access*, 9:28716–28728, 2021.

[14] Thomas C Kwee and Robert M Kwee. Chest ct in covid-19: what the radiologist needs to know. *RadioGraphics*, 40(7):1848–1865, 2020.

[15] Kunwei Li, Yijie Fang, Wenjuan Li, Cunxue Pan, Peixin Qin, Yinghua Zhong, Xueguo Liu, Mingqian Huang, Yuting Liao, and Shaolin Li. Ct image visual quantitative evaluation and clinical classification of coronavirus disease (covid-19). *European radiology*, 30(8):4407–4416, 2020.

[16] Hanxiao Liu, Zihang Dai, David R So, and Quoc V Le. Pay attention to mlps. *arXiv preprint arXiv:2105.08050*, 2021.

[17] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. *arXiv preprint arXiv:2103.14030*, 2021.

[18] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

[19] Ahmed Mohammed, Congcong Wang, Meng Zhao, Mohib Ullah, Rabia Naseem, Hao Wang, Marius Pedersen, and Faouzi Alaya Cheikh. Weakly-supervised network for detection of covid-19 in chest ct scans. *IEEE Access*, 8:155987–156000, 2020.

[20] Mohammed Bany Muhammad and Mohammed Yeasin. Eigen-cam: Class activation map using principal components. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2020.

[21] Denise Rey and Markus Neuhäuser. *Wilcoxon-Signed-Rank Test*, pages 1658–1659. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[22] Sertan Serte and Hasan Demirel. Deep learning for diagnosis of covid-19 using 3d ct scans. *Computers in Biology and Medicine*, 132:104306, 2021.

[23] Xinggang Wang, Xianbo Deng, Qing Fu, Qiang Zhou, Jiapei Feng, Hui Ma, Wenyu Liu, and Chuansheng Zheng. A weakly-supervised framework for covid-19 classification and lesion localization from chest ct. *IEEE transactions on medical imaging*, 39(8):2615–2625, 2020.

[24] Wikipedia contributors. Anatomical plane — Wikipedia, the free encyclopedia, 2021. [Online; accessed 16-July-2021].

[25] Yu-Huan Wu, Shang-Hua Gao, Jie Mei, Jun Xu, Deng-Ping Fan, Rong-Guo Zhang, and Ming-Ming Cheng. Jcs: An explainable covid-19 diagnosis system by joint classification and segmentation. *IEEE Transactions on Image Processing*, 30:3113–3126, 2021.

[26] Jinyu Zhao, Yichen Zhang, Xuehai He, and Pengtao Xie. Covid-ct-dataset: a ct scan dataset about covid-19. *arXiv preprint arXiv:2003.13865*, 2020.