

# Comprehensive Subset Selection for CT Volume Compression to Improve Pulmonary Disease Screening Efficiency

Qian Shao\*  
12221112@zju.edu.cn  
Zhejiang University  
Hangzhou, Zhejiang, China

Kai Zhang\*  
zhangkai1999@zju.edu.cn  
Zhejiang University  
Hangzhou, Zhejiang, China

Bang Du  
bangdu1994@zju.edu.cn  
Zhejiang University  
Hangzhou, Zhejiang, China

Zepeng Li  
lizepeng@zju.edu.cn  
Zhejiang University  
Hangzhou, Zhejiang, China

Yixuan Wu  
wux\_chloe@zju.edu.cn  
Zhejiang University  
Hangzhou, Zhejiang, China

Qiyuan Chen  
chenqiyuan1012@foxmail.com  
Zhejiang University  
Hangzhou, Zhejiang, China

Jian Wu<sup>†</sup>  
wujian2000@zju.edu.cn  
School of Public Health Zhejiang  
University  
Hangzhou, Zhejiang, China

Jintai Chen<sup>†</sup>  
jtchen721@gmail.com  
University of Illinois at  
Urbana-Champaign  
Champaign, Illinois, United States

## ABSTRACT

Deep learning models are widely used to process Computed Tomography (CT) data in the automated screening of pulmonary diseases, significantly reducing the workload of physicians. However, the three-dimensional nature of CT volumes involves an excessive number of voxels, which significantly increases the complexity of model processing. Previous screening approaches often overlook this issue, which undoubtedly reduces screening efficiency. Towards efficient and effective screening, we design a hierarchical approach to reduce the computational cost of pulmonary disease screening. The new approach re-organizes the screening workflows into three steps. First, we propose a Computed Tomography Volume Compression (CTVC) method to select a small slice subset that comprehensively represents the whole CT volume. Second, the selected CT slices are used to detect pulmonary diseases coarsely via a lightweight classification model. Third, an uncertainty measurement strategy is applied to identify samples with low diagnostic confidence, which are re-detected by radiologists. Experiments on two public pulmonary disease datasets demonstrate that our approach achieves comparable accuracy and recall while requiring approximately 4.5% of the time needed by the counterparts using full CT volumes. Besides, we also found that our approach outperforms previous cutting-edge CTVC methods in retaining important indications after compression.

\*Both authors contributed equally to this research.

<sup>†</sup>Corresponding Author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](https://permissions.acm.org).

*AIDSH 2024, August 25–29, 2018, Barcelona, Spain*

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-XXXX-X/18/06

<https://doi.org/XXXXXXXX.XXXXXXX>

## CCS CONCEPTS

• **Computing methodologies** → **Computer vision**.

## KEYWORDS

Pulmonary disease screening, Computed tomography, Comprehensive subset selection

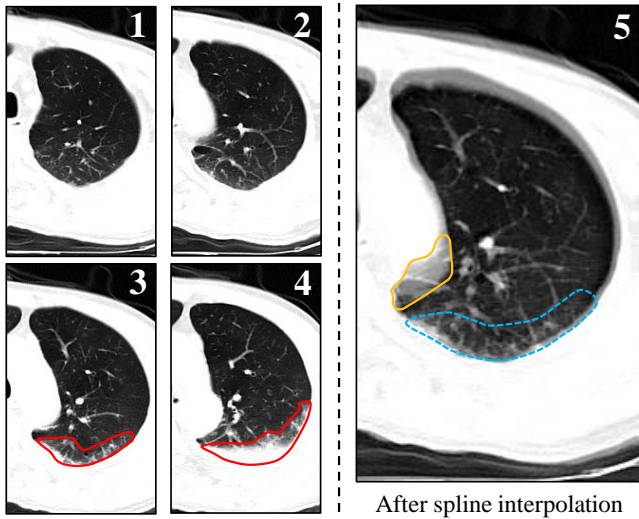
### ACM Reference Format:

Qian Shao, Kai Zhang, Bang Du, Zepeng Li, Yixuan Wu, Qiyuan Chen, Jian Wu, and Jintai Chen. 2024. Comprehensive Subset Selection for CT Volume Compression to Improve Pulmonary Disease Screening Efficiency. In *Proceedings of Artificial Intelligence and Data Science for Healthcare: Bridging Data-Centric AI and People-Centric Healthcare (AIDSH 2024)*. ACM, New York, NY, USA, 7 pages. <https://doi.org/XXXXXXXX.XXXXXXX>

## 1 INTRODUCTION

With the rapid advancements in medical imaging and deep learning algorithms, more researchers are utilizing deep learning methods to achieve expert-level disease screening, significantly reducing physicians' workloads [3]. Computed Tomography (CT) data are extensively used for diagnosing and screening various pulmonary diseases, such as lung carcinoma and pneumonia, due to CT's ability to reveal crucial details specific to individual patients [11]. However, pulmonary CT volumes, consisting of dozens or even hundreds of individual slices, present high computational complexity and require substantial memory capacity for processing—resources that are often limited in real-world medical settings.

To efficiently and accurately screen for pulmonary diseases, we design a novel screening approach. The core methodology insight is leveraging a Comprehensive Subset Selection (CSS) method to select representative slices for coarse detection, and then, during the inference phase, we devise an uncertainty strategy to identify samples with low diagnostic confidence. Radiologists will re-detect these samples. Our proposed CSS method ensures that the selected slices represent the entire CT volume, reducing the likelihood of omitting slices containing crucial indicators. While traditional slice



**Figure 1: Visualization results of interpolation-based methods.** Four pictures on the left represent four consecutive slices that contain CP lesions. The right picture illustrates the visualization results obtained from the spline interpolation. Within these visualizations, areas enclosed by red solid lines represent the CP lesions, while regions within yellow solid lines indicate artefacts generated after processing. And areas enclosed by blue dashed lines depict regions where lesions were eliminated after processing.

selection methods [26, 27] cannot always guarantee this. In addition, our method also avoids the problem of introducing artefacts that interpolation-based methods usually face [18, 5, 13], as shown in Fig. 1.

We conduct comparative experiments on two publicly available CT datasets to verify the accuracy and efficiency of our proposed screening approach, containing screening tasks for three types of pulmonary diseases, with a total of 2654 CT volumes. Specifically, we train classification networks using CT data compressed by different methods and compare their accuracy and recall. The experimental results demonstrate that our proposed CSS method significantly outperforms other methods. Furthermore, we explore the performance of the classification model under extreme conditions. For instance, when retaining only 4 slices from a single CT volume, the classification model still achieves a recall of 95.19% in the screening task for COVID-19. Based on this, we apply the uncertainty strategy to select low-confidence samples for re-detection, further improving the recall and meeting the clinical diagnosis standards. This proves that the uncertainty strategy screens out low-confidence cases that contain most of the misdiagnosed samples, which can provide reference value for actual disease screening tasks. Finally, we compare the inference time of different methods, validating the efficiency of our approach.

In summary, the main contributions of this article are as follows:

- (1) We design a novel hierarchical approach based on CT data for accurate and efficient screening of pulmonary diseases.
- (2) We propose a slice selection method CSS, which can select a representative slice subset from the whole CT volume, thereby enabling efficient screening.
- (3) We adopt an uncertainty strategy to select low-confidence samples for re-detection during the inference phase, ensuring the accuracy of the screening.
- (4) We validate the efficiency and accuracy of our proposed screening system on two public CT datasets. Experiments show that our screening method can achieve similar performance while requiring approximately 4.5% of the time needed by the method using full CT volumes.

## 2 RELATED WORK

Numerous studies have proposed various methods to process CT volumes for efficient screening of pulmonary diseases, which can be broadly categorized into three types:

*Patch-based methods.* These methods divide CT volumes into smaller patches for processing [8], which is similar to serializing parallel tasks. While this approach reduces the demand for large memory, it prolongs the inference time, resulting in inefficiency.

*Slice selection-based methods.* These methods select a subset of slices from the CT volume evenly to construct the desired volume [26, 27]. However, this approach lacks assurance that the selected slices represent the entire CT volume, as there is a high likelihood of omitting slices containing crucial indicators of the disease, leading to inaccurate screening results.

*Interpolation-based methods.* These methods adjust the Z-axis by either compressing or expanding it to reach the desired depth, such as Linear Interpolation (LI) [16, 18], Spline Interpolation (SI) [2, 5, 6], or Projection Interpolation (PI) [13]. However, these methods may distort the original pixel values, potentially introducing artefacts or losing lesion information, inevitably affecting screening accuracy.

## 3 METHODOLOGY

The overall process of our screening approach is divided into three steps, as shown in Fig. 2: First, we design CSS to select a slice subset that represents the entire CT volume. Second, we train a classification network to detect disease using the selected slices. Third, we adopt the uncertainty strategy to select low-confidence samples for re-detection.

### 3.1 Comprehensive Subset Selection

Assuming that each CT volume contains  $n$  slices (denoted as  $\{\mathbf{x}_i\}_{i=1}^n$ ), our goal is to select  $m (\ll n)$  slices that could represent the whole CT volume so that the performance of a model trained on the selected slice subsets approaches that on the whole CT volumes. The procedure of CSS is shown in Fig. 3.

*3.1.1 Representation Learning.* We leverage the pre-trained image feature extraction capabilities of MedCLIP-ViT [21] to obtain semantically meaningful representations containing key indicators. The image encoder of MedCLIP-ViT maps  $\mathbf{x}_i$  onto a  $d$ -dimensional hypersphere with  $L^2$  normalization, denoted as  $\mathbf{f}_i = f(\mathbf{x}_i)$ . Specifically, the [CLS] token features produced by the model's final output

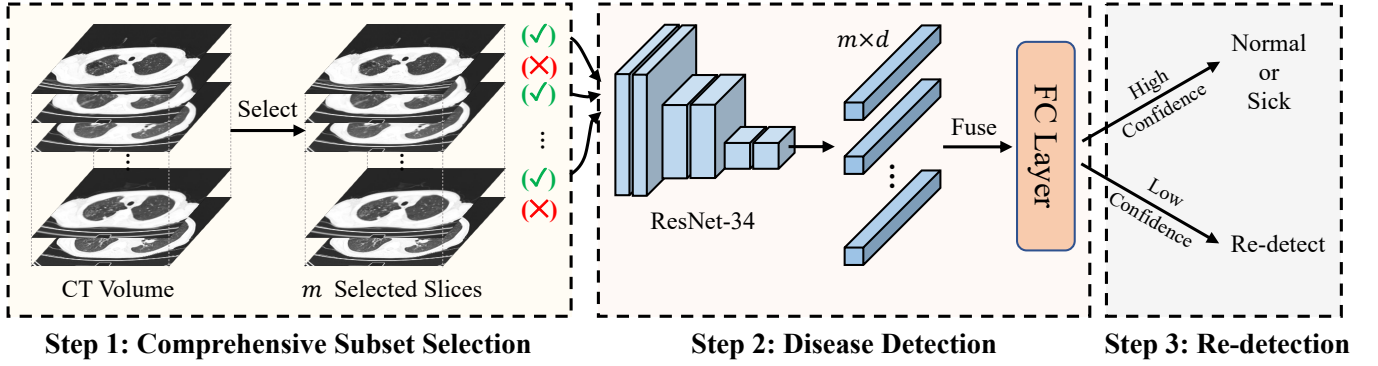


Figure 2: The overall process of our proposed screening approach.

with  $L^2$  normalization are employed for slice selection. It is noteworthy that although we utilize the larger model MedCLIP-ViT for feature extraction, which incurs additional computational resource consumption, this process is conducted before training the classification models. The efficiency should be evaluated based on the inference time of the models.

**3.1.2 Slice Selection.** We select CT slices according to two strategies that are representativeness and diversity [23, 22, 20, 25]. The former ensures that the selected slice subset can represent the whole CT volume containing key indicators. The latter minimizes the internal similarity of the selected subset to avoid redundancy, which provides the possibility of extremely compressing the CT volume.

First, we adopt  $K$ -Means clustering that partitions  $n$  instances (*i.e.*, the CT slices in the feature space, denoted as  $\{\mathbf{f}_i\}_{i=1}^n$ ) into  $m$  ( $m \leq n$ ) clusters, with each cluster represented by its centroid  $\mathbf{c}$  and every instance assigned to the cluster of the nearest centroid [4, 14]. Formally, we aim to find  $m$ -cluster partitioning  $\mathbb{S} = \{S_1, S_2, \dots, S_m\}$  that minimizes the within-cluster sum of squares [9]:

$$\min_{\mathbb{S}} \sum_{i=1}^m \sum_{\mathbf{f} \in S_i} \|\mathbf{f} - \mathbf{c}_i\|^2 = \min_{\mathbb{S}} \sum_{i=1}^m |S_i| \text{Var}(S_i). \quad (1)$$

It is optimized iteratively with EM [15] from random initial centroids.

Then, we select the density peaks within each cluster as the most representative slices since the density peaks are similar to as many instances as possible. The density is estimated by  $K$ -Nearest Neighbor ( $K$ -NN) [17] which is formulated as

$$\text{Den}(\mathbf{f}_i, k) = \frac{k}{n A_d \cdot D(\mathbf{f}_i, \mathbf{f}_i^k)}, \quad (2)$$

where  $A_d = \pi^{d/2} / \Gamma(\frac{d}{2} + 1)$  is the volume of a unit  $d$ -dimensional ball,  $d$  is the feature dimension,  $\Gamma(x)$  is the Gamma function, and  $D(\mathbf{f}_i, \mathbf{f}_i^k) = \|\mathbf{f}_i - \mathbf{f}_i^k\|$  is the feature distance between two instances,  $\mathbf{f}_i^k$  is the  $k$ th nearest neighbor of  $\mathbf{f}_i$ . Intuitively, the smaller the feature distance, the greater the similarity of the instances. However,  $\text{Den}(\cdot, \cdot)$  is very sensitive to noise, as it only takes the  $k$ th nearest neighbour into account. For robustness, we replace the  $k$ th neighbor distance  $D(\mathbf{f}_i, \mathbf{f}_i^k)$  with the average distance  $\bar{D}(\mathbf{f}_i, k)$  to all  $k$  nearest

neighbors instead:

$$\hat{\text{Den}}(\mathbf{f}_i, k) = \frac{k}{n A_d \cdot \bar{D}(\mathbf{f}_i, k)}, \quad (3)$$

where  $\bar{D}(\mathbf{f}_i, k) = \frac{1}{k} \sum_{j=1}^k D(\mathbf{f}_i, \mathbf{f}_i^j)$ . To sum up,  $\hat{\text{Den}}(\mathbf{f}_i, k)$  is used to measure the representativeness of instance  $\mathbf{f}_i$ . Since only the relative ordering matters in our selection process, the density peak corresponds to the instance with maximum  $\hat{\text{Den}}(\mathbf{f}_i, k)$ , *i.e.*, minimum  $\bar{D}(\mathbf{f}_i, k)$ .

However, when the clustering boundaries are located in high-density areas, the selected instances may align along these boundaries and become proximal to each other, leading to redundancy. Therefore, we apply a regularizer to diversify the selected instances in the feature space iteratively. In detail, let  $\hat{\mathbb{F}}^t = \{\hat{\mathbf{f}}_1^t, \hat{\mathbf{f}}_2^t, \dots, \hat{\mathbf{f}}_m^t\}$  denote the set of  $m$  instances selected at iteration  $t$ ,  $\hat{\mathbf{f}}_i^t$  is selected from clusters  $S_i$ , where  $i \in \{1, 2, \dots, m\}$ . For each candidate  $\mathbf{f}_i$  in cluster  $S_i$ , the farther it is away from those in other clusters in  $\hat{\mathbb{F}}^{t-1}$ , the more diversity it creates. We thus minimize the total inverse distance to others in a regularization loss  $\text{Reg}(\mathbf{f}_i, t)$ , with a sensitivity hyperparameter  $\alpha$ :

$$\text{Reg}(\mathbf{f}_i, t) = \sum_{j \neq i} \frac{1}{\|\mathbf{f}_i - \hat{\mathbf{f}}_j^{t-1}\|^\alpha}. \quad (4)$$

This regularizer is updated with an exponential moving average [10]:

$$\overline{\text{Reg}}(\mathbf{f}_i, t) = m_{\text{reg}} \cdot \overline{\text{Reg}}(\mathbf{f}_i, t-1) + (1 - m_{\text{reg}}) \cdot \text{Reg}(\mathbf{f}_i, t), \quad (5)$$

where  $m_{\text{reg}}$  is the momentum. At iteration  $t$ , we select instance  $i$  of the maximum combination of representativeness and diversity  $C(\mathbf{f}_i, t)$  within each cluster:

$$C(\mathbf{f}_i, t) = \frac{1}{\bar{D}(\mathbf{f}_i, k)} - \lambda \cdot \overline{\text{Reg}}(\mathbf{f}_i, t), \quad (6)$$

where  $\lambda$  is a hyperparameter that balances diversity and representativeness. At the last iteration, the instances in  $\hat{\mathbb{F}}$  are the selected CT slices for the downstream task.

## 3.2 Disease Detection

We train a classification model using the selected slices for disease detection. First, ResNet-34 [7] is used to extract image features of  $m$  slices in each case, denoted as  $f_{\text{Res}}(\mathbf{x}_i)$ , where  $i \in \{1, 2, \dots, m\}$ . Then, a light transformer-based network [19] is adopted to fuse

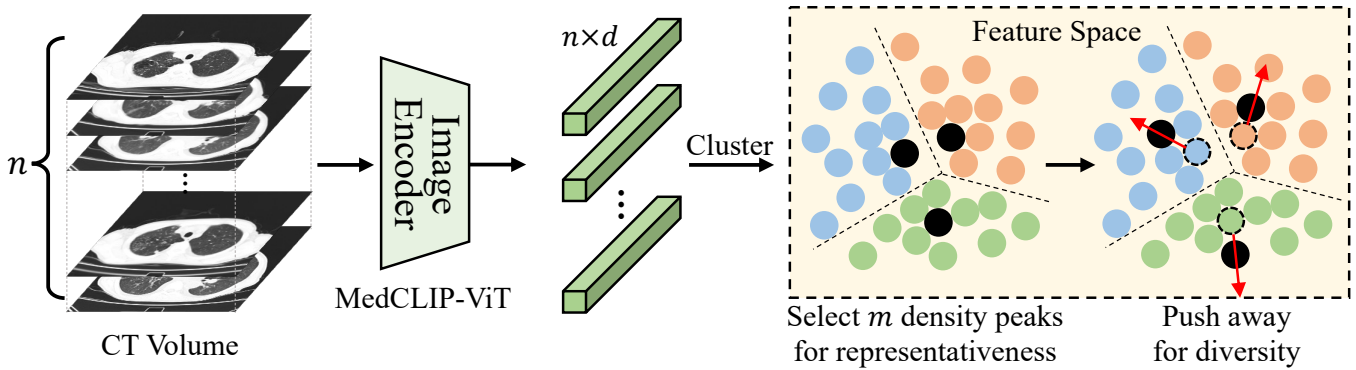


Figure 3: The procedure of CSS. The dark circles represent the selected slices.

these features. Finally, we perform a fully connected layer and Softmax on the output of the transformer-based network to predict the probability of disease. Details of the model can be found in Appendix A.

### 3.3 Low-confidence Case Re-detection

Low-confidence cases are re-detected using their corresponding full CT volumes by radiologists, specifically, those where the model’s predicted probability of disease is between 0.4 and 0.6. Although this step reduces some screening efficiency, it significantly improves accuracy to meet clinical requirements.

## 4 EXPERIMENTS

### 4.1 Dataset

We conduct comparative experiments on two publicly available datasets: SARS-CoV-2 dataset [24] and LUNG-PET-CT-Dx [12], which are obtained from the China Consortium of Chest CT Image Investigation (CC-CCII) and The Cancer Imaging Archive (TCIA) [1], respectively. Given our objective to compress CT volumes, we eliminated cases with a relatively low number of slices, specifically those with less than 64 slices. After that, 2654 cases are included in the study, containing 311 cases diagnosed with adenocarcinoma (AC), 747 cases with Novel Coronavirus Pneumonia (NCP, *i.e.*, COVID-19), 741 cases with Common Pneumonia (CP) and 855 normal cases. Here, a case refers to a 3D CT volume. The AC cases are from LUNG-PET-CT-Dx, while the NCP, CP and normal cases are from the SARS-CoV-2 dataset. Then, we design three binary classification tasks to detect the above three diseases. Specifically, Task 1 contains 747 NCP and 855 normal cases; Task 2 contains 741 CP and 855 normal cases; Task 3 contains 311 AC and 311 normal cases. The normal cases for Task 3 are randomly chosen from 855 normal cases of the SARS-CoV-2 dataset. Finally, we evenly divide the dataset into 5 parts and use 5-fold cross-validation to evaluate the accuracy and recall for each task.

### 4.2 Experimental Settings

All models are implemented in PyTorch and trained on an RTX 4090 with 24 GB memory. Considering the anatomical structure of the lungs—where the left lung is divided into upper and lower lobes with a volume ratio of approximately 0.4 : 0.6, and the right

lung into upper, middle, and lower lobes with a volume ratio of approximately 0.25 : 0.15 : 0.6, we divide the entire CT volume into three distinct parts based on the ratios of 0.25 : 0.15 : 0.6. Following this, the total budget of slice selection  $m$  is distributed across these parts in accordance with the same proportion. CSS is used to select a slice subset in each part. The hyperparameters set for CSS and training the diagnostic model can be found in Appendix B. In the efficiency evaluation experiments, we compared the inference times of different methods while ensuring the same memory usage.

### 4.3 Comparison with Other Compression Methods

In three diagnostic tasks, we compare the performance of diagnostic models when applied with different compression methods, including three interpolation-based methods: Projection Interpolation (PI) [13], Linear Interpolation (LI) [16] and Spline Interpolation (SI) [2], two slice selection-based methods: Subset Slice Selection (SSS) [26] and Even Slice Selection (ESS) [27], and full CT volume method. The implementation of the above methods can be found in Appendix C.

To ensure a fair comparison, we set the number of slices after compression to 64 for each method. The experimental results are shown in Table 1, from which we have several observations: (1) The performance of our method generally surpasses that of other compression methods, and it achieves performance levels very close to those of the full CT volume method across three tasks; (2) On Tasks 1 and 2, the performance of the second-best SI method is comparable to that of our method. On task 3, only our method achieves accuracy and recall above 90%. This may be attributed to the fact that Tasks 1 and 2 involve diagnosing pneumonia, where the lesion areas are larger and more distinct, thus easier to differentiate from normal cases. In contrast, Task 3 focuses on detecting adenocarcinoma, where the lesion areas are smaller and morphologically complex, making accurate detection challenging. The superior performance on difficult tasks demonstrates the superiority of our method.

### 4.4 Exploration Under Extreme Compression

We explore the accuracy and recall of the diagnostic model applied with CSS under extreme compression, that is, selecting 32, 16, 8, and 4 CT slices from each case. We compare CSS not only with the

**Table 1: Comparison with other methods. The best performance is bold, and the second best performance is underlined except for the method of full CT volume. Each result shows mean accuracy and standard deviation over 5-fold cross-validation.**

Method	Task 1		Task 2		Task 3	
	Accuracy	Recall	Accuracy	Recall	Accuracy	Recall
Full CT	99.12 $\pm$ 0.31	99.15 $\pm$ 0.35	99.35 $\pm$ 0.41	99.30 $\pm$ 0.38	92.47 $\pm$ 0.60	92.10 $\pm$ 0.65
PI [13]	90.36 $\pm$ 0.34	89.02 $\pm$ 0.41	88.92 $\pm$ 0.40	89.00 $\pm$ 0.30	84.59 $\pm$ 0.47	83.50 $\pm$ 0.52
LI [16]	78.52 $\pm$ 2.54	77.89 $\pm$ 2.67	73.67 $\pm$ 2.03	71.70 $\pm$ 2.30	68.81 $\pm$ 3.80	67.54 $\pm$ 4.81
SI [2]	<u>98.78<math>\pm</math>0.34</u>	<u>98.20<math>\pm</math>0.37</u>	<u>98.33<math>\pm</math>0.42</u>	<u>98.30<math>\pm</math>0.40</u>	87.24 $\pm$ 0.40	86.94 $\pm$ 0.37
SSS [26]	96.89 $\pm$ 0.56	96.28 $\pm$ 0.62	95.23 $\pm$ 0.97	95.12 $\pm$ 0.99	<u>89.34<math>\pm</math>0.90</u>	89.67 $\pm$ 1.24
ESS [27]	96.42 $\pm$ 0.86	96.85 $\pm$ 0.78	94.29 $\pm$ 1.37	94.50 $\pm$ 1.70	88.56 $\pm$ 1.56	86.78 $\pm$ 1.65
CSS (Ours)	<b>98.89<math>\pm</math>0.17</b>	<b>99.05<math>\pm</math>0.18</b>	<b>98.65<math>\pm</math>0.23</b>	<b>98.82<math>\pm</math>0.29</b>	<b>91.88<math>\pm</math>0.43</b>	<b>90.85<math>\pm</math>0.54</b>

**Table 2: Comparison under extreme compression. The best performance under different compression levels is bold. Each result shows mean accuracy and standard deviation over 5-fold cross-validation.**

Number	Method	Task 1		Task 2		Task 3	
		Accuracy	Recall	Accuracy	Recall	Accuracy	Recall
32	SI	97.99 $\pm$ 0.85	98.06 $\pm$ 0.54	96.73 $\pm$ 0.95	97.10 $\pm$ 0.79	83.63 $\pm$ 1.24	83.49 $\pm$ 1.33
	CSS	<b>98.12<math>\pm</math>0.55</b>	<b>98.19<math>\pm</math>0.36</b>	<b>97.89<math>\pm</math>0.69</b>	<b>97.74<math>\pm</math>0.77</b>	<b>90.01<math>\pm</math>0.97</b>	<b>90.27<math>\pm</math>0.81</b>
16	SI	96.99 $\pm$ 1.55	97.21 $\pm$ 1.10	93.87 $\pm$ 2.27	94.07 $\pm$ 2.03	76.71 $\pm$ 2.70	76.90 $\pm$ 2.56
	CSS	<b>97.26<math>\pm</math>1.54</b>	<b>97.58<math>\pm</math>1.21</b>	<b>94.00<math>\pm</math>2.15</b>	<b>94.28<math>\pm</math>2.04</b>	<b>85.60<math>\pm</math>2.37</b>	<b>85.54<math>\pm</math>2.40</b>
8	SI	95.96 $\pm$ 1.19	95.88 $\pm$ 1.22	92.09 $\pm$ 2.77	92.55 $\pm$ 2.49	67.34 $\pm$ 3.83	67.68 $\pm$ 3.41
	CSS	<b>96.57<math>\pm</math>1.24</b>	<b>96.89<math>\pm</math>1.27</b>	<b>93.21<math>\pm</math>3.72</b>	<b>93.02<math>\pm</math>3.44</b>	<b>82.55<math>\pm</math>2.32</b>	<b>82.63<math>\pm</math>2.17</b>
4	SI	93.86 $\pm$ 2.24	94.24 $\pm$ 1.89	90.03 $\pm$ 2.25	90.78 $\pm$ 2.30	64.41 $\pm$ 3.70	64.57 $\pm$ 3.54
	CSS	<b>94.90<math>\pm</math>2.12</b>	<b>95.19<math>\pm</math>2.01</b>	<b>91.17<math>\pm</math>2.54</b>	<b>91.76<math>\pm</math>2.51</b>	<b>78.64<math>\pm</math>3.19</b>	<b>78.89<math>\pm</math>2.98</b>
Full CT		99.12 $\pm$ 0.31	99.15 $\pm$ 0.35	99.35 $\pm$ 0.41	99.30 $\pm$ 0.38	92.47 $\pm$ 0.60	92.10 $\pm$ 0.65

full CT volume method but also with the second-best performing method described in Section 4.3, *i.e.*, SI method [2]. The experimental results are shown in Table 2, from which we have several observations: (1) On Tasks 1 and 2, the performance of models applied with two compression methods is similar, while on Task 3, our method shows a significant improvement over the SI method. This conclusion aligns with the second point discussed in Sec. 4.3; (2) On Task 3, as the number of target slices after compression decreases, the performance improvement of our method becomes increasingly significant compared to the SI method. For instance, at *Number* = 32 and 4, the recall increased by 6.78% and 14.32%, respectively, demonstrating that under extreme compression scenarios, our method holds a greater advantage over other methods.

#### 4.5 Comparison Before and After Re-detection

When selecting 8 and 4 CT slices, we verify the effectiveness of the uncertainty strategy by comparing the performance before and after the re-detection. We assume all low-confidence cases will be correctly diagnosed during the re-detection phase. The experimental results are shown in Table 3, from which we observe that the accuracy and recall improve after using the uncertainty strategy to filter out cases for re-detection. In particular, on Task 3, recall improves by 6.17% and 4.14% when retaining 8 and 4 slices,

respectively. This proves that low-confidence cases do contain most of the misdiagnosed cases, which can provide reference value for actual disease screening tasks.

**Table 3: Comparison before and after re-detection using uncertainty strategy.**

Task	Number	Re-detection	Accuracy	Recall
Task 1	8	Before	96.57	96.89
		After	97.24( $\uparrow$ 0.67)	97.88( $\uparrow$ 0.99)
	4	Before	94.90	95.19
		After	96.06( $\uparrow$ 1.16)	96.41( $\uparrow$ 1.22)
Task 2	8	Before	93.21	93.02
		After	94.27( $\uparrow$ 1.06)	94.59( $\uparrow$ 1.57)
	4	Before	91.17	91.76
		After	93.04( $\uparrow$ 1.87)	93.18( $\uparrow$ 1.42)
Task 3	8	Before	82.55	82.63
		After	88.53( $\uparrow$ 5.98)	88.80( $\uparrow$ 6.17)
	4	Before	78.64	78.89
		After	82.77( $\uparrow$ 4.13)	83.03( $\uparrow$ 4.14)

## 4.6 Efficiency Evaluation

We compared the inference time of CSS and the full CT method to evaluate the efficiency under different compression rates. The results are shown in Table 4, from which we can see that when retaining only 4 slices per CT volume, the inference time of our method is approximately 4.5% of that of the full CT approach.

**Table 4: Inference time under different compression ratios. The time shown in the table is the inference time of 50 cases.**

Method	Number	Task 1	Task 2	Task 3
CSS	32	38.74s	38.91s	38.70s
	16	20.40s	21.27s	20.82s
	8	11.39s	12.06s	11.58s
	4	7.69s	8.55s	8.04s
Full CT		169.24s	189.05s	179.25s

## 5 CONCLUSION

In this paper, we propose a novel hierarchical approach based on CT data for accurate and efficient screening of pulmonary diseases. The core insight is leveraging a CTVC method CSS to select representative slices from full CT volumes for coarse detection. During the inference phase, we adopt an uncertainty strategy to identify cases with low diagnostic confidence, which can be referred to radiologists for re-detection in clinical practice. Extensive experimental results show that our approach achieves comparable performance while requiring approximately 4.5% of the time the counterparts need to process full CT volumes. Besides, we also found that our approach outperforms previous SOTA CTVC methods in retaining crucial information after compression.

## REFERENCES

- [1] Kenneth Clark et al. 2013. The cancer imaging archive (tcia): maintaining and operating a public information repository. *Journal of digital imaging*, 26, 1045–1057.
- [2] Carl De Boor. 1962. Bicubic spline interpolation. *Journal of mathematics and physics*, 41, 1–4, 212–218.
- [3] Vijaypal Singh Dhaka, Geeta Rani, Meet Ganpatlal Oza, Tarushi Sharma, and Ankit Misra. 2021. A deep learning model for mass screening of covid-19. *International journal of imaging systems and technology*, 31, 2, 483–498.
- [4] Edward W Forgy. 1965. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *biometrics*, 21, 768–769.
- [5] Dong Han et al. 2023. Development and validation of a 3d-convolutional neural network model based on chest ct for differentiating active pulmonary tuberculosis from community-acquired pneumonia. *La radiologia medica*, 128, 1, 68–80.
- [6] Jinkui Hao et al. 2022. Uncertainty-guided graph attention network for paraneumonic effusion diagnosis. *Medical Image Analysis*, 75, 102217.
- [7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- [8] Thijs Kooi, Geert Litjens, Bram Van Ginneken, Albert Gubern-Mérida, Clara I Sánchez, Ritse Mann, Ard den Heeten, and Nico Karssemeijer. 2017. Large scale deep learning for computer aided detection of mammographic lesions. *Medical image analysis*, 35, 303–312.
- [9] Hans-Peter Kriegel, Erich Schubert, and Arthur Zimek. 2017. The (black) art of runtime evaluation: are we comparing algorithms or implementations? *Knowledge and Information Systems*, 52, 341–378.
- [10] Samuli Laine and Timo Aila. 2016. Temporal ensembling for semi-supervised learning. In *International Conference on Learning Representations*.

- [11] Ashia Lewis, Evanjin Mahmoodi, Yuyue Zhou, Megan Coffee, and Elena Sizikova. 2021. Improving tuberculosis (tb) prediction using synthetically generated computed tomography (ct) images. In *Proceedings of the IEEE/CVF international conference on computer vision*, 3265–3273.
- [12] P Li, S Wang, T Li, J Lu, Y HuangFu, and D Wang. 2020. A large-scale ct and pet/ct dataset for lung cancer diagnosis (lung-pet-ct-dx). *Cancer Imaging Arch*.
- [13] Vitali Liauchuk. 2019. Imageclef 2019: projection-based ct image analysis for tb severity scoring and ct report generation. In.
- [14] Stuart Lloyd. 1982. Least squares quantization in pcm. *IEEE transactions on information theory*, 28, 2, 129–137.
- [15] Geoffrey J McLachlan and Thriyambakam Krishnan. 2007. *The EM algorithm and extensions*. John Wiley & Sons.
- [16] Joseph Needham. 1959. *Science and Civilisation in China: Volume 3, Mathematics and the Sciences of the Heavens and the Earth*. Vol. 3. Cambridge University Press.
- [17] Jan Orava. 2011. K-nearest neighbour kernel density estimation, the choice of optimal k. *Tatra Mountains Mathematical Publications*, 50, 1, 39–50.
- [18] Qian Shao, Rongrong Xuan, Yutao Wang, Jian Xu, Menglin Ouyang, Caoqian Yin, and Wei Jin. 2021. Deep learning and radiomics analysis for prediction of placenta invasion based on t2wi. *Mathematical Biosciences and Engineering*, 18, 5, 6198–6215.
- [19] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.
- [20] Xudong Wang, Long Lian, and Stella X Yu. 2022. Unsupervised selective labeling for more effective semi-supervised learning. In *European Conference on Computer Vision*. Springer, 427–445.
- [21] Zifeng Wang, Zhenbang Wu, Dinesh Agarwal, and Jimeng Sun. 2022. Medclip: contrastive learning from unpaired medical images and text. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 3876–3887.
- [22] Xiaoyang Wu, Yuyang Huo, Haojie Ren, and Changliang Zou. 2023. Optimal subsampling via predictive inference. *Journal of the American Statistical Association*, just-accepted, 1–29.
- [23] Yichen Xie, Han Lu, Junchi Yan, Xiaokang Yang, Masayoshi Tomizuka, and Wei Zhan. 2023. Active finetuning: exploiting annotation budget in the pretraining-finetuning paradigm. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 23715–23724.
- [24] Kang Zhang et al. 2020. Clinically applicable ai system for accurate diagnosis, quantitative measurements, and prognosis of covid-19 pneumonia using computed tomography. *Cell*, 181, 6, 1423–1433.
- [25] Mei Zhang, Yongdao Zhou, Zheng Zhou, and Aijun Zhang. 2023. Model-free subsampling method based on uniform designs. *IEEE Transactions on Knowledge and Data Engineering*.
- [26] Hasib Zunair, Aimon Rahman, and Nabeel Mohammed. 2019. Estimating severity from ct scans of tuberculosis patients using 3d convolutional nets and slice selection. In.
- [27] Hasib Zunair, Aimon Rahman, Nabeel Mohammed, and Joseph Paul Cohen. 2020. Uniformizing techniques to process ct scans with 3d cnns for tuberculosis prediction. In *Predictive Intelligence in Medicine: Third International Workshop, PRIME 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 3*. Springer, 156–168.

## A DETAILS OF CLASSIFICATION MODEL

ResNet-34 is used to extract image features. Specifically, we replace the 1000 –  $d$  Fully-Connected (FC) layer in the penultimate layer with a 512 –  $d$  FC layer and delete the last SoftMax layer. The output of ResNet-34 is used as input to the Transformer Encoder, where both the layer number and the head number of the Transformer Encoder are set to 6. Finally, we perform an FC layer and Softmax on the output of the transformer-based network to predict the probability of disease. The prediction loss function is a cross-entropy loss function.

## B HYPERPARAMETER SETTINGS

For CSS, we set the hyperparameters  $k = 10$ ,  $\alpha = 0.5$ ,  $m_{\text{reg}} = 0.9$ ,  $\lambda = 0.5$  and  $t = 10$ . During the training stage of the diagnostic model, we use adaptive moment estimation (Adam) with an initial learning rate  $1e^{-3}$  to optimize the network. We set different batch sizes for experiments with different numbers of compressed selected

slices. The more slices selected for each case, the more GPU memory each case occupies; thus, the batch size should be smaller. In detail,  $batchsize = 1, 4, 8, 16, 16$  is set for  $m = 64, 32, 16, 8, 4$  respectively. For a fair comparison, the training epoch of all experiments is set to 400.

## C IMPLEMENTATION OF OTHER COMPRESSION METHODS

Interpolation-based methods calculate the new pixel value in the compressed image based on the values of surrounding pixels on the Z-axis. As for the slice selection-based method, SSS selects equal slices from the first, middle and last position of the CT volume, whereas ESS selects one after every specific number of slices. As a baseline method, the full CT volume method utilizes the whole CT dataset for training the classification model without compression.