IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 6 December 2023, accepted 19 December 2023, date of publication 25 December 2023,
date of current version 4 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3346910

RESEARCH ARTICLE

Representative Slice Selection and Multi-View Projection Learning for Pulmonary Tuberculosis Infectiousness Identification Using CT Volume

QIUSHUN BAI[1,2], YI GAO[3,4], FENG CHEN[5], YIWEN ZHANG[1,2], YUAN YANG[1,2], LIMING ZHONG[1,2], WEI YANG[1,2], AND YI YANG[1,2]

[1]School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China
[2]Guangdong Provincial Key Laboratory of Medical Image Processing, Guangzhou 510515, China
[3]Department of Infectious Disease, Hainan General Hospital, Hainan Affiliated Hospital of Hainan Medical University, Haikou 570311, China
[4]Department of Infectious Disease and Hepatology Unit, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China
[5]Department of Radiology, Hainan General Hospital, Hainan Affiliated Hospital of Hainan Medical University, Haikou 570311, China

Corresponding authors: Wei Yang (weiyanggm@gmail.com) and Yi Yang (yiyang20110130@163.com)

This work was supported in part by the Hainan Province Science and Technology Special Fund under Grant ZDYF2021SHFZ079, in part by the National Natural Science Foundation of China under Grant 82172020, and in part by the Guangdong Provincial Key Laboratory of Medical Image Processing under Grant 2020B1212060039.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Hainan General Hospital under Application No. 2021-314, and performed in line with the Declaration of Helsinki.

ABSTRACT Pulmonary tuberculosis (PTB) is a major global health threat. Diagnosing PTB infectiousness is vital for clinical decision-making, but existing etiological examination methods do not meet the requirements for speed, accuracy, and cost effectiveness. Developing deep learning models for infectiousness identification based on computed tomography (CT) volume measurements holds promise for meeting these requirements. However, with limited samples and coarse annotations, the large amount of information in the CT volume poses a challenge for models to distinguish information related to patient-level labels, which often leads to severe model overfitting. In this study, A dual-branch framework is developed for identifying PTB infectiousness using CT volume. To address the issue of imbalance between the CT volume information and the patient-level labels, we propose a method for selecting representative slices to reduce redundant information and adopt a multiple-instance learning framework to improve label supervision. Furthermore, we incorporate multi-view projection information to compensate for the deficiency of global information caused by using single-dimensional slices as the input. Experimental results demonstrate that our strategy effectively mitigates overfitting and achieves desirable performance on an external test set, with an area under the receiver operating characteristic curve of 80.48%. This performance is superior to that obtained for models using the 3D CT volume or 2D projection images alone as the input.

INDEX TERMS Pulmonary tuberculosis infectiousness, computed tomography, slice selection, multi-view projection, multiple instance learning.

I. INTRODUCTION

Pulmonary tuberculosis (PTB) is a leading cause of death and a major public health issue worldwide [1]. It is highly infectious and primarily transmitted through the exhalation of *Mycobacterium tuberculosis* by affected individuals [2]. PTB can be divided into infectious PTB and non-infectious PTB. The difference lies in the fact that individuals with infectious PTB exhale *M. tuberculosis*, leading to a high risk of transmission, while those with non-infectious PTB do not. Hence, the fast and accurate identification of infectious cases of PTB is crucial for effective disease

The associate editor coordinating the review of this manuscript and approving it for publication was Carmelo Militello.

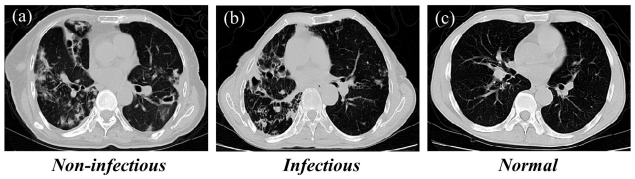


FIGURE 1. Slices fetched from (a) non-infectious PTB, (b) infectious PTB, and (c) normal patient's CT volume.

management and prevention of transmission. In clinical practice, the infectiousness of PTB is typically determined through etiological examination methods such as sputum smear tests. However, the available methods for etiological examination suffer from certain limitations. Sputum smear tests examine respiratory secretions under a microscope but have low sensitivity, potentially yielding false positives for non-tuberculous mycobacteria [3]. While sputum culture tests are precise, they take weeks for results [4]. Xpert Ultra assays, a quick and accurate molecular diagnostic test, are relatively expensive [5]. Consequently, the aforementioned etiological examination methods do not fulfill the requirements for rapid, accurate, and cost-effective diagnosis of PTB infectiousness.

Computed tomography (CT) plays a key role in the workup of patients with PTB by rapidly providing detailed information about the lesions [6]. It is frequently employed for diagnostic tasks, such as determining the type, severity, and activity of PTB. Accurate assessment of PTB activity can be achieved by identifying active lesions within CT volume [7]. However, diagnosing the infectiousness of PTB solely based on CT volume presents more challenges than evaluating the PTB activity. Specifically, it requires inferring pathogenic infectiousness information depending on CT volume. As shown in Fig. 1, the slices in the CT volumes of infectious and non-infectious PTB exhibit similar imaging appearances. Thus, accurately determining the infectiousness of PTB using the CT volume is a complex task that typically requires considerable expertise.

With the advent of deep learning, numerous studies have been conducted on diagnostic aids based on CT volume to improve the efficiency of radiologists, with a particular focus on the use of convolutional neural networks (CNNs). However, in the context of difficult medical sample collection and the limited availability of detailed annotations (such as specific lesion location, mask, type, etc.), training a CNNs classification model with only a small amount of CT volumes and patient-level labels can be challenging. The significant imbalance between the abundance of information in the CT volume and the patient-level labels, which we refer to as the information–label imbalance, inevitably results in learning difficulties and overfitting problems. These issues are further exacerbated by the limited amount of available data. Alleviating the information–label imbalance is a much more feasible solution than collecting more costly data. For example, [8] applied a modified 3D VoxNet to assess PTB severity using CT volume, achieving an area under the receiver-operating-characteristic curve (AUC) result of 73.00%. Despite using spline interpolation to preserve information when reducing CT volume slices, the authors faced challenges in improving performance. By contrast, [9] used a projection-based approach to squeeze the CT volume into 2D projection images, resulting in a 77.54% AUC for the same task. Although such projection helps to reduce the information–label imbalance, it inevitably sacrifices significant detail, complicating further performance enhancement.

In addition to using the 3D CT volume directly or employing 2D projection images as the input, another common method involves slicing the CT volume and using a 2D shared-weight network to extract features from each slice. These slice-level features are then aggregated to make a patient-level prediction. One representative approach is multiple-instance learning (MIL) [10], which has been widely adopted in classification tasks for CT volume. MIL alleviates the problem of the information–label imbalance by improving the supervision of label and has afforded remarkable results [11], [12], [13], [14]. What these works have in common is the use of MIL to identify key information that is relevant to the label from the CT volume. However, they usually select all slices/patches for one direction (typically the highest-resolution one) of CT volume as the input, which causes two issues that most studies have rarely discussed. 1) **Redundant information**: In general, the CT volume has high resolution in one or all directions, which enhances the visibility of lesions. However, massive slices (i.e., those with an abundance of information) may hinder the model from identifying the key information related to the label for MIL classification and ultimately undermine model performance. Meanwhile, processing hundreds of slices can result in a significant computational burden. According to our observations, only a small number of slices are relevant to the diagnostic decisions made by radiologists, and the remaining information can be considered redundant. As shown in Fig. 2 (a), there are two main sources of redundant information. The first one is the presence of adjacent slices with high similarity. These slices generally contain repetitive content, and this situation becomes more pronounced with decreasing slice thickness. The second one pertains to the normal slices, which consume a considerable amount of memory during training but actually make little contribution to the final decision made by radiologists. 2) **Global information deficiency**: As depicted in Fig. 2 (b), the MIL-based method leads to a deficiency of global information by discarding other dimensional slices. This results in a significant loss of information from other dimensions, which can be detrimental to the final aggregation process of MIL during CT volume classification tasks. Therefore, reducing the redundancy in the CT volume and compensating for the deficiency of global information may be helpful when using CNNs to evaluate the infectiousness of PTB.

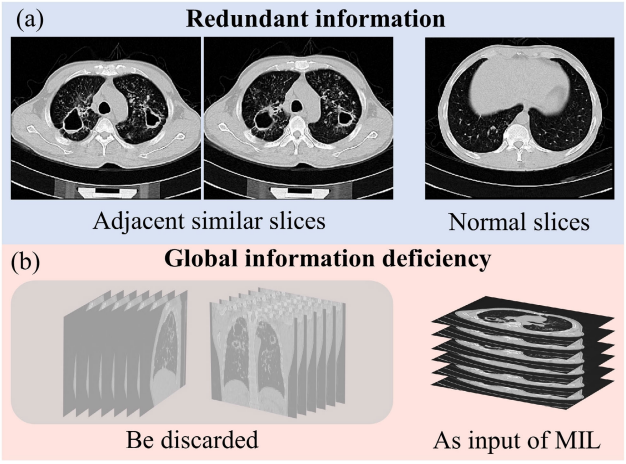


FIGURE 2. Two issues of MIL-based works in CT volume classification tasks. (a) Two types of redundant information: adjacent similar slices and normal slices; (b) Global information deficiency caused by selecting single-dimensional slice as input of MIL, which results in discarding information from other dimensions.

In the present study, a framework based on MIL was developed to achieve fast and accurate identification of PTB infectiousness using the CT volume. To overcome the aforementioned problems of MIL in conventional CT volume classification tasks, a method called representative slice selection (RSS) is proposed. RSS automatically selects a fixed number of representative slices from a CT volume, aiming to address the issue of redundant information. Moreover, multi-view projection (MVP) information is incorporated to minimize the global information deficiency. Specifically, we employ a pre-trained slice classifier in the RSS to reduce the redundancy caused by normal slices. Subsequently, we cluster the selected slices with lesions and then select a fixed number of representative slices to reduce the similar slice redundancy. These representative slices are selected based on their highest intra-lung mean intensity within each cluster. Furthermore, to extract global information from the CT volume without introducing excessive redundancy, we employ MVP to compensate for the global information deficiency caused by using single-dimensional slices as the input. The major contributions of this work can be summarized as follows:

- We propose a novel dual-branch framework to identify the infectiousness of PTB using CT volume.
- We design a CT slice selection method that eliminates unnecessary redundant information while preserving representative slices as much as possible. This approach improves the network identification performance, reduces memory consumption during training, and enhances the practical applicability of the model.
- We integrate MVP information into the MIL paradigm, mitigating the problem of global information deficiency caused by using single-dimensional slices as the input. This further improves the model performance.
- We trained and validated our model using data from one center (591 cases) and tested it with external data from another center (314 cases). Empirical studies show that our model exhibits satisfactory performance for the task of identifying the infectiousness of PTB.

II. RELATED WORK

A. DEEP LEARNING FOR PTB CLASSIFICATION

Owing to the high risk and transmissibility of PTB, numerous scholars have conducted research on medical-image-based diagnostic aids. Reference [15] first developed a deep learning model based on AlexNet to identify PTB from normal chest X-rays. Reference [16] utilized both transfer learning and ensemble learning to diagnose PTB in chest X-rays. Their work primarily demonstrated the potential of deep learning for PTB identification. Compared with chest X-rays, CT can reveal additional disease characteristics and more details regarding disease progression, thus affording more comprehensive information. This allows for more detailed PTB classification studies based on the CT volume. Examples include identification of drug-resistant PTB [17], differentiation from non-tuberculous mycobacteria pneumonia [18], and classification of various lesion types [8], [19], [20], [21], [22].

The topic that most closely resembles the identification of PTB infectiousness is the identification of PTB activity. Reference [23] conducted a comparative study using CT volume to differentiate between active PTB, other types of pneumonia, and normal individuals based on lesion segmentation and connectivity analysis. Reference [7] developed an integrated system for lesion detection, activity diagnosis, and severity assessment. Lesion detection and classification were performed on abnormal slices to determine if they were active. In contrast to the more general approach of [7] and [24] designed an activity recognition model for PTB based on a priori knowledge of the variability of tuberculomas using a fuzzy inference system. These studies concerning the classification of PTB activity demonstrate the potential for a meaningful classification system based on the appearance of lesions in the CT volume. However, the employed methods require the annotation of lesions, which is time consuming and labor intensive. In this study, we present an intelligent approach for identifying the infectiousness of PTB using the CT volume and propose an automatic method for selecting representative slices.

B. MULTIPLE-INSTANCE LEARNING

MIL is a paradigm of weakly supervised learning that falls under the category of ‘‘inexact supervision’’ [25] and is particularly well suited for medical images [26]. The goal of MIL is to identify instances that are relevant to the label of the bag, wherein the bag represents a collection of instances. In CT volume classification tasks with patient-level binary labels, CT volumes are treated as bags containing slices or patches as individual instances. The bag has a binary label, while individual instance label is unknown. Positively labeled bags contain at least one positive instance, whereas negatively

labeled bags consist solely of negative instances. MIL breaks down CT volumes into multiple instances, extracting their embeddings separately and establishing associations with label. This strengthens the impact of a single label during network training and mitigates the information–label imbalance from the label side. In addition, there are more pre-trained 2D models available for 2D instances. Owing to the superiority of MIL for medical image analysis, it has been utilized in recent studies of CT volume classification [12], [13], [14].

Despite the state-of-the-art performance of the aforementioned models, the issues pertaining to the input side, such as redundant information and global information deficiency, are rarely acknowledged. Reference [12] and [13] proposed novel frameworks using all-axial slices of CT volume and attention-based MIL [27], respectively, where the former aimed to evaluate the severity of COVID-19 in patients and the latter focused on the differentiation of COVID-19 from bacterial pneumonia. Reference [14] randomly selected a fixed number of patches to generate a bag for MIL. The authors did not consider the impact of redundant information on the input of MIL and discarded other dimensional information. In this work, we propose RSS to reduce redundant information in the CT volume by selecting a fixed number of representative slices as the input of MIL. This approach can enhance MIL's ability to address the information-label imbalance in the PTB infectiousness identification task. Furthermore, we incorporated MVP to mitigate the global information deficiency in the conventional MIL paradigm for CT volume classification.

C. CT VOLUME REDUCTION

Reducing redundant information is helpful when dealing with CT classification tasks with an information–label imbalance problem. The easiest and most viable reduction strategy is to reduce the number of CT slices. The simplest of these methods use a predetermined procedure for selection, such as random selection [14] or uniform selection [28]. However, these methods have a greater possibility of missing slices with lesions and lack controllability. The importance of the selected slices cannot be guaranteed. Reference [29] manually extracted slices with lesions to train the model, which reduced the amount of information that the model needed to process and partially addressed the aforementioned imbalance. However, manually selecting slices is time consuming and inefficient. Reference [7] designed a pre-trained slice classifier to filter out slices with lesions. Only 10 slices with the highest probability of containing lesions were selected for use. This method is likely to miss slices with lesions in other parts of the lung, especially in cases with multiple lesions. By contrast, we have developed a straightforward and efficient method for automatically selecting representative slices. This approach ensures that the selected slices are both important and comprehensive.

In addition to reducing the number of slices, squeezing of the CT volume can also lead to a certain reduction effect. Reference [9] divided the lung CT data into left and right lung regions and performed mean, standard deviation (SD), and maximum projections from the axial, coronal, and sagittal planes. This process simplified and compressed the CT volume into multi-view 2D projections while retaining some of the overall features. It achieved the first rank in that competition [30]. Although MVP significantly reduces the information–label imbalance and has afforded adequate results, it also leads to the loss of many details due to projection squeeze. By contrast, we do not solely rely on MVP information for inference, but consider it as an extra source to obtain basic 3D statistical features that act as supplementary global information.

III. METHOD

In this section, we elaborate on the proposed framework for the task of PTB infectiousness identification.

A. OVERVIEW

As illustrated in Fig. 3, our designed framework can be divided into two parts: CT volume reduction and infectious prediction. From the perspective of data flow, there is a dual-branch framework that contains the multiple-instance branch and the MVP branch. CT volume reduction includes preprocessing (not depicted in the figure), and two reduction method, namely, RSS and MVP. The masked and tailored CT volume obtained by preprocessing is conveyed into two handling modules for RSS and MVP. In the multiple-instance branch, RSS serves as a filter to select a fixed number of representative slices for the next feature extraction process. In the MVP branch, the CT volume is separated and projected into 18 individual 2D views. Correspondingly, slice (i.e., instance) embeddings and feature maps from three directional projections are extracted using four backbone networks in the infectiousness prediction part. In the IPFusion (Instance and Projection Fusion) module, the embeddings of slices are finally generated into a bag vector through MIL pooling. Next, the feature maps of the three directional projections are fused into a single global projection vector. Finally, a classifier outputs the prediction using the concatenated version of the bag vector and the global projection vector. The details of each component are elaborated as follows.

B. REPRESENTATIVE SLICE SELECTION

The principle of RSS is to reduce redundancy while retaining the importance and diversity of CT volume (i.e., keeping representative slices). For the two types of redundancy information (i.e., similar slices and normal slices), we believe that separate processing is required. To ensure that the process of reducing similar slices is not interfered with by normal slices, the primary step is to remove the normal slices. Therefore, we first trained a slice classifier using PTB lesion slices selected by radiologists from the CT volumes of PTB patients, as well as normal slices from the CT volumes of healthy patients. We then utilized this classifier to identify PTB lesion slices for the subsequent reduction step. As shown

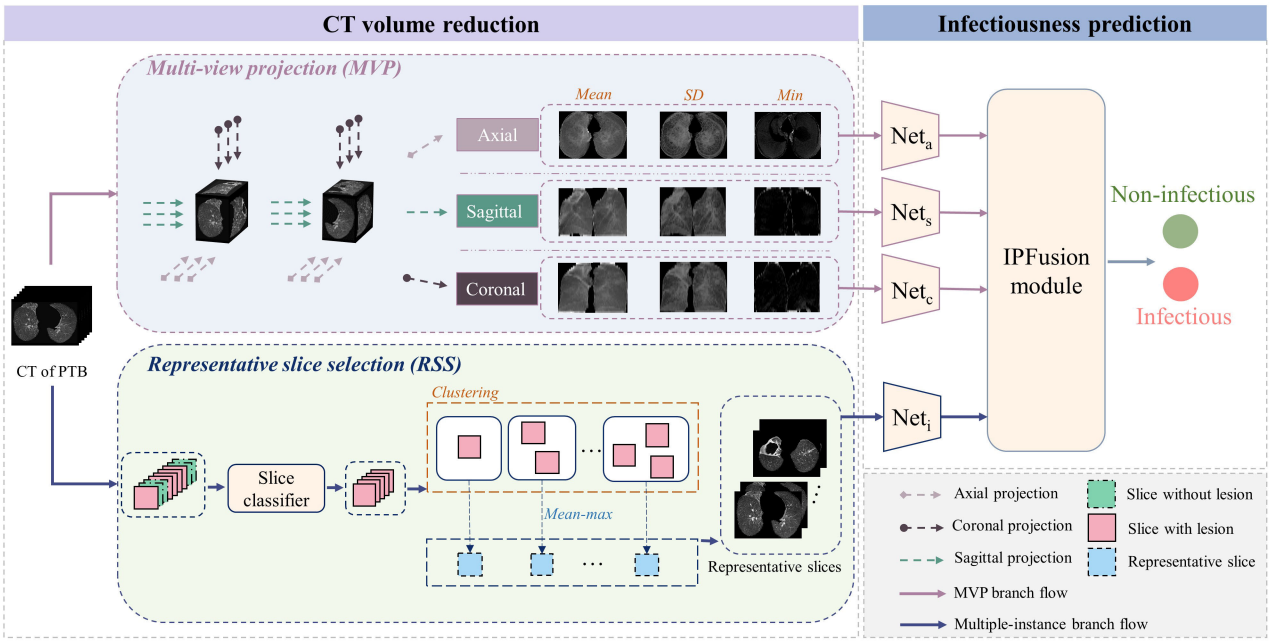


FIGURE 3. An overview of our proposed framework. After obtaining the pre-processed CT volume of the PTB, we reduce the CT volume in two way: RSS for reduction and MVP for squeeze. In the Infectiousness prediction part, we fuse the information from instances and projections and make prediction via the proposed IPFusion (Instance and Projection Fusion) module.

in Fig. 3, given a fixed K as the number of output slices, the lesion slices were selected by the pre-trained slice classifier. It is worth noting that it is not a difficult task to differentiate lesion slices from normal slices on account of the distinct appearance of PTB lesions in our dataset, as shown in Fig. 1; hence, a simple classifier is sufficient for this task. This step significantly reduces the number of unimportant slices. Next, to maintain input diversity, *K-means* clustering is performed to group the lesion slices into K clusters based on their gray intensity distributions within the lung field. We assume that lesions have a higher mean intensity because they typically appear as high-intensity areas in the CT volume [31], [32]. Therefore, we select the slices with the highest intra-lung mean intensity in each cluster as representative slices. In this manner, K representative slices are filtered out according to the principle of considering both importance and diversity, and they serve as instances for the model input. The details of the RSS are elaborated in Algorithm 1.

Algorithm 1 Representative Slice Selection

Input: The axial slices of a CT volume, S; The desired slice number K for output
Output: K slices

S_n: the set of normal slices;
S_l: the set of lesion slices;
S_r: the set of representative slices;
$S_n, S_l = slice\ classifier(S)$;
N = slice number of S_l;
if $N <= K$ **then**
  return $S_l + (K - N)$ slices randomly selected from S_n.
else
  $clusters = Kmeans(S_l)$;
  for *cluster in clusters* **do**
    if *only one slice in cluster* **then**
      put the slice into S_r.
    else
      put the slice with the highest intra-lung mean intensity among the *cluster* into S_r.
    end
  end
  return S_r.
end

C. MULTI-VIEW PROJECTION

To alleviate the issue of insufficient global information, we employ MVP to squeeze the CT volume and preserve some specific global features, which was inspired by the method of [9]. Projection information can be regarded as complementary information to the single-dimensional slice input, without adding excessive redundancy. Specifically, we first separate the CT volume into left and right lung regions after obtaining the masked and cropped CT volume. Then, projection operations are performed on each lung separately. These operations include calculating the mean, SD, and minimum values of all voxel points along the projection path from the axial, coronal, and sagittal planes.

The process of projection allows us to obtain a set of 2D projection images, as illustrated in Fig. 3. Because the resolution varies for each dimension, we resize the images to a consistent size for later combination. By using MVP with multiple directions and values, we are able to preserve some general global features while also enriching the input information. This approach is beneficial for alleviating the lack of insufficient dimensional information without introducing unnecessary redundancy.

D. FEATURES EXTRACTION AND IPFUSION MODULE

For the feature extraction of the MVP branch, we utilize three ResNet-18 [33] models pre-trained on ImageNet [34] (retaining only partial models and weights from the first convolutional layer to the second basic blocks (included)) as feature extractors: Net_a (axial), Net_s (sagittal) and Net_c (coronal), respectively. The input of these backbones is formed as the left and right lung stitched triple images (mean, SD, and minimum values). The output feature map size for each feature extractor is (C, H, W) = (128, 28, 28). Next, the feature maps from the three feature extractors are concatenated by channel and then fed into the subsequent layers, as depicted in Fig. 4. The convolution layer produces M feature maps. After applying group normalization (GN) and the ReLU activation function, an adaptive averaging pooling operation is performed. This operation averages each feature map into a single value, resulting in the global vector $Proj \in \mathbb{R}^M$.

Feature extractor Net_i of multiple-instance branch is an important component that greatly impacts the overall performance in our task, because infectious PTB and non-infectious PTB present similar CT manifestations, which requires that the feature extractor possess excellent ability to extract discriminatory features. One of the advantages of RSS is to reduce the number of slices, allowing us to use a better backbone for obtaining superior model performance under limited video memory. We selected EfficientNet-B4 [35] pre-trained on ImageNet as our feature extractor on account of its outstanding performance. Of note, while EfficientNet demonstrates superior classification ability, it consumes significantly more memory during training than ResNet because of its utilization of depthwise separable convolution. In our design, the original CT slice serves as the primary basis for prediction in the multiple-instance branch, while the projection image is considered as auxiliary information. Therefore, we allocate main memory to the multiple-instance branch and utilize a robust backbone network to acquire more effective information. The MVP branch consumes less memory and also extracts useful classification information.

The Gated-Attention [27] module was employed to calculate the attention weights and adaptively fuse the embeddings to generate a bag vector. The MIL pooling process is expressed in Equation 1. $Z \in \mathbb{R}^M$ is the final bag vector. $h_k \in \mathbb{R}^M$ denotes the feature embedding for each instance, and a_k represents the weight coefficient of each instance,

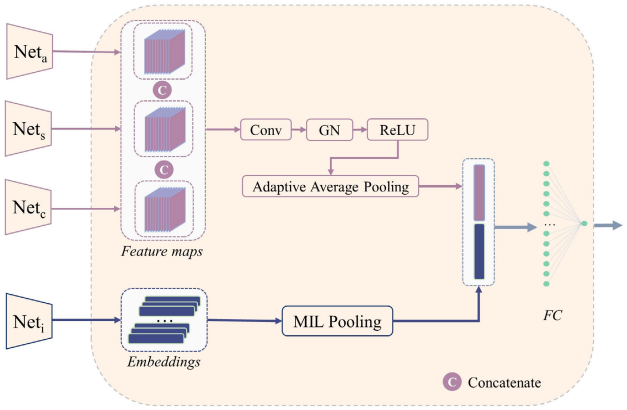


FIGURE 4. IPFusion module. The module accepts the feature maps from three MVP branch feature extractors and instance embeddings from the multiple-instance branch feature extractor.

which reflects its contribution to the bag vector.

$$z = \sum_{k=1}^{K} a_k h_k, \tag{1}$$

here, the a_k is calculated as follows:

$$a_k = \frac{exp\{w^T(tanh(Vh_k^T) \odot sigm(Uh_k^T))\}}{\sum_{k=1}^{K} exp\{w^T(tanh(Vh_k^T) \odot sigm(Uh_k^T))\}}, \tag{2}$$

where, *tanh* is the hyperbolic tangent function, *sigm* is the sigmoid function, and $w \in \mathbb{R}^L$, $V \in \mathbb{R}^{L\times M}$, $U \in \mathbb{R}^{L\times M}$ are trainable parameters. In this study, we set M and L as 512 and 128.

After obtaining the global vector *Proj* and the bag vector Z, we concatenate them and input the results into a fully connected layer. Finally, we output the infectious probability p using a sigmoid function. This process can be formulated as follows:

$$p = sigm(FC(concat(Proj, Z)). \tag{3}$$

IV. EXPERIMENTS

A. DATASETS

This retrospective study was approved by an ethics review board, and the requirement to obtain informed written consent was waived. For the purposes of this study, we collected data from a hospital in Hainan province, China, including 591 cases of PTB. Among these cases, 462 were infectious and 129 were non-infectious. The external test set originated from another hospital in Hainan province, consisting of 314 cases. Among these cases, 146 were infectious and 168 were non-infectious. To ensure the reliability of infectiousness labeling, each patient underwent multiple (three or more) sputum smear tests within one month. Cases with a positive result on any of the sputum smear tests were labeled as infectious; otherwise, they were labeled as non-infectious. All patients were scanned using spiral CT scanners, specifically the Philips Healthcare IQon Spectral CT and SOMATOM Force CT systems, following the same

TABLE 1. Details of the number of CT volumes in our dataset.

	Non-infectious	Infectious	Total
Training	129	462	591
External-Test	168	146	314

protocol. The CT volumes had a pixel spacing of 5 mm in the plane and a resolution of 512 × 512 in another plane, and the number of slices ranged from 47 to 70. The specific details are provided in Table 1.

B. IMPLEMENTATION

1) PREPROCESSING

We first truncated the CT volume window to a regular lung window [−1000, 400] to enhance the visibility of the lesions and ensure consistency in the intensity spectra. Then, to remove the non-lung area, lung field segmentation was performed using Lung-mask [36], a pre-trained model specifically designed for segmenting lung fields. Finally, we tailored the CT volume by the maximum circumscribed cube of the lung field to focus on the lung field.

2) TRAINING

For the training, the training set was randomly divided into five subsets for five-fold cross-validation to select the hyperparameters. Next, we tested each of the five models using cross-validation on the external test set. Finally, we obtained the average values of the test set metrics for each of the five models. For a fair comparison, we consistently retained most of the settings and made some changes to accommodate specific training requirements. In terms of the fixed settings, we performed the training for a maximum of 50 epochs using a PyTorch-based implementation on an NVIDIA GeForce GTX 2080Ti with 12 GB of GPU memory. We used weight resampling to alleviate the data imbalance and utilized the Adam optimizer [37] with a weight decay of 10^4. Additionally, we employed a polynomial learning rate scheduler with a maximum learning rate of 10^4 and a decay rate of 0.9. Data augmentations and the early-stop strategy were employed to mitigate overfitting. **Different methods comparison**: **3D**: all CT volumes were resampled to the average training set size of 208×336×46. Batch size 8 was used for the training. **2D projection only** [9]: projection images are resized to 224×224. Batch size of 16 for training. For our method, we resized the slices to 224 × 224 and duplicated them three times in the channel to fulfill the input requirements of the pre-trained model. The batch size was set to 1 for the training. Unless otherwise specified, there were 10 slices in each bag. We used the cross-entropy function as the loss function, which is defined as follows:

We used the cross-entropy function as the loss function, which is defined as follows:

$$L = \frac{1}{N}\sum_{i=1} -[y_i \cdot log(p_i) + (1 - y_i) \cdot log(1 - p_i)], \quad (4)$$

where N represents the number of samples, y_i denotes the label of the i^{th} sample, which $\in$ [0, 1], and p_i represents the positive probability of the model output for the i^{th} sample. The following metrics were utilized to evaluate model performance: AUC, accuracy, precision, recall, and F1 score (F1). All of the metrics are reported in the form of mean (± SD).

V. RESULTS

A. SLICE CLASSIFIER

We trained the slice classifier based on ResNet-18 using 20,587 slices containing lesions selected by physicians (from 591 PTB cases) and 31,819 normal slices (from 662 normal cases). We divided the training and validation sets by patient in a 4:1 ratio. The model achieved an AUC of 99.03% for the validation set, demonstrating its ability to accurately distinguish between lesion slices and normal slices from the CT volume. For RSS, we used the version trained on all of the data (52,406 slices).

B. ABLATION STUDY

To validate the efficacy of our proposed strategy, we initially employed 40 slices obtained through uniform selection. These slices were then used as the input for the multiple-instance branch in our framework, thus serving as a baseline. From the results presented in Table 2, we can see that we achieved better performance when we extracted 15 slices using RSS than that obtained using the 40 uniformly selected slices. This demonstrates the effectiveness of our strategy for filtering slices. We then incorporated the information from MVP to enhance the global information, resulting in further improvement in the classification performance. It is notable that after incorporating the MVP information, the performance of the model with 10 input slices in the multiple-instance branch was superior to that with 15 input slices, which exhibited a decline in performance. We speculate that there are two potential reasons for this. First, both the slices and the projection images are derived from the CT volume. Therefore, the information obtained through projection may already be present in the slices and introducing projection information in this case may not lead to further improvement. Second, the incorporation of projection information may introduce redundancy, which could exacerbate the information–label imbalance, ultimately diminishing the model effectiveness. Consequently, the influence of introducing projection information was found to be inconsistent.

C. INFLUENCE OF SLICE NUMBER

To further test our hypothesis regarding the detrimental effects of redundant information on the model, we conducted experiments using various numbers of slices as the input for the multiple-instance branch, without adding MVP information. Fig. 5 shows that the model performance initially increased and then decreased with increasing number

TABLE 2. Ablation study on our proposed method for the identification of PTB infectiousness. K is the slice number.

	AUC (%)	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
Baseline*	76.98 (±2.29)	73.69 (±2.71)	68.23 (±4.05)	**82.19 (±4.67)**	74.41 (±2.07)
Baseline+RSS(K=15)	79.37 (±1.30)	**76.05 (±1.52)**	72.03 (±2.07)	79.45 (±3.28)	**75.51 (±1.60)**
Baseline+RSS(K=15)+MVP	77.57 (±2.08)	74.71 (±2.56)	71.10 (±4.33)	77.95 (±5.00)	74.16 (±1.58)
Baseline+RSS(K=10)+MVP	**80.48 (±2.42)**	75.80 (±2.49)	**72.43 (±5.00)**	78.36 (±3.20)	75.11 (±1.54)

* using 40 slices obtained through uniform selection as input.

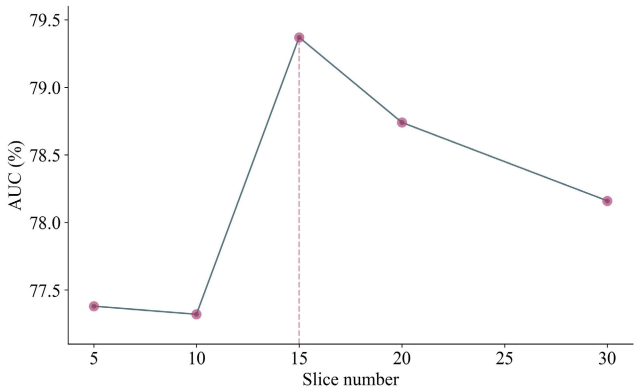


FIGURE 5. Identification performance of different selected slice numbers in RSS.

of slices. The initial increase was ascribed to the fact that when the number of slices is too small (e.g., five slices), there is insufficient information for the model to fully learn about the entire CT volume. As the number of slices increases (e.g., 20-30 slices), however, the excess information begins to interfere with the ability of the model to focus on the key information, leading to a decline in performance.

The above results illustrate two important points. First, more slices are not always better. The increase in slices may introduce redundancy when the fundamental information has already been secured. Although providing more information can expand the search scope of the network, it can also increase the learning difficulty and negatively affect the model performance. In the case of our task, 15–20 slices afforded better results. Increasing the number of slices also leads to performance degradation and higher memory consumption. Second, RSS can effectively retain the crucial information from the CT volume via extracting representative slices to accurately identify the infectiousness of PTB while reducing redundancy.

D. COMPARISON WITH OTHER SLICE SELECTION METHODS

We compared our slice selection method with several alternative approaches:

a) Random: Selecting K slices randomly from the CT Volume.
b) Uniform: Selecting K slices from the CT Volume with nearly equal intervals between each slice.
c) Mean-max: This method involves calculating the mean gray intensity of the intra-lung region for each slice and selecting the K slices with the highest intensities.
d) Classifier-mean-max: After selecting lesion slices (using a threshold of 0.5) using the slice classifier, we performed mean-max selection on the lesion slices.
e) Classifier-prob-max: Selecting K slices with the highest positive probability predicted by the slice classifier among all of the slices. This approach has been utilized in previous research [7]
f) RSS

As illustrated in Fig. 6, RSS excludes the normal slices such as the first three slices obtained by the random (a), uniform (b), and mean-max (c) methods, while maintaining the diversity of selected slices compared with the classifier-mean-max (d) and classifier-prob-max (e) methods. In this manner, RSS ensures a more representative selection of slices in the event of a significant reduction in CT capacity slices.

In Table 3, the following results are observed. First, the mean AUC values of the first three methods without a slice classifier (72.71%) were significantly lower than those of the last three methods (76.55%), demonstrating the effectiveness of the strategy of initially selecting slices containing lesions. In cases where only a small number of sections are taken, it is helpful for classification to retain as many lesion slices as possible. Second, the mean-max and classifier-prob-max selection methods afforded relatively poor results that were only slightly superior to those obtained by random sampling. In Fig. 6(c) and (e), we observe that these two methods do not guarantee the diversity of slices at all, i.e., they missed some parts of representative slices. This demonstrates that slice diversity is an important aspect of slice selection. Third, the uniform selection method ensures slice diversity and preserves the original distribution information to a greater extent. This is the reason why it outperformed other methods, with the exception of RSS. However, this approach does not guarantee the significance of the selected slices, i.e., whether they contain lesions. Finally, our method exhibited the best results. Compared with the second-best classifier-mean-max method, we used K-means clustering to ensure the diversity of the final filtered slices. This approach allows for a better representation of the key information in the CT volume.

E. EFFECT OF DIFFERENT PROJECTION COMBINATIONS

As shown in Fig. 7, projection images obtained using different values can reflect distinct characteristics about lesions within the CT volume. In addition to the underlying mean and SD projection images, we found that the minimum-value projection provided some interesting information. Specifically, it revealed some penetrating cavity

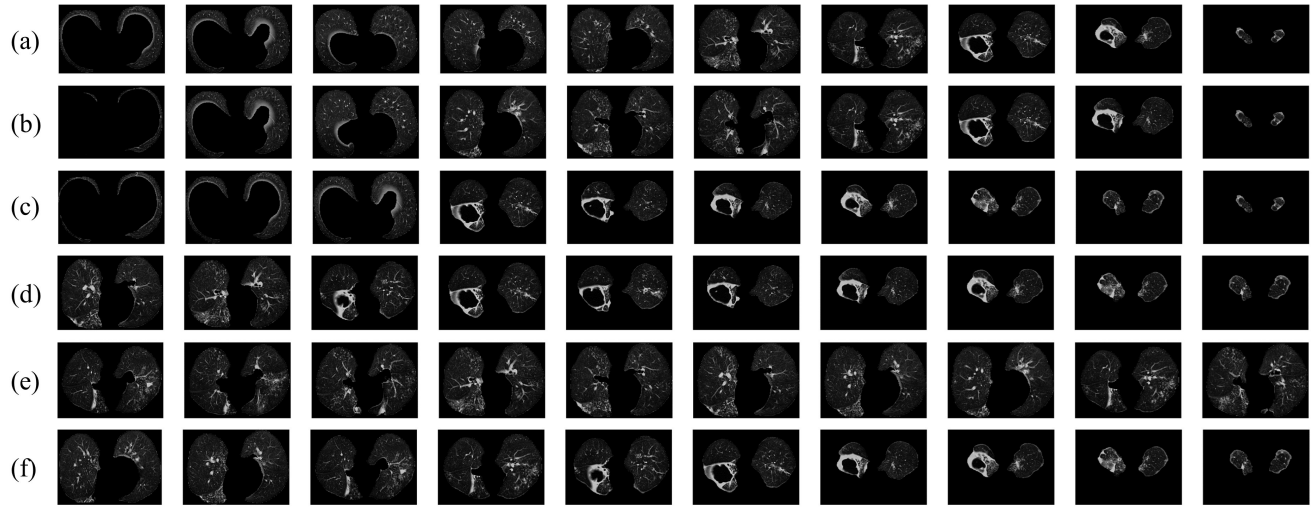


FIGURE 6. 10 slices were selected using different slice selection methods from the same CT volume of PTB. (a)-(f) are Random, Uniform, Mean-Max, Classifier-mean-max, Classifier–prob-max selection and RSS, respectively. RSS presents a more comprehensive selection capability compared with other methods.

TABLE 3. Classification performance of using different slice selection methods.

Methods	AUC (%)	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
Random	70.33 (±1.86)	66.69 (±1.05)	62.08 (±1.56)	73.29 (±4.96)	67.12 (±1.65)
Mean-max	72.13 (±2.38)	69.36 (±2.18)	64.63 (±2.63)	75.75 (±3.80)	69.68 (±2.01)
Uniform	75.69 (±2.22)	70.76 (±2.03)	64.69 (±2.44)	**82.19 (±2.52)**	72.35 (±1.35)
Classifier-prob-max	72.20 (±1.99)	68.85 (±1.13)	63.54 (±3.53)	79.86 (±12.12)	70.19 (±3.21)
Classifier-mean-max	76.97 (±1.38)	72.68 (±0.85)	70.03 (±3.50)	73.15 (±7.46)	71.23 (±2.03)
RSS(Ours)	**80.48 (±2.42)**	**75.80 (±2.49)**	**72.43 (±5.00)**	78.36 (±3.20)	**75.11 (±1.54)**

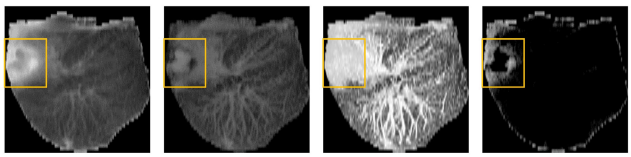


FIGURE 7. Different projection images. Maximum value projection masks more information compared with other value projections.

lesions, as indicated by the yellow box, which were not captured by the maximum-value projection. As can be seen in Fig. 8, the projection combination of mean, SD, and minimum values afforded better results than the combination of maximum values. The latter even displayed a lower AUC than the best result obtained without adding projection information, further suggesting that the addition of projection information does not always lead to enhanced performance. In the case of projections that introduce too much noise, such as the maximum value, the model performance may instead deteriorate.

F. COMPARISON WITH OTHER METHODS FOR IDENTIFYING PTB INFECTIOUSNESS

Table 4 shows the results obtained using different forms of input for PTB infectiousness identification. Neither the use of the 3D CT volume directly nor the use of the 2D projection images only provided good results in our task. Using the 3D CT volume as the input faces obstacles in attaining good generalization performance due to several factors, including the multitude of parameters in 3D networks, the absence of detailed label, and the limited amount of available data. Using the 2D projection images as the input afforded better results than using the 3D CT volume via squeezing the CT volume while preserving the global information. However, this approach also leads to a significant loss of detail, making it impossible to improve the classification performance. By taking advantage of the combination of MIL and MVP, our method outperformed the other two methods with respect to all five metrics. Fig. 9 presents the receiver operating characteristic (ROC) curves obtained for the different methods.

VI. DISCUSSION

In this paper, we have proposed a dual-branch framework for identifying the infectiousness of PTB using the CT volume. Compared with previous CT volume classification studies based on the MIL paradigm [12], [13], [14], our proposed framework pays greater attention to the optimization of the input side to overcome the challenges posed by redundant information present in the CT volume and the global information deficiency due to using single-dimensional slices

TABLE 4. Comparison of using 3D volume and merely 2D projection (images) as input for the identification of PTB infectiousness.

Methods	AUC (%)	Accuracy(%)	Precision(%)	Recall(%)	F1(%)
3D ResNet-34	63.14 (±4.65)	61.34 (±2.71)	58.71 (±3.14)	59.45 (±20.42)	57.50 (±9.86)
3D EfficientNet-B4	62.86 (±2.45)	60.89 (±2.97)	56.10 (±3.26)	77.53 (±8.56)	64.77 (±1.37)
2D projection	72.80 (±2.06)	68.73 (±1.07)	65.39 (±1.90)	70.27 (±8.85)	67.44 (±3.27)
Ours	**80.48 (±2.42)**	**75.80 (±2.49)**	**72.43 (±5.00)**	**78.36 (±3.20)**	**75.11 (±1.54)**

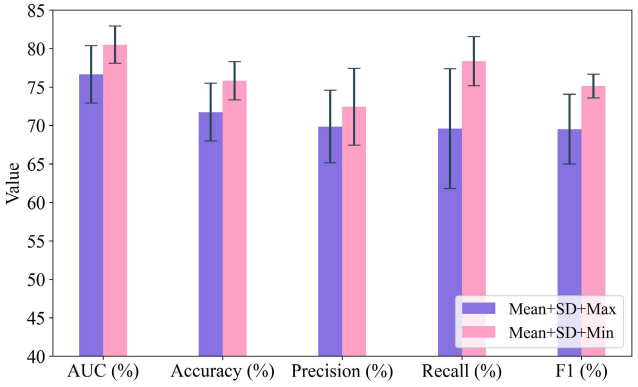


FIGURE 8. Comparision of different projection combinations for MVP branch.

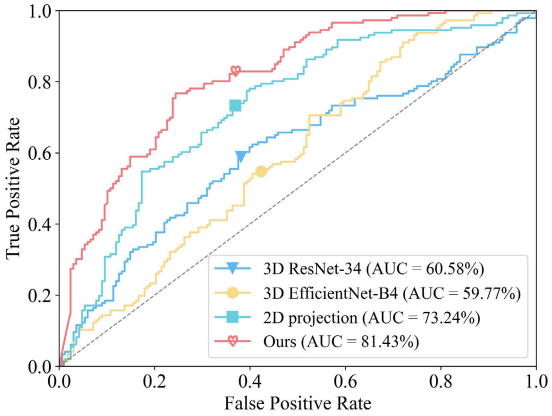


FIGURE 9. ROC curves of different methods on external test set.

as the input. For the former aspect, we employed RSS to reduce the redundant information in the CT volume by keeping a fixed number of representative slices. For the latter aspect, we incorporated MVP to squeeze the CT volume while preserving some projection information from multiple views to mitigate the global information deficiency problem.

Table 2 demonstrates the effectiveness of our CT volume reduction strategy. We further confirmed the performance of RSS with respect to two considerations, namely, the influence of different numbers of slices and comparison with other slice selection methods. From Fig. 5, we see that the gain from the increased amount of information is not always proportional. As the number of slices continues to increase, redundant information may start to diminish the model performance. As depicted in Fig. 6, RSS exhibits improved selection capability compared with other selection methods by retaining the most representative slices. This superior selection capability ultimately leads to enhanced performance for the PTB infectiousness identification task, as shown in Table 3. In addition, RSS is designed based on clinical prior knowledge. Therefore, the slice selection results have a certain degree of clinical explainability as previous works [38], [39]. Furthermore, the effect of MVP was explored. We find that the performance improvement achieved by incorporating MVP is not very stable. This instability may originate from the redundancy introduced by MVP, which potentially leads to impaired model performance. We also identified a better projection combination for our task. Finally, we compared our proposed approach with other methods. Table 4 suggests that the use of the 3D CT volume or 2D projection images as the input does not yield good results, in accordance with the findings of [29]. In the context of difficult medical sample collection and the limited availability of detailed annotations, our strategy is considerably more feasible for the identification of PTB infectiousness than other methods using the 3D CT volume or 2D projection images alone as the input.

Nonetheless, there is some room for improvement that may be explored in the future. First, there exists the possibility of losing the most important slices during the slice selection step. Although the slices within the cluster can partially compensate for this, it may still impact the final prediction of the network. Slice selection is a trade-off between capturing more details and reducing redundancy. The obtained results suggest that retaining important information and reducing redundancy are both important. Therefore, we intend to explore the possibility of keeping more key information during the process of reducing redundancy. Furthermore, the introduction of clinical information such as examination metrics and main patient symptoms may further improve the identification performance of the model. How to improve the fusion of the available information (both imaging data and clinical information) is also something that warrants further exploration in the future.

VII. CONCLUSION

We have presented a novel deep learning framework for determining PTB infectiousness by exploiting the attention-based MIL paradigm. To alleviate the imbalance between the abundant information in the CT volume and the patient-level labels, we used RSS to reduce the redundant information in the CT volume. In addition, we incorporated MVP to overcome the problem of deficient global information caused by using single-dimensional slices as the input. Comprehensive experiments were conducted to demonstrate

the effectiveness of the proposed method in terms of identification performance. Visualization and ablation studies were also performed to facilitate a more thorough analysis and provide a greater understanding of our method. Despite the effectiveness of our framework, there appears to be room for improvement in assessing PTB infectiousness. Possible approaches include enhancing the accuracy and efficiency of slice selection and incorporating available clinical information.

ACKNOWLEDGMENT

(Qiushun Bai and Yi Gao contributed equally to this work.)

REFERENCES

[1] *Global Tuberculosis Report 2021*, World Health Organization, Geneva, Switzerland, 2022.

[2] E. Temesgen, Y. Belete, K. Haile, and S. Ali, "Prevalence of active tuberculosis and associated factors among people with chronic psychotic disorders at St. Amanuel mental specialized hospital and Gergesenon mental rehabilitation center, Addis Ababa, Ethiopia," *BMC Infectious Diseases*, vol. 21, no. 1, p. 1100, Dec. 2021.

[3] K. F. Laserson, N. T. N. Yen, C. G. Thornton, V. T. C. Mai, W. Jones, D. Q. An, N. H. Phuoc, N. A. Trinh, D. T. C. Nhung, T. X. Lien, N. T. N. Lan, C. Wells, N. Binkin, M. Cetron, and S. A. Maloney, "Improved sensitivity of sputum smear microscopy after processing specimens with C 18-carboxypropylbetaine to detect acid-fast bacilli: A study of United States-bound immigrants from Vietnam," *J. Clin. Microbiol.*, vol. 43, no. 7, pp. 3460–3462, Jul. 2005.

[4] S. Ghosh, D. Felix, J. S. Kammerer, S. Talarico, R. Brostrom, A. Starks, and B. Silk, "Evaluation of sputum-culture results for tuberculosis patients in the United States-affiliated Pacific islands," *Asia Pacific J. Public Health*, vol. 34, nos. 2–3, pp. 258–261, Mar. 2022.

[5] N. Mafirakureva, E. Klinkenberg, I. Spruijt, J. Levy, D. Shaweno, P. de Haas, N. Kaswandani, A. Bedru, R. Triasih, M. Gebremichael, P. J. Dodd, and E. W. Tiemersma, "Xpert ultra stool testing to diagnose tuberculosis in children in Ethiopia and Indonesia: A model-based cost-effectiveness analysis," *BMJ Open*, vol. 12, no. 7, Jul. 2022, Art. no. e058388.

[6] X. W. Gao, C. James-Reynolds, and E. Currie, "Analysis of tuberculosis severity levels from CT pulmonary images based on enhanced residual deep learning architecture," *Neurocomputing*, vol. 392, pp. 233–244, Jun. 2020.

[7] C. Yan, L. Wang, J. Lin, J. Xu, T. Zhang, J. Qi, X. Li, W. Ni, G. Wu, J. Huang, Y. Xu, H. C. Woodruff, and P. Lambin, "A fully automatic artificial intelligence–based CT image analysis system for accurate detection, diagnosis, and quantitative severity evaluation of pulmonary tuberculosis," *Eur. Radiol.*, vol. 32, no. 4, pp. 2188–2199, Apr. 2022.

[8] H. Zunair, A. Rahman, N. Mohammed, and J. P. Cohen, "Uniformizing techniques to process CT scans with 3D CNNs for tuberculosis prediction," in *Predictive Intelligence in Medicine*, vol. 12329, I. Rekik, E. Adeli, S. H. Park, and M. D. C. Valdés Hernández, Eds. Cham, Switzerland: Springer, 2020, pp. 156–168.

[9] V. Liauchuk, "ImageCLEF 2019: Projection-based CT image analysis for TB severity scoring and CT report generation," in *Proc. Conf. Labs Eval. Forum (CLEF)*, 2019, p. 13.

[10] M.-A. Carbonneau, V. Cheplygina, E. Granger, and G. Gagnon, "Multiple instance learning: A survey of problem characteristics and applications," *Pattern Recognit.*, vol. 77, pp. 329–353, May 2018.

[11] Z. Han, B. Wei, Y. Hong, T. Li, J. Cong, X. Zhu, H. Wei, and W. Zhang, "Accurate screening of COVID-19 using attention-based deep 3D multiple instance learning," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2584–2594, Aug. 2020.

[12] Z. Li, W. Zhao, F. Shi, L. Qi, X. Xie, Y. Wei, Z. Ding, Y. Gao, S. Wu, J. Liu, Y. Shi, and D. Shen, "A novel multiple instance learning framework for COVID-19 severity assessment via data augmentation and self-supervised learning," *Med. Image Anal.*, vol. 69, Apr. 2021, Art. no. 101978.

[13] P. Chikontwe, M. Luna, M. Kang, K. S. Hong, J. H. Ahn, and S. H. Park, "Dual attention multiple instance learning with unsupervised complementary loss for COVID-19 screening," *Med. Image Anal.*, vol. 72, Aug. 2021, Art. no. 102105.

[14] K. He, W. Zhao, X. Xie, W. Ji, M. Liu, Z. Tang, Y. Shi, F. Shi, Y. Gao, J. Liu, J. Zhang, and D. Shen, "Synergistic learning of lung lobe segmentation and hierarchical multi-instance classification for automated severity assessment of COVID-19 in CT images," *Pattern Recognit.*, vol. 113, May 2021, Art. no. 107828.

[15] S. Hwang, H.-E. Kim, J. Jeong, and H.-J. Kim, "A novel approach for tuberculosis screening based on deep convolutional neural networks," *Proc. SPIE*, vol. 9785, Mar. 2016, Art. no. 97852W.

[16] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, Aug. 2017.

[17] X. W. Gao and Y. Qian, "Prediction of multidrug-resistant TB from CT pulmonary images based on deep learning techniques," *Mol. Pharmaceutics*, vol. 15, no. 10, pp. 4326–4335, Oct. 2018.

[18] Q. Yan, W. Wang, W. Zhao, L. Zuo, D. Wang, X. Chai, and J. Cui, "Differentiating nontuberculous mycobacterium pulmonary disease from pulmonary tuberculosis through the analysis of the cavity features in CT images using radiomics," *BMC Pulmonary Med.*, vol. 22, no. 1, p. 4, Dec. 2022.

[19] L. Li, H. Huang, and X. Jin, "AE-CNN classification of pulmonary tuberculosis based on CT images," in *Proc. 9th Int. Conf. Inf. Technol. Med. Educ. (ITME)*, Oct. 2018, pp. 39–42.

[20] X. Li, Y. Zhou, P. Du, G. Lang, M. Xu, and W. Wu, "A deep learning system that generates quantitative CT reports for diagnosing pulmonary tuberculosis," *Int. J. Speech Technol.*, vol. 51, no. 6, pp. 4082–4093, Jun. 2021.

[21] S. C. Wu, X. J. Wang, J. Y. Ji, G. Geng, Z. H. Zhang, and D. L. Hou, "A preliminary investigation on a deep learning convolutional neural networks based pulmonary tuberculosis CT diagnostic model," *Chin. J. Tubere. Respir. Dis.*, vol. 44, no. 5, pp. 450–455, May 2021.

[22] A. Lewis, E. Mahmoodi, Y. Zhou, M. Coffee, and E. Sizikova, "Improving tuberculosis (TB) prediction using synthetically generated computed tomography (CT) images," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 3258–3266.

[23] L. Ma, Y. Wang, L. Guo, Y. Zhang, P. Wang, X. Pei, L. Qian, S. Jaeger, X. Ke, X. Yin, and F. Y. M. Lure, "Developing and verifying automatic detection of active pulmonary tuberculosis from multi-slice spiral CT images based on deep learning," *J. X-Ray Sci. Technol.*, vol. 28, no. 5, pp. 939–951, Sep. 2020.

[24] V. Sineglazov, K. Riazanovskiy, A. Klanovets, E. Chumachenko, and N. Linnik, "Intelligent tuberculosis activity assessment system based on an ensemble of neural networks," *Comput. Biol. Med.*, vol. 147, Aug. 2022, Art. no. 105800.

[25] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 44–53, Jan. 2018.

[26] G. Quellec, G. Cazuguel, B. Cochener, and M. Lamard, "Multiple-instance learning for medical image and video analysis," *IEEE Rev. Biomed. Eng.*, vol. 10, pp. 213–234, 2017.

[27] M. Ilse, J. M. Tomczak, and M. Welling, "Attention-based deep multiple instance learning," 2018, *arXiv:1802.04712*.

[28] X. Wang, L. Jiang, L. Li, M. Xu, X. Deng, L. Dai, X. Xu, T. Li, Y. Guo, Z. Wang, and P. L. Dragotti, "Joint learning of 3D lesion segmentation and classification for explainable COVID-19 diagnosis," *IEEE Trans. Med. Imag.*, vol. 40, no. 9, pp. 2463–2476, Sep. 2021.

[29] R. Miron, C. Moisii, and M. Breaban, "Revealing lung affections from CTs. A comparative analysis of various deep learning approaches for dealing with volumetric data," 2020, *arXiv:2009.04160*.

[30] Y. D. Cid, V. Liauchuk, D. Klimuk, A. Tarasau, V. A. Kovalev, and H. Müller, "Overview of imagecleftuberculosis 2019—Automatic CT-based report generation and tuberculosis severity assessment," in *Proc. Conf. Labs Eval. Forum (CLEF)*, 2019.

[31] I. Yurdaisik, F. Nurili, A. G. Agirman, and S. H. Aksoy, "The relationship between lesion density change in chest computed tomography and clinical improvement in COVID-19 patients," *Int. J. Clin. Pract.*, vol. 75, no. 9, Sep. 2021, Art. no. e14355.

[32] Q. Yang, R. Zhang, Y. Gao, C. Zhou, W. Kong, W. Tao, G. Zhang, and L. Shang, "Computed tomography findings in patients with pulmonary tuberculosis and diabetes at an infectious disease hospital in China: A retrospective cross-sectional study," *BMC Infectious Diseases*, vol. 23, no. 1, p. 436, Jun. 2023.

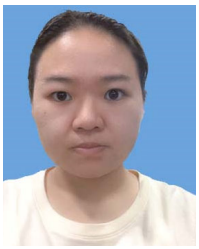
[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*.
[34] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[35] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," 2019, *arXiv:1905.11946*.
[36] J. Hofmanninger, F. Prayer, J. Pan, S. Röhrich, H. Prosch, and G. Langs, "Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem," *Eur. Radiol. Experim.*, vol. 4, no. 1, p. 50, Aug. 2020.
[37] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[38] C. Combi, B. Amico, R. Bellazzi, A. Holzinger, J. H. Moore, M. Zitnik, and J. H. Holmes, "A manifesto on explainability for artificial intelligence in medicine," *Artif. Intell. Med.*, vol. 133, Nov. 2022, Art. no. 102423.
[39] F. Prinzi, C. Militello, N. Scichilone, S. Gaglio, and S. Vitabile, "Explainable machine-learning models for COVID-19 prognosis prediction using clinical, laboratory and radiomic features," *IEEE Access*, vol. 11, pp. 121492–121510, 2023.

QIUSHUN BAI received the B.S. degree in biomedical engineering from Southern Medical University, Guangzhou, China, in 2021. He is currently pursuing the M.E. degree with the Department of Biomedical Engineering, Southern Medical University. His research interests include medical image analysis, machine learning, deep learning, and computerized-aided diagnosis.

YI GAO received the master's degree from the Department of Infectious Disease, Xiangya Medical College, Central South University, Changsha, China, in 2012. Her research interests include tuberculosis, infectious diseases, deep learning, and medical model construction.

FENG CHEN received the B.S. degree in clinical medicine from the Hubei University of Medicine, in 2004, the master's degree in medical imaging from Wuhan University, in 2006, and the Ph.D. degree in medical imaging from Southern Medical University, in 2015. Her research interests include neuroradiology, psychiatry, behavioral science, medical image analysis, computer-aided diagnosis, and machine learning.

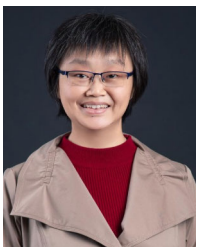
YIWEN ZHANG received the bachelor's degree in biomedical engineering from Southern Medical University, Guangzhou, China, in 2019, where he is currently pursuing the Ph.D. degree in engineering with the Department of Biomedical Engineering. His research interest includes the synthesis and segmentation of medical images.



YUAN YANG received the B.S. degree from the Department of Biomedical Engineering, Southern Medical University, Guangzhou, China, in 2018, where she is currently pursuing the Ph.D. degree. Her research interests include medical image analysis and computerized-aided diagnosis.

LIMING ZHONG received the B.S. and Ph.D. degrees in biomedical engineering from the Department of Biomedical Engineering, South Medical University, Guangzhou, China, in 2013 and 2019, respectively. Her research interests include medical image analysis, machine learning, deep learning, computerized-aided diagnosis, and medical image reconstruction.

WEI YANG received the bachelor's degree in industrial automation from the Wuhan University of Technology, China, the master's degree in control theory and control engineering from Xiamen University, China, and the Ph.D. degree in biomedical engineering from Shanghai Jiao Tong University, China. He is currently a Professor with the College of Biomedical Engineering, Southern Medical University, China. His research interests include medical image modality synthesis, intelligent analysis of medical images, intelligent analysis of medical signals, and radiomics. He is also a Council Member of the Guangdong Biomedical Engineering Society, China, the Standing Committee Member of the Medical Robotics and Artificial Intelligence Branch of the Guangdong Biomedical Engineering Society, China, and an Executive Member of the Medical Imaging Computing Seminar (MICS), China.



YI YANG received the bachelor's and master's degrees in computer science and technology from the National University of Defense Technology, China, and the Ph.D. degree in quantitative pathology from Southern Medical University, China. She is currently a Professor with the College of Biomedical Engineering, Southern Medical University. Her research interests include medical data analysis and processing, and medical information system design and development.