Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Segmentation of Pulmonary Cavity in Lung CT Scan for Tuberculosis Disease

Zhuoyi Tan[1], Hizmawati Madzin[1,*], Fatimah Khalid[1], Ng Seng Beng[1]

[1] Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang, 43400, Malaysia

ARTICLE INFO

Article history:
Received 23 June 2023
Received in revised form 25 September 2023
Accepted 10 October 2023
Available online 31 October 2023

Keywords:
Pulmonary tuberculosis; semantic segmentation; Deep Learning; object detection

ABSTRACT

The complexity of pulmonary tuberculosis (TB) lung cavity lesion features significantly increase the cost of semantic segmentation and labelling. However, the high cost of semantic segmentation has limited the development of TB automatic recognition to some extent. To address this issue, we developed an algorithm that automatically generates a semantic segmentation mask of TB from the TB target detection boundary box. Pulmonologists only need to identify and label the location of TB, and the algorithm can automatically generate the semantic segmentation mask of TB lesions in the labelled area. The algorithm, first, calculates the optimal threshold for separating the lesion from the background region. Then, based on this threshold, the lesion tissue within the bounding box is extracted and forms a mask that can be used for semantic segmentation tasks. Finally, we use the generated TB semantic segmentation mask to train Unet and Vnet models to verify the effectiveness of the algorithm. The experimental results demonstrate that Unet and Vnet achieve mean Dice coefficients of 0.612 and 0.637, respectively, in identifying TB lesion tissue.

1. Introduction

Tuberculosis is a common infectious disease. The computer-aided decision support system can effectively assist doctors in the diagnosis and treatment of tuberculosis [1-4]. In such a system, the segmentation of pulmonary tuberculosis tissue is a crucial step, which can separate the lesion area from the normal area in the CT scan image, thus helping doctors to diagnose and treat more accurately. Segmentation of tuberculosis lesions have been a challenging problem due to the complexity and variability of lung images. Therefore, it is necessary to develop an efficient and accurate segmentation method for pulmonary tuberculosis lesions.

Medical imaging, especially CT scanning, is an important tool for diagnosing and managing TB. However, the complexity of pulmonary lesion features significantly increases the cost of semantic segmentation and labeling. Manual segmentation and marking by trained medical professionals are time-consuming and costly, which is difficult to implement in resource-limited settings [5]. Therefore,

* *Corresponding author.*
E-mail address: hizmawati@upm.edu.my

https://doi.org/10.37934/araset.33.2.98106

an automatic segmentation and labeling method that can accurately identify and segment lung cavity lesions in lung CT scan image is needed.

In this paper, we propose an algorithm to automatically generate TB semantic segmentation mask from the TB target detection bounding box. The algorithm only needs medical professional to mark the TB area and the algorithm can automatically generate a semantic segmentation mask of TB within the marked area. Our algorithm can reduce the cost and time required for manual segmentation and labeling and achieve the development of automatic TB recognition.

The implementation principle of our algorithm: First, use the discrete gradient histogram [6] to find the optimal threshold to separate the lesion from the background area, and then extract the lesion within the bounding box according to the threshold and form a mask that can be used for semantic segmentation tasks. Finally, we train 3D Unet [7] and Vnet [8] models using the generated TB semantic segmentation masks to verify the effectiveness of the algorithm.

Experimental results show that our proposed algorithm able to segment and label pulmonary cavity lesions in lung CT scan based on bounding box labelled by medical professional. It can help the medical professional to improve the accuracy and efficiency of TB diagnosis and treatment procedure.

2. Methodology

In this research paper, ImageCLEF2022-TB dataset [9] is used to segmentize TB lesion tissue in lung CT scan image. The dataset contains lung CT scan image with lesion area bounding box. Then, the optimal threshold for CT scan image is calculated. Based on the threshold value, the lesion tissue of the TB target detection marked is extracted within the bounding box area.

2.1 Determination of TB Lesion Threshold

The pixel value distributions of TB CT scan images in the ImageCLEF2022-TB dataset are vary. This is because each of TB CT scan image contains of 125 slices of CT scan. Figure 1 illustrates the mean distribution of pixel values in four randomly selected CT scan images from the dataset. It is evident from the figure that the pixel value distributions in these four images are distinct. The calculation principle (pseudocode) of the discrete mean pixel histogram is shown in Figure 2.

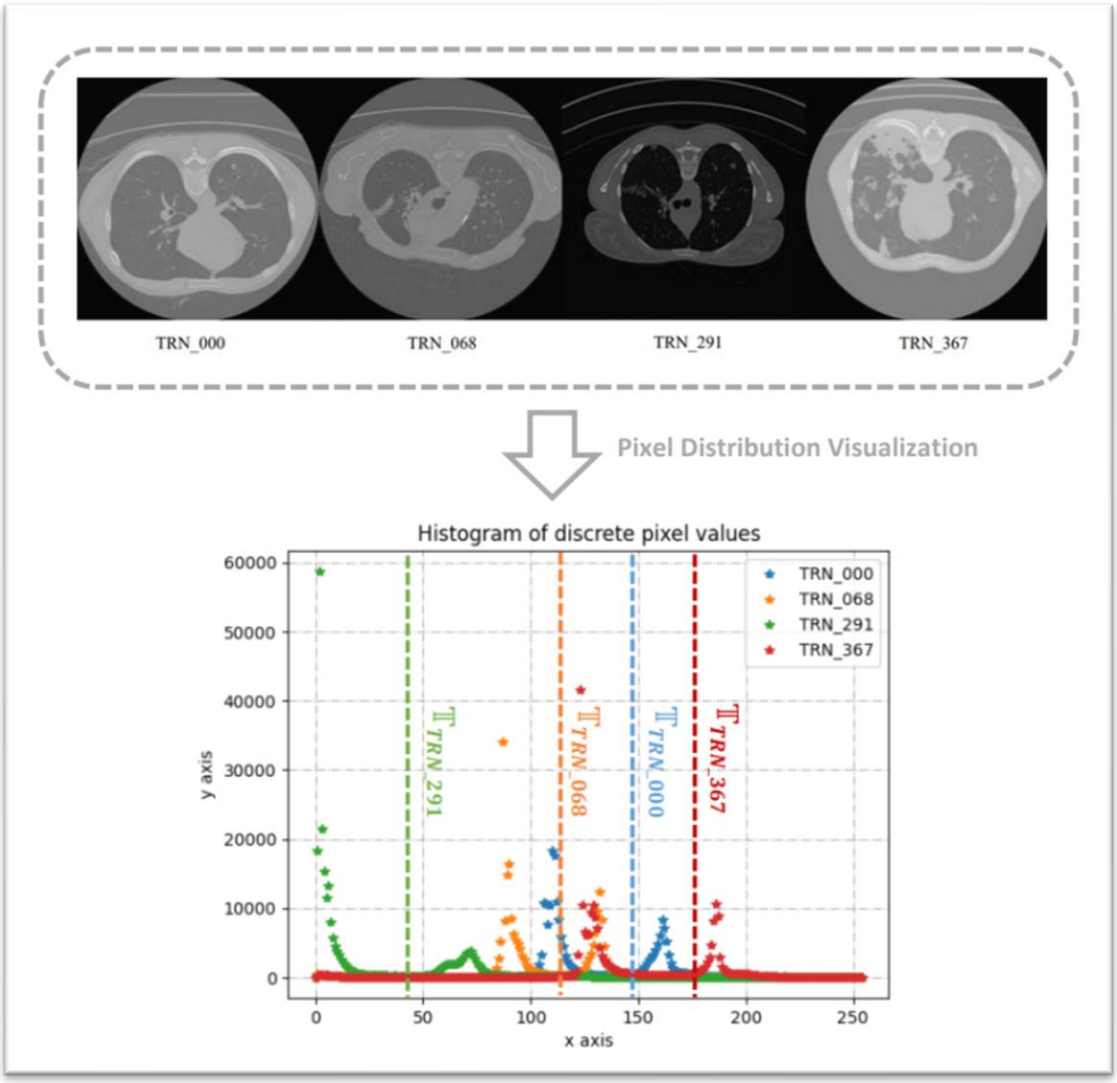


Fig. 1. ImageCLEF2022 sample pixel value distribution. TRN_000, TRN_068, TRN_291, TRN_367 represent the CT scan images from the ImageCLEF2022 dataset. The x axis is the pixels value, and the y axis is the frequency of counted pixels values

Algorithm Calculation of Mean Distribution of Pixel Values

Require: Let $\mathbb{C}$ be the entire pixel value distribution matrix of the CT scan

Require: Let n be the total number of slices in $\mathbb{C}$

for $i = 1$ *to* n **do**

Let $F_j^{S_i}$ be the frequency of pixel value j appearing in the i-th slice of $\mathbb{C}$

for $j = 0$ *to* 255 **do**

Calculate F_j as the average pixel value frequency of all slices of pixel value j in a CT scan image $\mathbb{C}$:

$$F_j = \frac{\sum_{i=1}^{n} F_j^{S_i}}{n}$$

Fig. 2. The calculation principle (pseudocode) of the discrete mean pixel histogram

In Figure 2, the experimental principle of the algorithm can be explained as follows:

First, assuming that the entire pixel value distribution matrix of the CT scan is denoted as $\mathbb{C}$ (as shown in Eq. (1))

$$\mathbb{C} = \begin{bmatrix} \mathrm{F}_0^{S_1} & \cdots & \mathrm{F}_{255}^{S_1} \\ \vdots & \ddots & \vdots \\ \mathrm{F}_0^{S_n} & \cdots & \mathrm{F}_{255}^{S_n} \end{bmatrix} \tag{1}$$

where, n represents the total number of slices that exist in $\mathbb{C}$. $\mathrm{F}_j^{S_i}$ represents the frequency of pixel value j appearing in the i^{th} slice ($S_i \in \mathbb{C}, i = 1, \cdots, n$) of $\mathbb{C}$.

Then, the calculation principle of the mean distribution of pixel values can be represented by Eq. (2)

$$\mathrm{F}_j = \frac{\mathrm{F}_j^{S_0} + , \cdots , + \mathrm{F}_j^{S_n}}{n} = \frac{\sum_{i=1}^{n} \mathrm{F}_j^{S_i}}{n}, j = 0, \cdots, 255, i = 1, \cdots, n \tag{2}$$

where, F_j represents the average pixel value frequency of all slices of pixel value j in a CT scan image $\mathbb{C}$.

In Figure 1, it shows there are differences in the mean distribution of pixel values of different CT scan images. However, the pixel value distribution and the average value of pixel values for slices in the same CT scan image is basically consistent. Therefore, the average value of pixel values can well reflect the distribution of pixel values of each slice in the entire CT scan image. Therefore, to determine the pixel threshold for TB lesion in lung CT scan image is by calculating the mean value of all slices in a CT scan image.

In this research paper, histogram of discrete pixel value (HDPV) is used to calculate distribution of pixel values. There are two sets of pixel value waves (excluding 0-value pixels) produced from the CT scan image. For these two sets of pixel value waves, we replace the first and second sets of pixel value waves with symbols $\mathbb{W}_1$ and $\mathbb{W}_2$ respectively, as shown in Figure 3. The peak values of these two groups of pixel value waves often have large fluctuations, some can reach up to 60,000 number of pixels, however, some only have more than 1,000 number of pixels. Although in some cases, there are several groups of wavelets within a group of waves, overall, these wavelets are continuous. The law shows that $\mathbb{W}_1$ and $\mathbb{W}_2$ respectively reflect the basic distribution of pixel values in the background area and diseased tissue area in the tuberculosis CT image. That is, the pixel values of the pulmonary tuberculosis lesion area are basically distributed after the dividing line $\mathbb{T}$ between $\mathbb{W}_1$ and $\mathbb{W}_2$ (this dividing line can also be understood as a specific threshold for extracting lesion tissue). Therefore, to realize the extraction of lesion tissue in pulmonary tuberculosis CT, we need to separate the pixel value area behind the dividing line $\mathbb{T}$. Figure 3 shows the dividing line $\mathbb{T}$ must be between the two groups of peaks, at the starting point of $\mathbb{W}_2$. It can easily determine that the position of $\mathbb{T}$ is half of the distance between these two sets of peaks. The specific implementation process for determining the dividing line $\mathbb{T}$ is as follows:

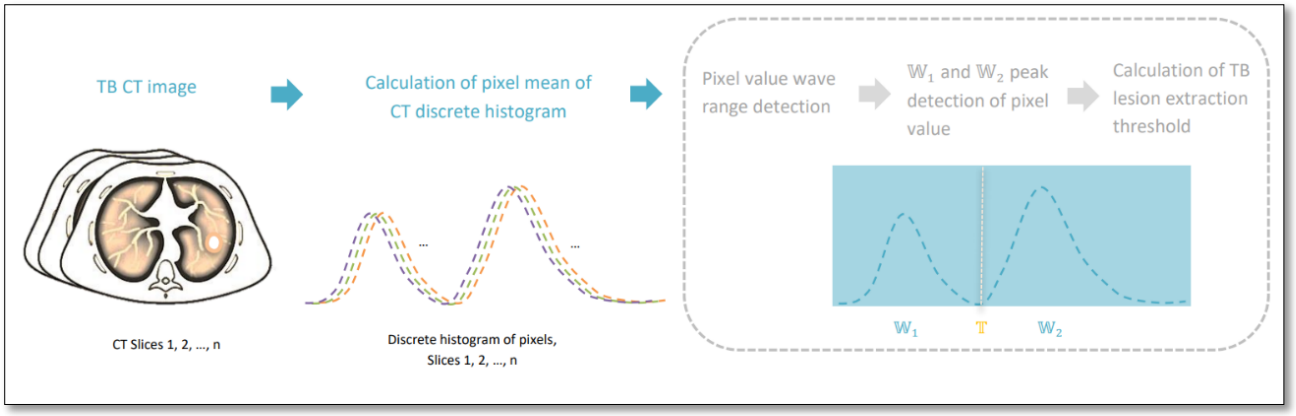


Fig. 3. Determination of TB lesion extraction threshold

2.2 Extraction of TB Lesions

Once the threshold value $\mathbb{T}$ is identified, the lesion tissue in the boundary frame region is extracted. The specific implementation of this process is as follows: First, we use $A\ (x_1, y_1)$ as the upper-left coordinate point and $B\ (x_2, y_2)$ as the lower-right coordinate point to construct a bounding box. Second, extract all pixel value points arranged after $\mathbb{T}$ ($\mathbb{T}$ is a exact pixel value in the $HDPV$) in the bounding box (as shown in Eq. (3))

$$\mathbb{X}_{\text{pv}}^{(x,y)} \geq \mathbb{T}, x_2 \geq x \geq x_1,\ y_2 \geq y \geq y_1 \tag{3}$$

where (x, y) are the coordinates of all pixel-value points within the bounding box. $\mathbb{X}_{\text{pv}}^{(x,y)}$represents the pixel value point with coordinates (x, y), which is reflected on the abscissa on HDPV.

2.3 Model Training Framework

Figure 4 shows the TB lesion extraction training framework constructed in this research paper. The framework can be divided into two stages:

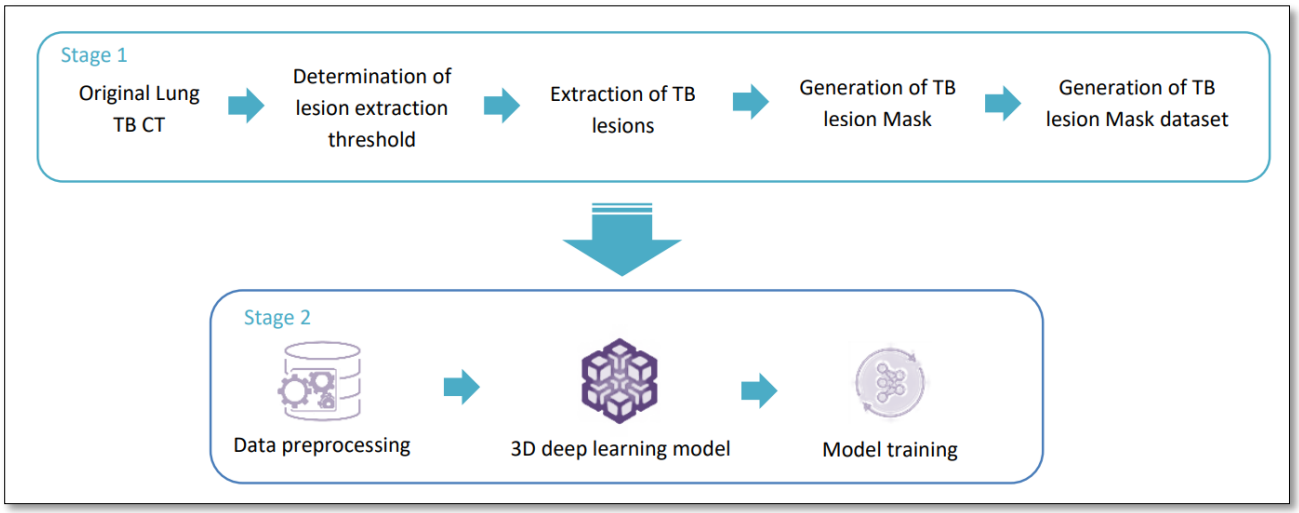
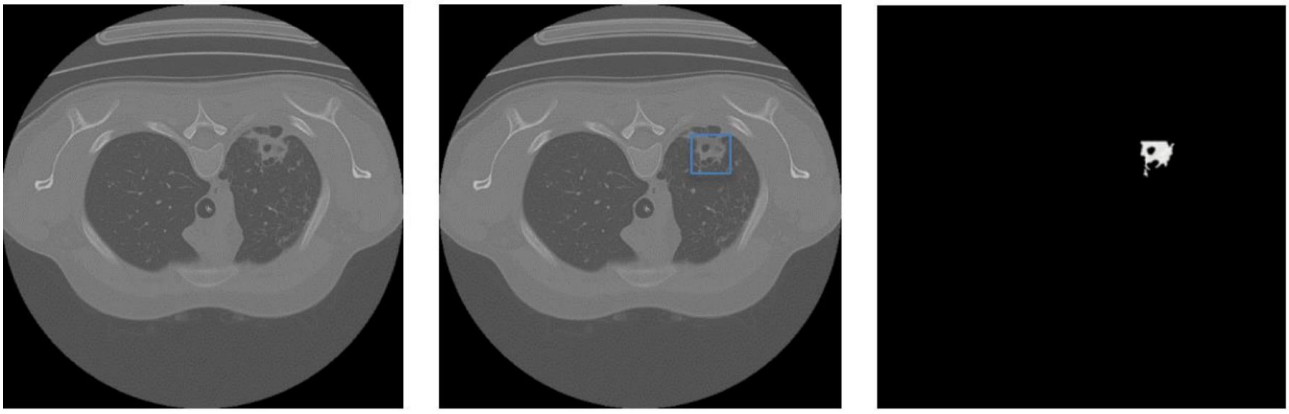


Fig. 4. TB lesion extraction model training framework

In the first stage, we complete the generation of the data set. The implementation process of this stage is as follows: First, determine the TB lesion threshold of the lesion tissue in lung CT scan image. Secondly, extract the areas where TB lesions exist in all slices in the lung CT scan image. Finally, a semantic segmentation mask of TB lesions is generated based on the extracted regions, as shown in Figure 5.



(a) (b) (c)

Fig. 5. TB lesion extraction. Panels (a) - (c) represents the original image, the image with bounding boxes, and the extracted TB lesion tissue mask, respectively

In the second stage, we assessed the performance of 3D Unet and Vnet models in identifying TB lesion tissue. We selected Unet and Vnet models as evaluation models for TB lesion tissue recognition due to their encoder-decoder structure, which enables efficient learning and extraction of lesion features from TB images. These models can capture features at various scales and acquire spatial information from the images, thereby facilitating precise segmentation of target structures in medical images. This structural characteristic makes Unet and Vnet models well-suited for accurately segmenting TB images, especially when dealing with complex-shaped or size-variable target structures.

The implementation process of this stage is as follows: First, preprocess the data, and then train the data into the deep learning model, and the model is trained. Finally, Figure 6 presents the ground truth of TB lesions alongside the recognition outcomes of Unet and Vnet. In the figure it shows that Vnet achieves superior performance in TB lesion recognition compared to 3D Unet, displaying a higher level of agreement with the ground truth. The segmentation results produced by Vnet closely resemble the actual TB lesions.

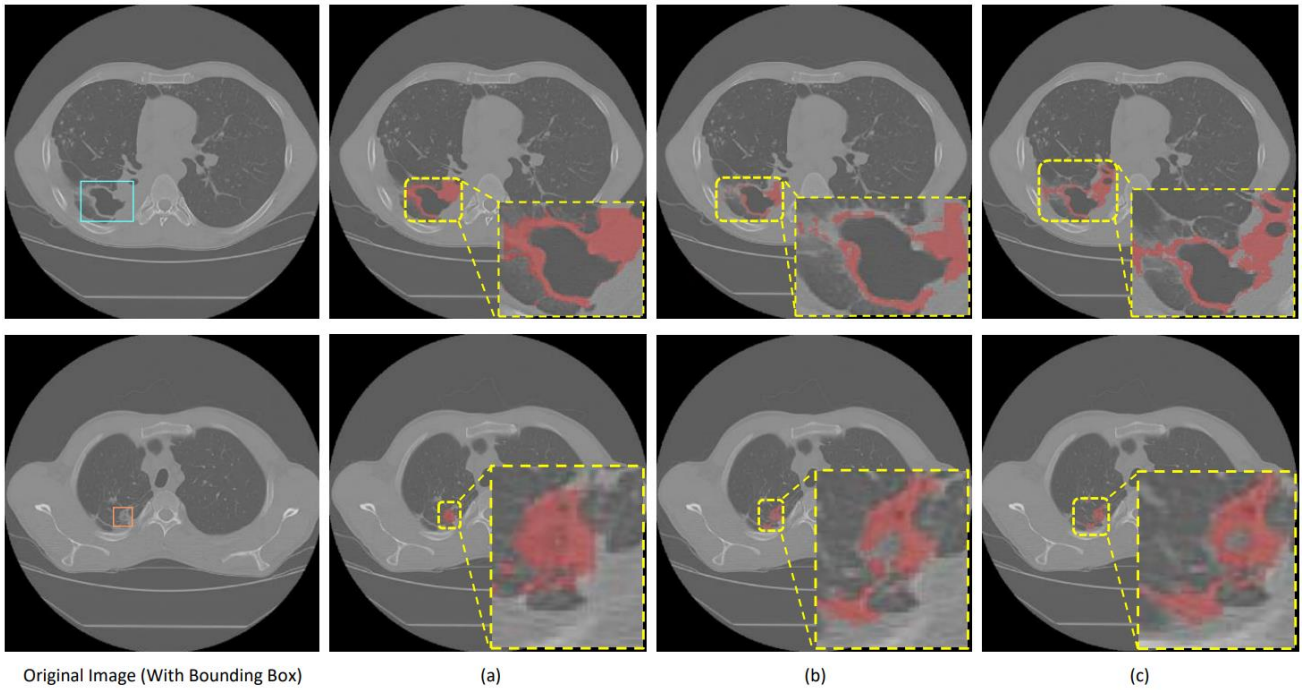


Fig. 6. TB lesion tissues extracted by different methods. Panels (a)-(c) represent the ground truth extracted using bounding box, the upper panels are recognition result of 3D Unet, and the below panels are recognition result of Vnet

3. Experimental Setting

3.1 Dataset and Data Pre-processing

As mentioned above, this research paper used ImageCLEF2022 TB dataset. The dataset participating in the experiment has a total of 75 CT scan images; each CT scan image encompass of 135 slices of image. Therefore, the total of CT scan images used in this experiment is 10,125 images. The training and test set are divided in a ratio of 8:2, which, 80% of data for training and the rest will be the testing data for verification.

In the pre-processing stage, all input images are uniformly adjusted the size to (96, 96, 96) [10,11]. Additionally, to avoid training getting stuck in overfitting [12], the images are randomly rotate and label data on the three axes by 90 degrees at a time [13]. Then, the pixel intensities of the input images are randomly shifted to make it more suitable for training [14,15].

3.2 Model Training

In the experiment, there are two models involved in the training, namely 3D Unet, and Vnet, and their training iterations are 6000 and 8000 respectively. The two models are consistent in the following training parameters: Batch size is 2; the learning rate is 1e-4, and the weight decay is 1e-5. The evaluation index for this training model performance verification is mean Dice Similarity Coefficient (mDice) [16,17].

4. Results and Discussion

Table 1 shows the mDice of 3D Unet and Vnet in training the CT scan images. It can be seen from the table of experimental results that the mDice index of the 3D Unet model is 0.612, and the mDice index of the Vnet model is 0.637, and the Vnet model is slightly better than the 3D Unet model. In

this experiment, the mDice index was used to evaluate the segmentation accuracy of the model, that is, the accuracy of the model's segmentation of the lesion area and the normal area. Since the Vnet model can better capture the local and global features in TB CT images [18-22], it has higher segmentation accuracy than 3D Unet.

The results of this experiment show that the Vnet model has better performance in identifying TB lesion tissue, and its mDice index is higher 3.9% than that of the 3D Unet model. This provides a certain reference for the research in the field of medical image segmentation.

Table 1
The performance of 3D Unet and Vnet in identifying TB lesion tissue

Model	3D Unet	Vnet
mDice	0.612	0.637

5. Conclusions

This paper presents an algorithm to automatically generate TB semantic segmentation masks from TB object detection bounding boxes to aid in the diagnosis and treatment of pulmonary tuberculosis. Compared with traditional manual segmentation and labelling methods, this algorithm can significantly reduce the cost and time of lung CT scan lung cavity lesion segmentation and improve the accuracy and efficiency of tuberculosis diagnosis and treatment planning. The algorithm extracts the lesion tissue within the bounding box by finding the optimal threshold and forms a mask that can be used for semantic segmentation tasks and uses the generated TB semantic segmentation mask to train 3D Unet and Vnet models to verify the effectiveness of the algorithm. Experimental results show that the Vnet model exhibits better performance than 3D Unet in the segmentation task of TB lesion tissue. The method proposed in this study could provide a more accurate and efficient aid in the diagnosis and treatment of tuberculosis, especially in resource-limited settings.

Acknowledgement
This research was funded by a grant from Ministry of Higher Education of Malaysia (FRGS Grant FRGS/1/2022/ICT01/UPM/02/1).

References
[1] Hrizi, Olfa, Karim Gasmi, Ibtihel Ben Ltaifa, Hamoud Alshammari, Hanen Karamti, Moez Krichen, Lassaad Ben Ammar, and Mahmood A. Mahmood. "Tuberculosis disease diagnosis based on an optimized machine learning model." Journal of Healthcare Engineering 2022 (2022). https://doi.org/10.1155/2022/8950243

[2] Tan, Zhuoyi, Hizmawati Madzin, and Zeyu Ding. "Semi-supervised Semantic Segmentation Methods for UW-OCTA Diabetic Retinopathy Grade Assessment." In Mitosis Domain Generalization and Diabetic Retinopathy Analysis: MICCAI Challenges MIDOG 2022 and DRAC 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18–22, 2022, Proceedings, pp. 97-117. Cham: Springer Nature Switzerland, 2023. https://doi.org/10.1007/978-3-031-33658-4_10

[3] Tan, Zhuoyi, Yuping Hu, Dongjun Luo, Man Hu, and Kaihang Liu. "The clothing image classification algorithm based on the improved Xception model." International Journal of Computational Science and Engineering 23, no. 3 (2020): 214-223. https://doi.org/10.1504/IJCSE.2020.111426

[4] Tan, Zhuoyi, Hizmawati Madzin, and Zeyu Ding. "Image Quality Assessment Based on Multi-model Ensemble Class-Imbalance Repair Algorithm for Diabetic Retinopathy UW-OCTA Images." In Mitosis Domain Generalization and Diabetic Retinopathy Analysis: MICCAI Challenges MIDOG 2022 and DRAC 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18–22, 2022, Proceedings, pp. 118-126. Cham: Springer Nature Switzerland, 2023. https://doi.org/10.1007/978-3-031-33658-4_11

[5] Farhat, Hanan, George E. Sakr, and Rima Kilany. "Deep learning applications in pulmonary medical imaging: recent updates and insights on COVID-19." Machine vision and applications 31 (2020): 1-42. https://doi.org/10.1007/s00138-020-01101-5

[6] Zuo, Wangmeng, Lei Zhang, Chunwei Song, and David Zhang. "Texture enhanced image denoising via gradient histogram preservation." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1203-1210. 2013. https://doi.org/10.1109/CVPR.2013.159

[7] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234-241. Springer International Publishing, 2015. https://doi.org/10.1007/978-3-319-24574-4_28

[8] Milletari, Fausto, Nassir Navab, and Seyed-Ahmad Ahmadi. "V-net: Fully convolutional neural networks for volumetric medical image segmentation." In 2016 fourth international conference on 3D vision (3DV), pp. 565-571. Ieee, 2016. https://doi.org/10.1109/3DV.2016.79

[9] Ionescu, Bogdan, Henning Müller, Renaud Péteri, Johannes Rückert, Asma Ben Abacha, Alba G. Seco de Herrera, Christoph M. Friedrich et al. "Overview of the ImageCLEF 2022: Multimedia retrieval in medical, social media and nature applications." In Experimental IR Meets Multilinguality, Multimodality, and Interaction: 13th International Conference of the CLEF Association, CLEF 2022, Bologna, Italy, September 5–8, 2022, Proceedings, pp. 541-564. Cham: Springer International Publishing, 2022. https://doi.org/10.1007/978-3-031-13643-6_31

[10] Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., Xu, D., Nath, V. and Hatamizadeh, A., 2022. Self-supervised pre-training of swin transformers for 3d medical image analysis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 20730-20740). https://doi.org/10.1109/CVPR52688.2022.02007

[11] Hatamizadeh, Ali, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R. Roth, and Daguang Xu. "Unetr: Transformers for 3d medical image segmentation." In Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp. 574-584. 2022. https://doi.org/10.1109/WACV51458.2022.00181

[12] Bejani, Mohammad Mahdi, and Mehdi Ghatee. "A systematic review on overfitting control in shallow and deep neural networks." Artificial Intelligence Review (2021): 1-48. https://doi.org/10.1007/s10462-021-09975-1

[13] Li, Kuan, Yi Jin, Muhammad Waqar Akram, Ruize Han, and Jiongwei Chen. "Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy." The visual computer 36 (2020): 391-404. https://doi.org/10.1007/s00371-019-01627-4

[14] Chaman, Anadi, and Ivan Dokmanic. "Truly shift-invariant convolutional neural networks." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3773-3783. 2021. https://doi.org/10.1109/CVPR46437.2021.00377

[15] Dawood, Muhammad, Amina Asif, and Fayyaz ul Amir Afsar Minhas. "Deep-PHURIE: deep learning based hurricane intensity estimation from infrared satellite imagery." Neural Computing and Applications 32 (2020): 9009-9017. https://doi.org/10.1007/s00521-019-04410-7

[16] Tanabe, Yoshinori, Takayuki Ishida, Hidetoshi Eto, Tatsuhiro Sera, and Yuki Emoto. "Evaluation of the correlation between prostatic displacement and rectal deformation using the Dice similarity coefficient of the rectum." Medical Dosimetry 44, no. 4 (2019): e39-e43. https://doi.org/10.1016/j.meddos.2018.12.005

[17] Korepanova, Anastasia A., Valerii D. Oliseenko, and Maxim V. Abramov. "Applicability of similarity coefficients in social circle matching." In 2020 XXIII International Conference on Soft Computing and Measurements (SCM), pp. 41-43. IEEE, 2020. https://doi.org/10.1109/SCM50615.2020.9198782

[18] Cao, Bingyi, Andre Araujo, and Jack Sim. "Unifying deep local and global features for image search." In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16, pp. 726-743. Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-58565-5_43

[19] Qiu, Zhaofan, Ting Yao, Chong-Wah Ngo, Xinmei Tian, and Tao Mei. "Learning spatio-temporal representation with local and global diffusion." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12056-12065. 2019. https://doi.org/10.1109/CVPR.2019.01233

[20] Li, Leida, Tianshu Song, Jinjian Wu, Weisheng Dong, Jiansheng Qian, and Guangming Shi. "Blind image quality index for authentic distortions with local and global deep feature aggregation." IEEE Transactions on Circuits and Systems for Video Technology 32, no. 12 (2021): 8512-8523. https://doi.org/10.1109/TCSVT.2021.3112197

[21] Wang, Dan, Terh Jing Khoo, and Zhangfei Kan. 2020. "Exploring the Application of Digital Data Management Approach for Facility Management in Shanghai's High-Rise Buildings". Progress in Energy and Environment 13 (May):1-15. https://www.akademiabaru.com/submit/index.php/progee/article/view/1063.

[22] Cheok, Choon Yoong, and Nyuk Ling Chin. 2017. "Antitumoral Properties of Xanthones from Mangosteen (Garcinia Mangostana L.) Hull". Progress in Energy and Environment 1 (June):20-24. https://www.akademiabaru.com/submit/index.php/progee/article/view/1031.