Multi-Classification Study of the Tuberculosis with 3D CBAM-ResNet and EfficientNet

Xing Lu[2], Eric Y Chang[1,2], Chun-Nan Hsu[1,2], Jiang Du[2] and Amilcare Gentili[1,2]

[1]*San Diego VA Health Care System, San Diego, CA, USA*

[2]*University of California, San Diego, CA, USA*

Abstract

The detection and characterization of tuberculosis along with the evaluation of tuberculosis lesion characteristics are challenging. To provide a solution for this multi-classification task, we performed a deep learning study that relied on the use of 3D Resnet and EfficientNet. With proper application of the provided masks, lung images were cropped, masked, and rearranged with different windowing. Stratified sampling for train/validation split and a balanced sampler in each batch sampler during training were used to address the data imbalance problem. A convolutional block attention model (CBAM) was used to add an attention mechanism in each block of the Resnet to further improve the performance of the convolutional neural network (CNN).

Keywords

Tuberculosis, Computed Tomography, Image Classification, Tuberculosis Type, CBAM, Efficient Net

1. Introduction

Tuberculosis (TB) is a bacterial infection caused by the germ Mycobacterium tuberculosis, and is a leading cause of death from infectious disease worldwide. An epidemic in many developing regions such as Africa and Southeast Asia, TB was responsible for 1.6 million deaths in 2017 alone. There are different manifestations of TB which require different treatments, making the detection and characterization of TB disease and the evaluation of lesion characteristics critically important tasks in the monitoring, control, and treatment of the disease. An accurate and automated method for classification of TB from computed tomography (CT) images would be especially useful in regions of the world with few radiologists.

The ImageCLEF 2021 Tuberculosis – CT report challenge [1] aimed to automatically categorize each TB case into one of the following five types: (1) Infiltrative, (2) focal, (3) tuberculoma, (4) miliary, and (5) fibro-cavernous [2]. In this edition, a dataset containing chest CT scans of 1338 TB patients was used: 917 images for the Training (development) dataset and 421 for the Test set. Because each CT scan corresponded to only one patient, each CT image corresponded to only one TB type at a time.

2. Methods

2.1. Data

The datasets provided for the tuberculosis task training set contained a total of 917 patients, with labeling provided for five categories. To avoid bias in the training and validation cohorts, a balanced train/validation strategy was employed to split each class according to an 8:2 ratio, as shown in Figure 1. Figure 1(a) shows the results of a random train/validation split, whereas Figure 1(b) shows the balanced split. As can be seen in Figure 1(a), in the fibro-cavernous class, there were only a few examples in the validation dataset, whereas the balanced split shown in Figure 1(b) resulted in each class having similar validation and training dataset.
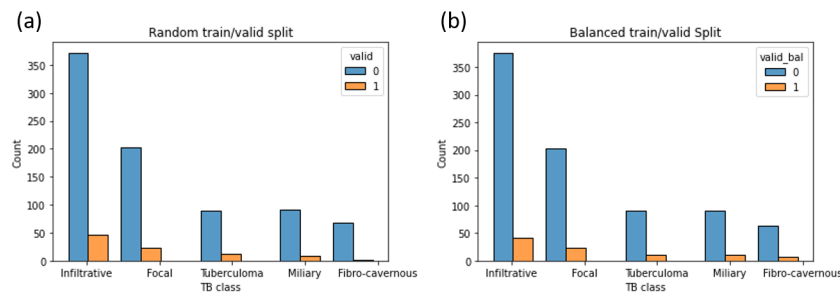

Figure 1: Stratified sampling of different classes for train/validation split

2.2. Preprocessing

The preprocessing of the images for the deep learning model is shown in Figure 2. The images for the ImageCLEF TB task were provided as NIFTI 3D datasets. Two versions of lung segmentation masks were also provided. The first version of segmentation (denoted as Mask-1) provided more accurate masks, containing individual masks for left and right laterality (values equal 1 for left and 2 for right), but in the most severe TB cases, there was a tendency to miss large abnormal regions in the lungs [3]. On the other hand, the second segmentation (denoted as Mask-2) provided less precise boundaries, given that it contained the entire lung area (i.e., both left and right sides of the lung), but was more stable in terms of including lesion areas [4]. As there was no need to locate lesions in terms of lung side, only Mask-2 was used in this study.

As shown in Fig. 2, the original NIFTI-formatted dataset was transformed into image data by first applying the NiBabel package. Next, the reformatted images were adjusted to three different window levels, namely baseline, lung, and soft tissue, and then normalized. For baseline window level, the foreground was obtained via the Otsu thresholding algorithm provided in the openCV package; for lung and soft tissue, the image levels were set as [-600,1500] and [50,350], respectively. Then, images were normalized to [0,1] with their mean and std values. Finally, all three windows and levels of data were saved, and annotation files were rearranged for use in further training.
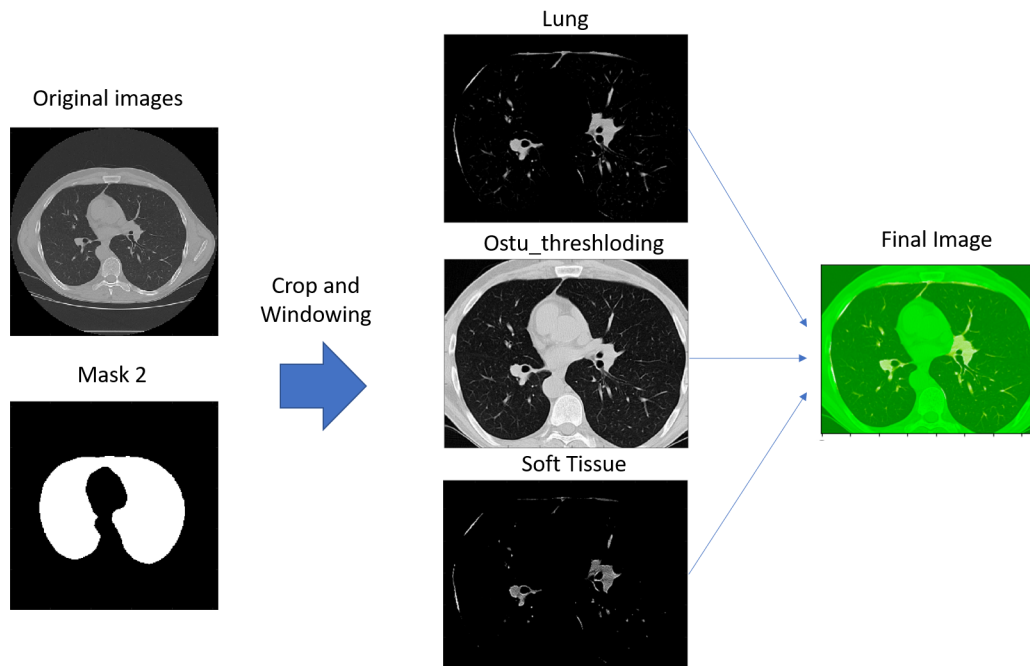
Figure 2: Preprocessing for the images.

2.3. Network and Training

In this study, a 3D convolutional block attention module (CBAM)-Resnet and a 3D EfficientNet were employed to train the model for 5-class classification based on the PyTorch framework. Similar to our last year's work [5], a standard 3D-resnet34 [6] was used as the convolutional neural network (CNN) backbone, with three fc layers as the classifier. CBAM [7]was used to implement channel and spatial attention mechanisms for each block of the Resnet, and sigmoid was used as the activation function for binary classification. according to our computing resources, EfficientNet B5 was the optimal 3D EfficientNet for training [8].

To train the neural networks, we used a workstation with 4 Nvidia GTX 1080 Ti video cards, 128 GB RAM, and a 1 TB solid state drive. During the training process, to avoid overfitting, image augmentation and a balanced sampler were implemented in each batch. For the image augmentation, traditional data augmentation methods, including brightness, shear, scale, and flip, were applied. The balanced sampler strategy, which equalized the data sampled from all five classes for each batch, was adopted during the training process.

2.4. Experiments and Model Selection

Three experiments were conducted during model training. For 3D Resnet, 20 datasets that fed into the network were dynamically generated from saved metadata with different window levels as a single channel and were interpolated into two kinds of data size: $3\times64\times256\times256$ and $3\times16\times384\times384$. For 3D EfficientNet, only $3\times64\times256\times256$ was used. For each experiment,

60 epochs with a cosine annealing warm-up learning rate were performed to train the model. To find the best model for each experiment during training, epochs with either minimum loss or highest accuracy were selected and saved for further submission.

3. Results and Submissions

The provided TST dataset included 421 image files for testing. With our preprocessing pipeline, the TST data were cropped according to Mask-2 to generate calibrated image files. After evaluation of the trained model, the results were rearranged according to the requirement and saved as a .txt file to be submitted. As was mentioned in the Methods, we saved six models with different metrics for evaluating the TST datasets; their performances are displayed in Table 1.

Table 1
Submission model types and results

Submission name	Model type	Tensor size	Accuracy	Kappa
194900_loss	Resnet34_loss	3×16×384×384	**0.371**	0.179
194900_acc	Resnet34_acc	3×16×384×384	0.354	0.163
154940_loss	Resnet34_loss	3×16×384×384	**0.371**	**0.190**
154940_acc	Resnet34_acc	3×16×384×384	0.363	0.177
060907_loss	Efficient B5_loss	3×64×256×256	0.352	0.130
060907_acc	Efficient B5_acc	3×64×256×256	0.361	0.155
135762	Model_ensembling	-	0.359	0.171

Per the submitted results, 3D Resnet34 achieved both better accuracy and Kappa than EfficientNet B5. For 3D Resnet34, the model with tensor size 3×64×256×256 and a loss-based selection model achieved a superior Kappa result of 0.190 and accuracy of 0.371, with submission name of 154940_loss. We also tried to assemble the results from different models into a single submission,137652, but the result was not significantly improved.

4. Discussion and Conclusion

To provide a deep learning solution for a multi-classification task of tuberculosis, we performed experiments using 3D CBAM Resnet and 3D EfficientNet as CNN backbones. There were several challenges for this task, such as the severe class imbalance and 3D dimensionality of the CT images, so we tried several techniques to improve the models' performance. First, we properly applied stratified sampling of each class for the train/validation split to mitigate bias in the training and validation cohorts. Furthermore, a balanced sampler in each batch sampler was used to address the data imbalance problem. Second, CBAM was used to add an attention mechanism to each block of the Resnet to further improve the performance of the CNN. Third, different windowings of the CT images were concatenated to further focus the CNN on features of the illness according to a radiologist. Using all the aforementioned techniques, we achieved a kappa of 0.190 in the evaluation of the test dataset and placed third in this competition.

5. Acknowledgments

References

[1] B. Ionescu, H. Müller, R. Peteri, A. Ben Abacha, M. Sarrouti, D. Demner-Fushman, S. A. Hasan, S. Kozlovski, V. Liauchuk, Y. Dicente, V. Kovalev, O. Pelka, A. G. S. de Herrera, J. Jacutprakart, C. M. Friedrich, R. Berari, A. Tauteanu, D. Fichou, P. Brie, M. Dogariu, L. D. Ştefan, M. G. Constantin, J. Chamberlain, A. Campello, A. Clark, T. A. Oliver, H. Moustahfid, A. Popescu, J. Deshayes-Chossart, Overview of the ImageCLEF 2021: Multimedia retrieval in medical, nature, internet and social media applications, in: Experimental IR Meets Multi-linguality, Multimodality, and Interaction, Proceedings of the 12th International Conference of the CLEF Association (CLEF 2021), LNCS Lecture Notes in Computer Science, Springer, Bucharest, Romania, 2021.

[2] S. Kozlovski, V. Liauchuk, Y. Dicente Cid, V. Kovalev, H. Müller, Overview of ImageCLEFtuberculosis 2021 - CT-based tuberculosis type classification, in: CLEF2021 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Bucharest, Romania, 2021.

[3] Y. Dicente Cid, O. A. Jiménez del Toro, A. Depeursinge, H. Müller, Efficient and fully automatic segmentation of the lungs in ct volumes, in: O. Goksel, O. A. Jiménez del Toro, A. Foncubierta-Rodríguez, H. Müller (Eds.), Proceedings of the VISCERAL Anatomy Grand Challenge at the 2015 IEEE ISBI, CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, 2015, pp. 31–35.

[4] V. Liauchuk, V. Kovalev, Imageclef 2017: Supervoxels and co-occurrence for tuberculosis ct image classification, in: CLEF2017 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Dublin, Ireland, 2017.

[5] X. Lu, E. Y. Chang, Z. Liu, C. Hsu, J. Du, A. Gentili, Imageclef2020: Laterality-reduction three-dimensional cbam-resnet with balanced sampler for multi-binary classification of tuberculosis and CT auto reports, in: L. Cappellato, C. Eickhoff, N. Ferro, A. Névéol (Eds.), Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, September 22-25, 2020, volume 2696 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020. URL: http://ceur-ws.org/Vol-2696/paper_70.pdf.

[6] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, 2015. arXiv:1512.03385.

[7] S. Woo, J. Park, J.-Y. Lee, I. S. Kweon, Cbam: Convolutional block attention module, 2018. arXiv:1807.06521.

[8] M. Tan, Q. V. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, 2020. arXiv:1905.11946.