Lijie at ImageCLEFmed Tuberculosis 2021: EfficientNet Simplified Tuberculosis Case Classification

Jie Li[1], Li Yang[2] and Yang Bai[3]

[1]*School of Information Science and Engineering, Yunnan University, Kunming 650091, P.R.China*

[2]*Computer Science and Engineering,North Minzu University, Ningxia 750021, P.R.China*

[3]*School of Information Science and Engineering, Yunnan University, Kunming 650091, P.R.China*

Abstract

Tuberculosis has been a bacterial infectious disease known to humans for a long time, and different types of tuberculosis cases have different treatment methods. Therefore, rapid and accurate positioning of its classification is essential for the treatment of tuberculosis. This paper describes our experiment in ImageCLEF 2021 Tuberculosis-TBT classification. We first use two versions of automatically extracted masks of the lungs which are provided by the organizer to process all the CT images. Then the image is transformed into a 2-dimensional through data preprocessing that is projection, and data enhancement is used in the training process. Through the modified version of the EfficientNet network pre-trained on Image Net, and the introduction of the ghost module, and fine-tuning the model after migration learning, the one classification task with 5 labels required by this challenge was completed. The final kappa score of our team in ImageCLEF 2021 Tuberculosis-TBT classification is 0.015, and the acc score is 0.380.

Keywords

EfficientNet, Transfer learning, Tuberculosis classification, Ghost

1. Introduction

Tuberculosis is a chronic, late-onset, highly morbid, and dangerous respiratory infectious disease. It is a bacterial disease, an airborne disease that uses droplets released when a patient coughs to attack the respiratory system. According to statistics, about 1.5 million people died of chronic tuberculosis in 2018, and the number of people infected with tuberculosis is about 6 times more than the number of people who die from tuberculosis. However, with the advancement of medicine, most patients can be cured clinically if they can be diagnosed in time, treated rationally, improved resistance, or reduced cell-mediated allergic reactions. Therefore, it is very important to detect and diagnose tuberculosis cases quickly and accurately.

At present, the clinical diagnosis of tuberculosis is mainly a comprehensive analysis of the patient's clinical manifestations, imaging data, and laboratory test results. Among them, the use of CT with higher resolution and more detailed evaluation is more common. When tuberculosis

affects the lungs, several visual patterns can be seen in CT images. These models can better reflect the characteristics of the type of tuberculosis, and using these characteristics to determine the classification of the specific type of tuberculosis. Due to the large delay between the onset of disease symptoms and the clinical manifestations or diagnosis of tuberculosis, it is important to find a quick and convenient early diagnosis method within this delay. In this way, the use of antibiotics, a general treatment method that harms the body, can be reduced, and the efficiency of early diagnosis of tuberculosis can be improved.

ImageCLEF [1] began to organize image classification and retrieval challenges in 2003 [2]. Among them, the tuberculosis task only started in 2007 and the tasks vary from year to year. The previous tasks include multidrug resistance (MDR) tuberculosis detection, tuberculosis type (TBT) classification and severity score (SVR), etc. The other difference lies in the changes in the data set. Due to the difficulty of each task, the official plan is to solve them one by one. Therefore, the task of 2021 [3] was focused on the classification of tuberculosis types (TBT).

The rest of this article is organized as follows: In the second part, we briefly introduce the literature review on ImageCLEF Tuberculosis. The third part introduces the specific tasks of this competition and analyses the data set in detail. Then, in the fourth part, we introduce the network model and transfer learning method. The fifth part introduces the relevant details of the experiment. In the sixth part, we report the results of our experiment. Finally, in the seventh part, we summarize and discuss the current and future development directions.

2. Related Work

Image recognition is considered to be the earliest application field of machine learning. In recent years, deep learning [4] has achieved amazing results in the field of medical imaging, and machine learning and deep learning have gradually appeared in tuberculosis diagnosis. Among them, deep learning is also frequently used to diagnose CT scan images of lung diseases. With the rapid development and improvement of the mathematics and computing power of human computers, CNN is considered to be an advanced feedforward neural network containing convolution calculation and its deep structure [5]. It also has good image classification performance and can be used for various image classification tasks. Therefore, CNN is also widely used in the diagnosis of tuberculosis.

Since the appearance of the tuberculosis task in the ImageCLEF competition in 2017, the approximate methods are to manually select meaningful axial CT slices first. Then using the semantic features extracted by two-dimensional CNN to describe these slices. Finally, the neural network sets the classifier to make the final category classification representation, such as the method of the MostaganemFSEI group [6]. Or by dividing the lung field into several sub-regions, and treating these sub-regions as nodes of a graph, the lungs can be modeled as a graph. Then, they defined the weighted edges between adjacent sub-regions, where the weighted edges encode the distance between the 3D texture descriptors obtained in each sub-region (node). To compare the obtained pictures, they converted these pictures into a lung description vector and used a support vector machine to classify them, such as MedGIFT [7].
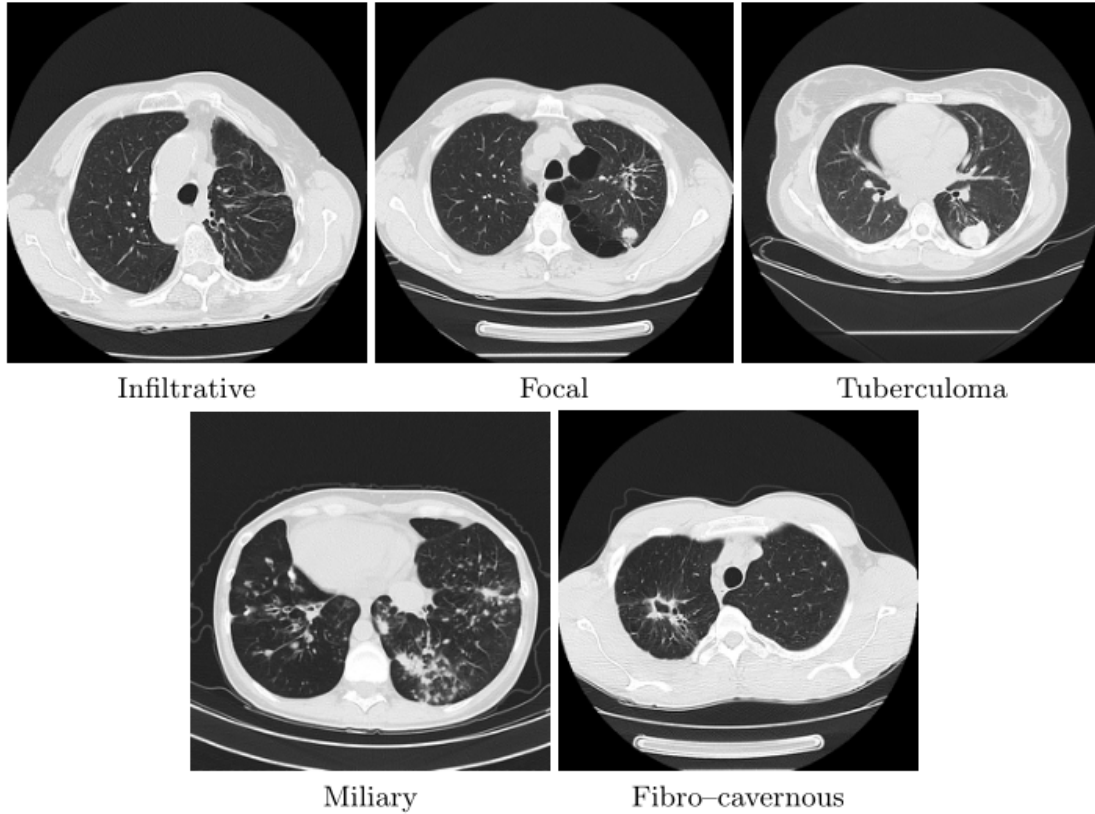
Figure 1: The specific images of each TB category

3. Task and Dataset

The tuberculosis task in ImageCLEF 2021 is based on 1,338 chest 3D CT image data of TB patients. There are 5 labels: Infiltrative, Focal, Tuberculoma, Miliary, Fibro-cavernous". Those types are originally labeled manually, and the specific images of each category are shown in Figure 1. Compared with 2017 and 2018, the data set has been updated. The training set provides metadata items including LeftLungAffected, RightLungAffected, LungCapacityDecrease, Calculation, Pleurisy, Caverns.

The tuberculosis classification task in 2021 is actually one of the 2017 and 2018 subtasks. The original file stores the original voxel intensity and the corresponding image metadata. Among them, there are 917 images in the training set and 421 images in the test set. In addition, the organizer provides two different segmentation methods for us to use the data. The mask#1 [8] provided by the first segmentation method is more accurate but the most severe TB cases will miss larger lung abnormal areas. The second mask#2 [9]inaccurately provides rougher boundaries, but it is more stable in terms of including the lesion area. In the experiment, we conducted experiments on the data processed by the two methods and also proved that the effect of the second method is better than the first one. Table 1 lists the number of metadata

Table 1
Metadata the amount of information in the training set

Filename	In Training set
RightLungAffected	197
LeftLungAffected	181
LungCapacityDecrease	69
Calcification	28
Pleurisy	13
Caverns	104

Table 2
The number of classified label information in the training set

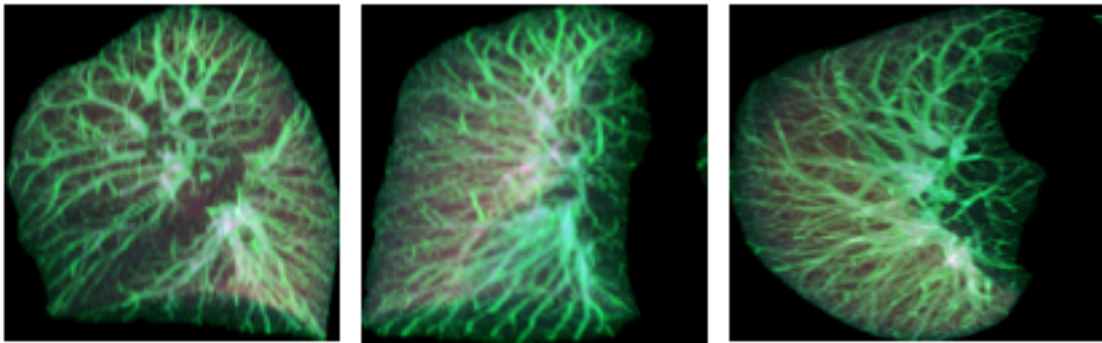| Filename	In Training set
Infiltrative	420
Focal	226
Tuberculoma	101
Miliary	100
Fibro-cavernous	70

Figure 2: The image projected on the x, y, z axis with mask#1

information, and Table 2 lists the number of classified tag information.

The ImageCLEF2021 tuberculosis task initially provided a data set in the form of NIfTI 3D, and we processed the data according to the two officially provided masks before starting the experiment. Then project on each coordinate axis to transform into a two-dimensional image. In this way, we can better invest in the neural network for training. Figure 2 is the image projected on the x, y, z-axis with mask#1, and Figure 3 is the image projected on the x, y, z-axis with mask#2.
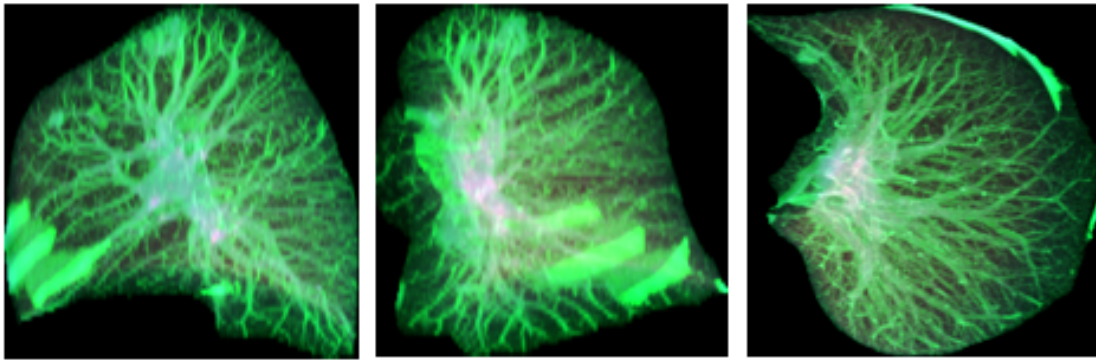
Figure 3: The image projected on the x, y, z axis with mask#2

4. Methods

4.1. Model

In recent years, when dealing with image classification problems, we tend to choose networks with a large number of calculations for processing. Facts have proved that network structures like ResNet combined with DenseNet and Attention tend to show good results in the extraction of image features. But those have some disadvantage such as the amount of calculation is too large, and the network transmission speed will be much slower than that of a lightweight network. Therefore, to meet the requirements of better feature extraction and improved speed, we use the EfficientNet+Ghost module.

Convolutional neural networks are generally developed under limited hardware resources and budgets. If more resources are available in the experiment, which can help us get better accuracy by expanding the scale of the model. The expanded model is generally divided into three aspects: model width, depth, and input image resolution. Most of the existing models adjust these three aspects manually, but such a combination space is too large to exhaust manpower, so the EfficientNet model came into being. It mainly uses a composite coefficient to simultaneously enlarge the network from the three dimensions of model depth, width, and resolution. Based on neural structure search technology, an optimal set of parameters (composite coefficients) can be obtained [10]. Figure 4 is the basic model, and Figure 5 is the EfficientNet model. We can see that the three dimensions have been expanded, and how much is expanded is obtained through the compound model expansion method combined with neural structure search technology.

Since the existing CNN extracts a large number of redundant features for image cognition, Han [11]et al. proposed GhostNet, which effectively reduces the computational cost. They found that there are many similar feature maps in the traditional convolutional layer, and these feature maps are called ghosts. The main contributions are as follows: They divide the traditional convolutional layer into two parts. In the first part, there are fewer convolution kernels directly used for feature extraction, which is the same as the original convolution. Then, these features are linearly transformed to obtain multiple feature maps, as shown in Figure 5.
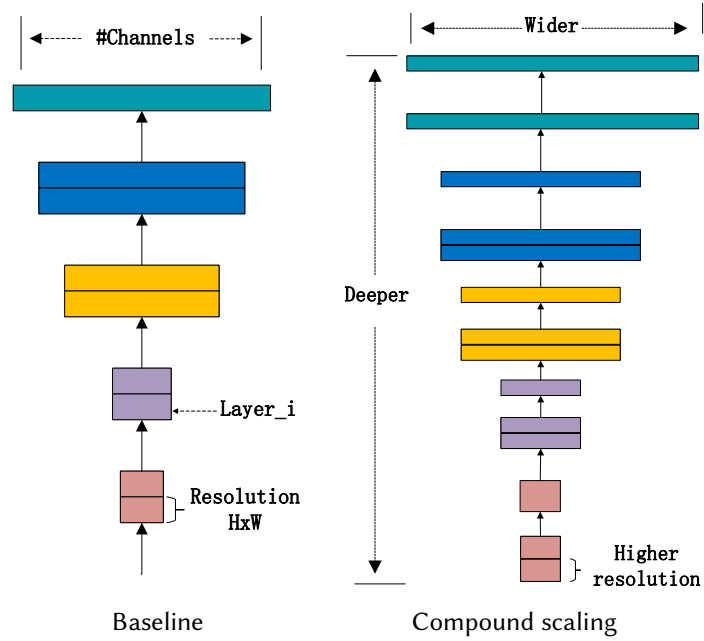
Figure 4: Comparison of basic model and EfficientNet

They proved that the Ghost module is also applicable to other CNN models. Ghost is different from 1*1 convolution in that it can control the size of the convolution kernel without restriction and use the same method of parallel mapping and linear transformation. We can use the Ghost module as a simple and convenient and effective component to improve the convolutional neural network.
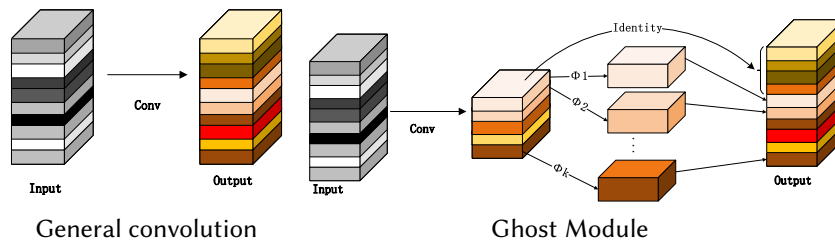

Figure 5: Comparison of basic model and Ghost

4.2. Transfer Learning

Due to the complexity of 3D CT image data and the diversity of data representation, including changes in the size and shape of CT image voxels, many image analysis algorithms are difficult to apply correctly, and the complexity and high cost of manually labeling 3D CT images, resulting in a lack of a large amount of high-quality labeling data. So transfer learning can transform existing data to acquire basic knowledge to assist future learning. Transfer Learning [12]aims to use the knowledge gained from the current context as the main means to complete learning in

another context. It is also a method of machine learning. The parameters of the training model developed for task A can be used as the initial point and transferred to the model developed for task B to help train the new model. Initializing the network with migration parameters can improve generalization performance. Even migrating from tasks that are not particularly similar is better than using random filters (or random parameters). Because deep learning models require a lot of training time and training data to get good results, and transfer learning can alleviate this necessity. This method has shown remarkable performance in a variety of medical image classification frameworks [13, 14].

In the experiment, we transfer the feature extraction capabilities learned by the neural network on the ImageNet dataset as prior knowledge to the feature extraction of CT image projection. After retaining the weights of EfficientNet on its data set, first, we choose to freeze the weights of the first k layers in the pre-trained model. Then we train the remaining n-k layers. Finally, it needs to be adjusted according to the corresponding output and format to make it meet the output of the task.

5. Experiment

5.1. Data preprocessing

In the experiment, because the official data set is the 3D image, and more computational overhead is required. In the case of limited GPU memory, heavy memory and computational requirements need to be overcome. Therefore, it is necessary for us to first convert the image into the 2D image in the image preprocessing, and it is also necessary to preserve the information of each dimension of the 3D image as much as possible. The method we use is to first process the data according to the two officially provided mask methods. Then the image is projected on the three coordinate axes, namely sagittal, frontal and axial projection, which makes the image easier to process and also increases the amount of data, which makes the training information richer. In addition, some data enhancement methods are used to prevent overfitting and refine image details, such as image enhancement by rotation, cropping, or whitening techniques. Since there is no verification set, we will separate the training set from the training set for verification at a ratio of 3:1.

The main steps that need to be performed to generate a CT projection image are: the voxel intensity value of the CT image is increased by 1024 Hounsfield Units (HU) to ensure that it is only a positive intensity value. To eliminate the effect of image segmentation defects on the edge of the lung mask, an ellipsoid structure element with a radius of 10 is used to corrode the lung membrane along the XY plane. The radius along the Z-axis is calculated as the inter-layer distance of a specific CT image. The image intensity value outside the corroded lung mask is set to zero. The intensity threshold can be used to filter low-intensity noise voxels. The advantage of this method over traditional slice representation is that each projection obtained by averaging the value of voxel intensity contains information about all slices in the image. On the one hand, this method greatly reduces the complexity of input and training data transmission and greatly reduces the task of learning on limited training data. On the other hand, generating projections along different axes (X, Y, Z) provides additional "native" enhancements to the input data.
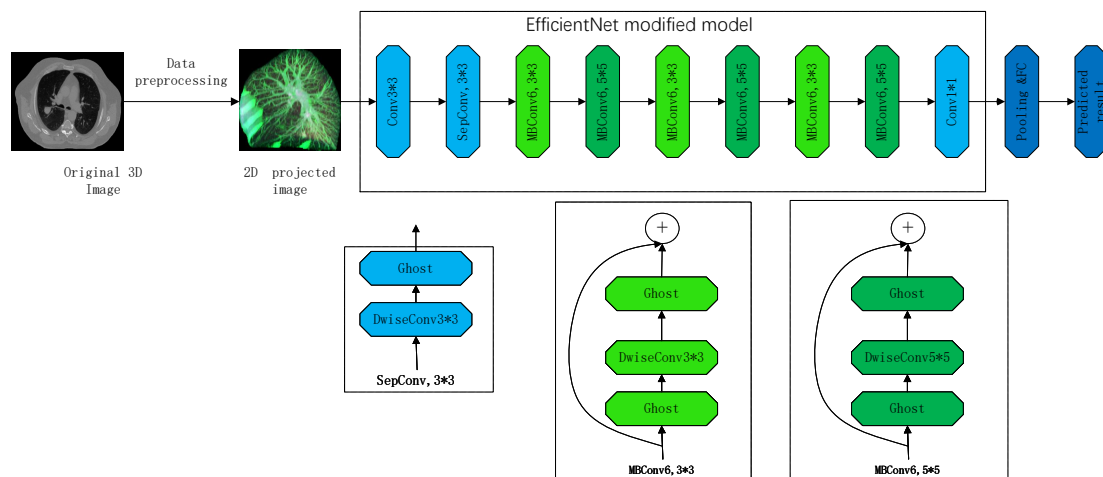
Figure 6: The actual model we used in the competition

5.2. Experimental model and analysis

The experimental process is shown in figure 6. We first convert 3D images into 2D images through preprocessing methods such as projection. Then through the EfficientNet network, and the network uses transfer learning, and transfers the feature extraction capabilities learned on the Image Net dataset as prior knowledge to the feature extraction of CT image projection. Here we simplify the structure of the original EfficientNet, remove the final fully connected layer, and use the new linear layer to implement the one classification task with 5 labels of this year's task. Subsequently, we unfroze the final convolutional layer of EfficientNet and iterated normally. This prevents the network from overfitting during training and allows us to train the model for a longer time for better generalization. In addition, we turn the 1∗1 convolution in EfficientNet into a ghost module to effectively reduce costs. Due to the official provision of two masks, we use the processed data of the two masks to train separately, and the final results show that mask2 performs better. The specific model structure of the experiment is shown in Figure 6. Among them, the EfficientNet module has been streamlined and improved.

In our experiment, a total of 100 epochs were trained. The Adam used by the optimizer has a learning rate of 1e-05 and a batch size of 32. During the training process, it stabilized at about 60 epochs, and then we fine-tuned the model, and finally performed best at 100 epochs.

6. Results

After all the CT images have been processed by the two masks, then the training set is divided into the verification set according to 3:1, and the accuracy rates of the final verification sets of the two mask data are shown in Table 3. The evaluation methods used in this ImageCLEF 2021 Tuberculosis-TBT classification task are kappa and acc. Our final kappa score is 0.015 and our acc score is 0.380, as shown in Figure 7.

Table 3
Results achieved on local validation set for two mask types

ValidationSet	Accuracy
MASK#1	0.416
MASK#2	0.432
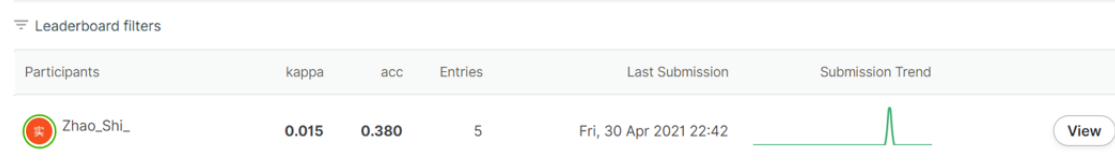
Figure 7: The final official score of this challenge

7. Perspectives For Future Work

In this task, before experimenting with images, the way of preprocessing the data needs to be improved. For example, the model used can only process 2D images, so that a lot of important feature information is lost when the 3D projection is changed to 2D. In addition, the pre-training before transfer learning is performed on the ImageNet dataset. However, general image data sets and medical data sets are still quite different. Therefore, there is still much room for improvement in these two specific aspects of data preprocessing.

In this competition, since we are only limited to the implementation of deep learning knowledge. However, due to our lack of medical knowledge and in-depth understanding of CT images, the importance of medical image features or label information has been neglected, resulting in poor results. In addition, we should be aware of the latest developments in deep learning, and we should not only limit the design of models to CNN. Recently, many excellent teams have begun to study the work of visual Transformer. The Transformer module has achieved amazing results in the field of natural language processing and has become a standard configuration in the field of NLP, but in the field of imagery, CNN (such as ResNet, DenseNet, etc.) occupies most of the SOTA results. In future work, we will apply the Transformer in the NLP field to the image field as much as possible. I believe that shortly, Transformer can also achieve the same excellent results in the image field as the NLP field.

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

Multilinguality, 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] H. Müller, P. Clough, T. Deselaers, B. Caputo, ImageCLEF: experimental evaluation in visual information retrieval, volume 32, Springer Science & Business Media, 2010.

[3] 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, Bucharest, Romania, 2021.

[4] X. Zhang, X. Zhou, M. Lin, J. Sun, Shufflenet: An extremely efficient convolutional neural network for mobile devices, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 6848–6856.

[5] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, et al., Recent advances in convolutional neural networks, Pattern Recognition 77 (2018) 354–377.

[6] A. Hamadi, N. B. Cheikh, Y. Zouatine, S. M. B. Menad, M. R. Djebbara, Imageclef 2019: Deep learning for tuberculosis ct image analysis., in: CLEF (Working Notes), 2019.

[7] Y. D. Cid, H. Müller, Lung graph-model classification with svm and cnn for tuberculosis severity assessment and automatic ct report generation., in: CLEF (Working Notes), 2019.

[8] 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.

[9] 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.

[10] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: International Conference on Machine Learning, PMLR, 2019, pp. 6105–6114.

[11] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, C. Xu, Ghostnet: More features from cheap operations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 1580–1589.

[12] L. Torrey, J. Shavlik, Transfer learning, in: Handbook of research on machine learning applications and trends: algorithms, methods, and techniques, IGI global, 2010, pp. 242–264.

[13] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, et al., Identifying medical diagnoses and treatable diseases by image-based deep learning, Cell 172 (2018) 1122–1131.

[14] M. Raghu, C. Zhang, J. Kleinberg, S. Bengio, Transfusion: Understanding transfer learning for medical imaging, arXiv preprint arXiv:1902.07208 (2019).