ImageCLEF 2020: Deep Learning for Tuberculosis in Chest CT Image Analysis based on multi-axis projections

Tetsuya Asakawa[1] and Masaki Aono[2]

[1] Department of Computer Science and Engineering, Toyohashi University of Technology, Aichi, Japan
asakawa@kde.cs.tut.ac.jp
[2] aono@tut.jp

Abstract. ImageCLEF 2020 Tuberculosis Task is an example of the challenging research problem in the field of CT image analysis. The purpose of this research is to make accurate estimates for the three labels (affected, pleurisy, caverns) for each of the lungs. We describe the tuberculosis task and approach for chest CT image analysis, then perform multi-label CT image analysis using the task dataset. We propose fine-tuning deep neural network model that uses inputs from multiple CNN features. In addition, this paper presents two approaches for applying mask data to the extracted 2D image data and for extracting a set of 2D projection images along multi-axis based on the 3D chest CT data. Our submissions on the task test dataset reached a mean AUC value of about 75% and a minimum AUC value of about 69%

Keywords: Computed Tomography, Tuberculosis, Deep Learning, Multi-label classification.

1 Introduction

With the spread of various virus (such as Tuberculosis, Coronavirus, and Influenza), medical researchers perform to give the necessary treatment for viruses in recent years. However, there is nothing to identify the disease early. Early diagnosis needed to give the necessary treatment, develop specific medicine, and prevent the death of patients. Therefore, several researchers have invested their efforts in recent years, especially within the medical image analysis community. In fact, a task dedicated to the tuberculosis had been adopted as part of the ImageCLEF evaluation campaign in its editions of the four last years [1][2][3][4]. In ImageCLEF 2020 the main task [5], "ImageCLEFmed Tuberculosis" is considered to be CT Report (CTR). In the task, the problem consists of generating an automatic report that includes the following information in binary form (0 or 1): Left Lung Affected, Right Lung Affected, Caverns Left, Caverns Right, Pleurisy Left, Pleurisy Right. The purpose of this research is to automatically analyze the 3D CT images of TB patients to detect semantic information for the type of Tuberculosis.

In this paper, we also employ a new fine-tuning neural network model which uses features coming from pre-trained CNN models as input. In addition, existing deep learning MODELS had weak classifications, therefore we propose a new fully connected 2 layers. The new contributions of this paper is to propose a novel feature building techniques, which incorporates features from two CNN models to predict Tuberculosis from images, unlike most recent research only concerned with adopting single CNN features. In the following, we first describe the tasks which were conducted in Section 2 followed by dataset of ImageCLEF2020, In Section 3, we introduce masking the dataset, experimental settings, and feature used in this research . In Section 4, we describe experiments we have carried out. In Section 5 we conclude this paper .

2 Dataset of ImageCLEF 2020

The tuberculosis task of ImageCLEF 2020 Challenge included part of chest in format of 3D CT images [6][5]. A dataset contains the chest CT scan imaging data which included 283 images in the Training (also referred as Development) dataset and 120 in the Test dataset. Since the labels are provided on lung-wise scale rather than CT-wise scale, the total number of cases is virtually increased twice.

This task participants have to generate automatic lung-wise reports based on CT image data. Each report should include the probability scores (ranging from 0 to 1) for each of the three labels and for each of the lungs (resulting in 6 entries per CT). The resulting list of entries includes: **LeftLungAffected, RightLungAffected, CavernsLeft, CavernsRight, PleurisyLeft, PleurisyRight**. Table 1 shows labels for the chest CT scan in the Training dataset.

Table 1. Presence of labels for the chest CT scan in the Training dataset.

Label	In Training set
LeftLungAffected	211
RightLungAffected	233
CavernsLeft	66
CavernsRight	79
PleurisyLeft	7
PleurisyRight	14

3 Proposed Method

We propose a multi-label analysis system to predict Tuberculosis from CT scan images. The first step is the input data pre-processing. After pre-processing input data, we will describe our deep neural network model that enables the multi-label outputs, given CT scan images. In addition, we add an optional step to the first step. We use a CT scan movie not CT scan images. We will detail our proposed system in the following section.

3.1 Input Data Pre-processing

First, we remind the reader that in train and test data, 3D CT scans are provided in compressed Nifti format. We decompressed the files and extracted the slices of x-axis, y-axis, and z-axis from the three dimensions of the 3D image shown in Fig. 1. For each dimension for each Nifti image, we obtained a number of slices ranging according to the dimension: 512 images for the X and Y dimensions, and from 110 to 250 images for the Z dimension.

After extracting slices along x-axis, y-axis, and z-axis, we propose to filter the slices of each patient using mask data [7][8]. We extract a filtering CT scan image, as shown in Fig. 2. Indeed, we can notice that many slices contain relevant information including bone, space, fat, and skin except for the lungs that could help to classify the samples. This is why we added a step to the filter and selected a number of slices per patient.

3.2 Proposed deep neural network model

To solve our multi-label problem, we propose new combined neural network models which allow inputs coming from End-to-end (CNN) features.

Training and Validation sets The training dataset consists of 108 891, 77 468, 31 497 images extracted from the filtered CT image for x, y and z axis respectively.

We have divided the train data into training and validation data with 8:2 ratio in random. CNN features were extracted using pre-trained CNN-based neural networks, including VGG16, ResNet50, NasNet-Large and EfficientNet B07. In order to deal with the above feature, we propose a deep neural network architecture where we allow multiple inputs and a multi-hot vector output.

Our system incorporates CNN features, which can be extracted using deep convolutional neural networks pre-trained on the ImageNet [9] such as VGG16 [10], ResNet50[11], NasNet-Large [12] and EfficientNet B07[13]. Because of the lack of dataset in visual sentiment analysis, we adopt transfer learning for the feature extraction to prevent over fitting. We decreased the dimensions of fully-connected layers used in CNN models. In addition, we reduced the vector to 2048 dimensions. This was introduced with the expectation of reducing the number of parameters and unifying the dimensions.

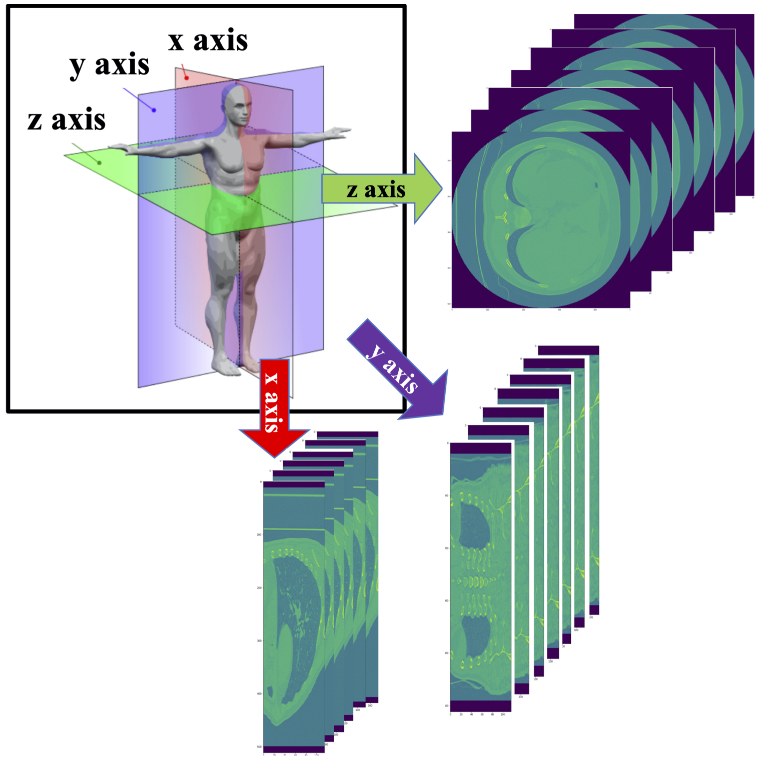

Fig. 1. Extraction by x-axis, y-axis, and z-axis slices.

Training and Validation sets and Test data We employ from the top AUC for four fine-tuning the CNN model from above. As illustrated in Fig. 3, CNN feature is combined and represented by an integrated feature as a linearly weighted average, where weights are w_3 for CNN features, respectively. CNN feature is passed out on "Fusion" processing to generate the integrated features, followed by "softmax" activation function.

3.3 Probability of multi-label

We propose a method illustrated in Algorithm 1. The input is a collection of features extracted from each image with K kinds of sentiments, while the output is a K-dimensional multi-hot vector.

In Algorithm 1, we assume that the extracted CNN feature is represented by their probabilities. For each Tuberculosis, we sum up the features, followed by median of the result, which is denoted by T_i^k in Algorithm 1. In short, the vector S_i represents the output multi-hot vector. We repeat this computation until all the test (unknown) images are processed.

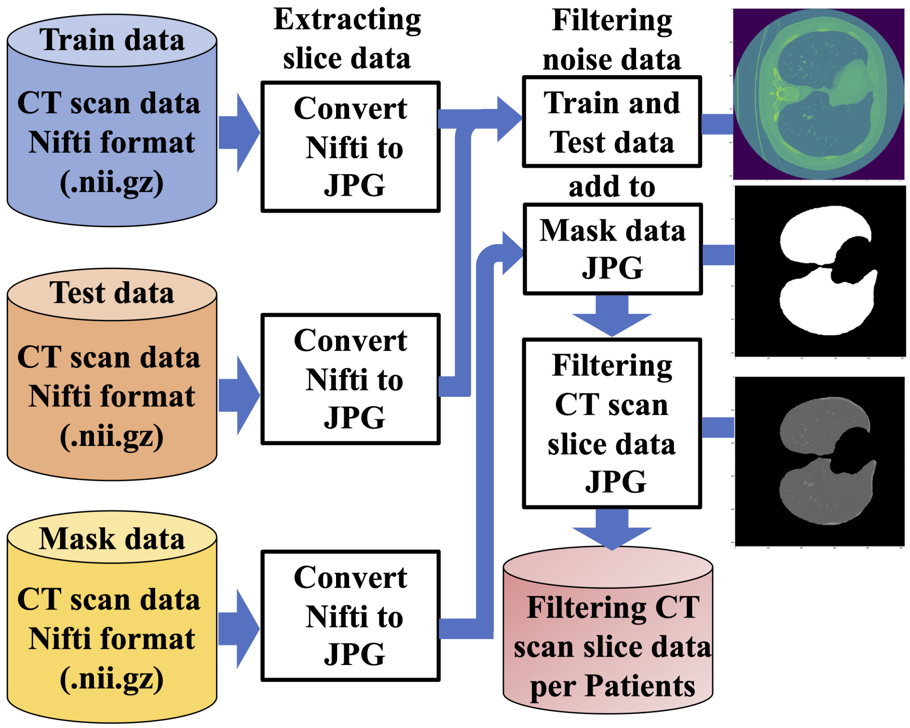

Fig. 2. Pre-processing of input data using mask data.

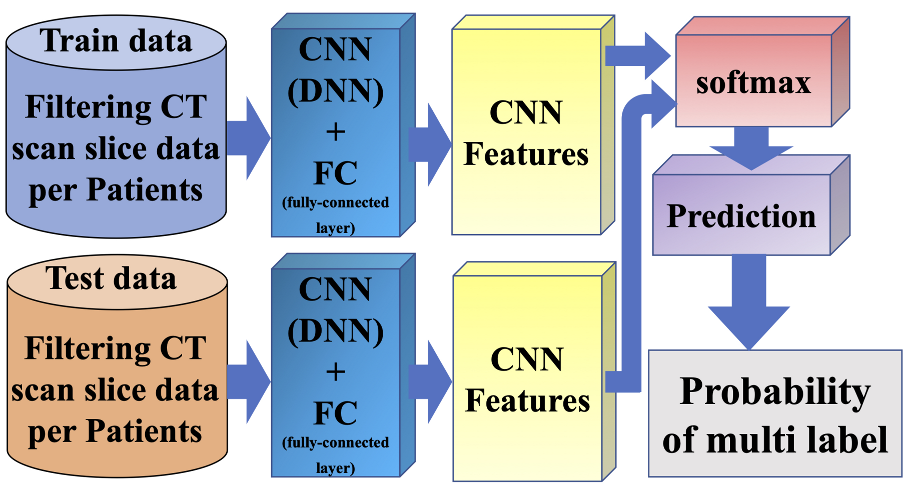

Fig. 3. Our proposed feature for multi-label feature extraction.

4 Experiments

4.1 AUC of training and validation sets

The train dataset consists in filtering CT image on x-axis, y-axis, or z-axis. The train dataset consists of 108 891, 77 468, 31 497 images extracted from the filtered CT image for x, y and z axis respectively.

Algorithm 1 Predicting multi hot vector for an image

Input: Image data i including K kinds of disease for Lungs
Output: Multi hot vector S_i
1: **for** k **do** range (K):
2: $Prob_{i,k}$=$FeatureExtraction_{i,k}$
3: T_i^k=median($Prob_{i,k}$)
4: **end for**

Here, we have divided the filtering data into training and validation data with 8:2 ratio. We determined the following hyper-parameters; batch size as 256, optimization function as "SGD" with a learning rate of 0.001 and momentum 0.9, and the number of epochs 200. For the implementation, we employ TensorFlow[14] as our deep learning framework. For the evaluation of multi-label classification, we employ mean Area Under Curve (AUC). Table 2 shows the results. Here we compare in terms of AUC for multiple axes. For fine-tuning EfficientNet B07 in x, y, and z-axis, it turns out that our proposed CNN model has the best AUC. Finally, we employ EfficientNet B07 for training and validation sets and test data. The result shows as below (4.2).

Table 2. Validation accuracy of four models (VGG16, ResNet50, NasNet-Large and EfficientNet B07) on multi-axis projections

axis	Model	Dimension	AUC
x-axis	VGG16	2048	0.901
	ResNet50	2048	0.907
	NasNet-Large	2048	0.905
	EfficientNet B07	2048	**0.908**
y-axis	VGG16	2048	0.916
	ResNet50	2048	0.917
	NasNet-Large	2048	0.915
	EfficientNet B07	2048	**0.918**
z-axis	VGG16	2048	0.976
	ResNet50	2048	0.957
	NasNet-Large	2048	0.955
	EfficientNet B07	2048	**0.978**

4.2 The result for training and validation sets and test data using our proposed model

The test dataset consists of 46 605, 32 901, 13 938 images extracted from the filtered CT image for x, y and z axis respectively.

We expect that our proposed models could give better results after a more advanced data preprocessing including the use of filtering image, and data augmentation for multi-axis. Here, we described above, we employ fine-tuning CNN models in EfficientNet B07 based on multi axis. Table 3 shows the results. "x-axis and y-axis" mean the probabilities of x-axis and y-axis. "y-axis and z-axis" mean the probabilities of y-axis and z-axis. "x-axis, y-axis, and z-axis" mean the probabilities of x-axis, y-axis, and z-axis.

Here we compare in terms of AUC. For z-axis on fine-tuning EfficientNet B07, it turns out that our proposed CNN model has the good mean AUC and minimum AUC.

Table 3. The results of doing experiment for multi-label classification and AUC for training and validation sets

Model	axis	meanAUC	minAUC
EfficientNet B07	x-axis	0.633	0.573
	y-axis	0.692	0.635
	z-axis	**0.753**	**0.698**
	x-axis and y-axis	0.642	0.596
	y-axis and z-axis	0.735	0.664
	x-axis, y-axis, and z-axis	0.654	0.615

In addition, results of the participants' submissions with the highest AUC values are shown in Table 4 [4]. Here we compare in terms of mean AUC and minimum AUC. For KDE-lab team, it turns out that our proposed CNN model has the best mean AUC and minimum AUC. The results achieved by our submissions are well ranked compared to those of the top of the list, we can notice that several runs belong to the same teams that had good results, and they probably do not differ too much. Our rank is 5th.

Table 4. The best participants' runs submitted for the CTR subtask

Group Name	Rank	meanAUC	minAUC
SenticLab.UAIC	1	0.923	0.885
agentili	2	0.875	0.811
chejiao	3	0.791	0.682
CompElecEngCU	4	0.767	0.733
KDE-lab	5	**0.753**	**0.698**
Waqas-sheikh	6	0.705	0.644
uaic	7	0.659	0.562
JBTTM	8	0.601	0.432
sztaki-dsd	9	0.595	0.546

5 Conclusions

In this research, we proposed a model for predicting each of the three labels and for each of the lungs as a multi-label problem from chest CT images. We performed Chest CT Image analysis where we proposed a combined deep neural network model which enabled inputs to come from CNN features. In multi-label Chest CT Image analysis, we also introduced a threshold-based multi-label prediction algorithm. Specifically, after training our deep neural network, we could predict the existence of a disease for given unknown CT scan images. Experimental results demonstrate that all our proposed models outperform the individual pre-trained CNN model in terms of mean AUC and minimum AUC.

In this research, we proposed a model for Tuberculosis CT Image analysis which accurately estimates multi-label problems from given images. The multi-label problems are evoking multiple different types of Tuberculosis findings simultaneously.

In the future, given an arbitrary CT or X-ray image might be included the optimal weights for the neural networks. Moreover, we hope our proposed model can encourage further research on the early detection of several viruses or unknown diseases. We also expect that our proposed model will be widely used in the field of medical computing.

Acknowledgment

A part of this research was carried out with the support of the Grant-in-Aid for Scientific Research (B) (issue number 17H01746), and Grant for Education and Research in Toyohashi University of Technology.

References

1. Yashin Dicente Cid, Alexander Kalinovsky, Vitali Liauchuk, Vassili Kovalev, , and Henning Müller. Overview of ImageCLEFtuberculosis 2017 - predicting tuberculosis type and drug resistances. In *CLEF2017 Working Notes*, CEUR Workshop Proceedings, Dublin, Ireland, September 11-14 2017. CEUR-WS.org <http://ceur-ws.org>.
2. Bogdan Ionescu, Henning Müller, Mauricio Villegas, Alba García Seco de Herrera, Carsten Eickhoff, Vincent Andrearczyk, Yashin Dicente Cid, Vitali Liauchuk, Vassili Kovalev, Sadid A. Hasan, Yuan Ling, Oladimeji Farri, Joey Liu, Matthew Lungren, Duc-Tien Dang-Nguyen, Luca Piras, Michael Riegler, Liting Zhou, Mathias Lux, and Cathal Gurrin. Overview of ImageCLEF 2018: Challenges, datasets and evaluation. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, Proceedings of the Ninth International Conference of the CLEF Association (CLEF 2018), Avignon, France, September 10-14 2018. LNCS Lecture Notes in Computer Science, Springer.
3. Bogdan Ionescu, Henning Müller, Renaud Péteri, Yashin Dicente Cid, Vitali Liauchuk, Vassili Kovalev, Dzmitri Klimuk, Aleh Tarasau, Asma Ben Abacha, Sadid A. Hasan, Vivek Datla, Joey Liu, Dina Demner-Fushman, Duc-Tien Dang-Nguyen, Luca Piras, Michael Riegler, Minh-Triet Tran, Mathias Lux, Cathal Gurrin, Obioma Pelka, Christoph M. Friedrich, Alba García Seco de Herrera, Narciso

Garcia, Ergina Kavallieratou, Carlos Roberto del Blanco, Carlos Cuevas Rodríguez, Nikos Vasillopoulos, Konstantinos Karampidis, Jon Chamberlain, Adrian Clark, and Antonio Campello. ImageCLEF 2019: Multimedia Retrieval in Medicine, Lifelogging, Security and Nature: Multimedia Retrieval in Medicine, Lifelogging, Security and Nature. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, volume 2380 of *Proceedings of the 10th International Conference of the CLEF Association (CLEF 2019)*, Lugano, Switzerland, September 9-12 2019. LNCS Lecture Notes in Computer Science, Springer.
4. Obioma Pelka, Christoph M Friedrich, Alba García Seco de Herrera, and Henning Müller. Medical image understanding: Overview of the ImageCLEFmed 2020 concept prediction task. In *CLEF2020 Working Notes*, Workshop Proceedings, Thessaloniki, Greece, September 22-25 2020. CEUR-WS.org.
5. Bogdan Ionescu, Henning Müller, Renaud Péteri, Asma Ben Abacha, Vivek Datla, Sadid A. Hasan, Dina Demner-Fushman, Serge Kozlovski, Vitali Liauchuk, Yashin Dicente Cid, Vassili Kovalev, Obioma Pelka, Christoph M. Friedrich, Alba García Seco de Herrera, Van-Tu Ninh, Tu-Khiem Le, Liting Zhou, Luca Piras, Michael Riegler, Pål Halvorsen, Minh-Triet Tran, Mathias Lux, Cathal Gurrin, Duc-Tien Dang-Nguyen, Jon Chamberlain, Adrian Clark, Antonio Campello, Dimitri Fichou, Raul Berari, Paul Brie, Mihai Dogariu, Liviu Daniel Ştefan, and Mihai Gabriel Constantin. Overview of the ImageCLEF 2020: Multimedia Retrieval in Medical, Lifelogging, Nature, and Internet Applications. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, volume 12260 of *Proceedings of the 11th International Conference of the CLEF Association (CLEF 2020)*, Thessaloniki, Greece, September 22-25 2020. LNCS Lecture Notes in Computer Science, Springer.
6. Serge Kozlovski, Vitali Liauchuk, Yashin Dicente Cid, Aleh Tarasau, Vassili Kovalev, and Henning Müller. Overview of ImageCLEFtuberculosis 2020 - automatic CT-based report generation. In *CLEF2020 Working Notes*, CEUR Workshop Proceedings, Thessaloniki, Greece, September 22-25 2020. CEUR-WS.org <http://ceur-ws.org>.
7. Yashin Dicente Cid, Oscar Alfonso Jiménez del Toro, Adrien Depeursinge, and Henning Müller. Efficient and fully automatic segmentation of the lungs in ct volumes. In Orcun Goksel, Oscar Alfonso Jiménez del Toro, Antonio Foncubierta-Rodríguez, and Henning Müller, editors, *Proceedings of the VISCERAL Anatomy Grand Challenge at the 2015 IEEE ISBI*, CEUR Workshop Proceedings, pages 31–35. CEUR-WS, May 2015.
8. Vitali Liauchuk and Vassili Kovalev. Imageclef 2017: Supervoxels and cooccurrence for tuberculosis CT image classification. In Linda Cappellato, Nicola Ferro, Lorraine Goeuriot, and Thomas Mandl, editors, *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum, Dublin, Ireland, September 11-14, 2017*, volume 1866 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2017.
9. Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.
10. Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2014.
11. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2016.

12. Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. Learning transferable architectures for scalable image recognition. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun 2018.
13. Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *ICML 2019*, 05 2019.
14. Google. Tensorflow. https://github.com/tensorflow.