AN ENHANCED DEEP LEARNING ARCHITECTURE FOR CLASSIFICATION OF TUBERCULOSIS TYPES FROM CT LUNG IMAGES

Xiaohong Gao[1], Richard Comley[1], Maleika Heenaye-Mamode Khan [2]

[1]Department of Computer Science, Faculty of Science and Technology, Middlesex University, London NW4 4BT, UK. {x.gao@mdx.ac.uk , r.comley@mdx.ac.uk }
[2]Department of Software and Information Systems, Faculty of Information Communication and Digital Technologies, University of Mauritius, Mauritius. {m.mamodekhan@uom.ac.mu}

ABSTRACT

In this work, an enhanced ResNet deep learning network, depth-ResNet, has been developed to classify the five types of Tuberculosis (TB) lung CT images. Depth-ResNet takes 3D CT images as a whole and processes the volumatic blocks along depth directions. It builds on the ResNet-50 model to obtain 2D features on each frame and injects depth information at each process block. As a result, the averaged accuracy for classification is 71.60% for depth-ResNet and 68.59% for ResNet. The datasets are collected from the ImageCLEF 2018 competition with 1008 training data in total, where the top reported accuracy was 42.27%.

Index Terms — deep learning, Tuberculosis classification, CT lung images, 3D image analysis

1. INTRODUCTION

Tuberculosis (TB) is a bacterial infectious disease caused by Mycobacterium (M.) Tuberculosis is contracted through inhaling tiny droplets from the coughs or sneezes of an infected person and remains one of the top 10 causes of death worldwide. In 2015, 10.4 million people fell ill with TB, among them 1.8 million died of the disease [1], including 0.4 million HIV patients. While most of the TB cases occur in developing countries, this Victorian disease has not been eradicated in the developed countries. On the contrary, the rate of the disease has risen recently in some parts of western countries, for example, in London UK [2], as a result of a number of reasons, including drug abuse and sleeping rough.

Although TB remains a serious contagious condition, it can be cured if treated in a timely manner with the right antibiotics. Hence knowing the types of TB plays an important first step. To assist clinicians to analyze, diagnose and deliver optimal treatment for TB patients, high resolution Computed Tomography (CT) imaging is one of the tools. This paper focuses on the implementation of a state of the art deep learning technique to classify five TB types based on CT pulmonary images, namely Infiltrative (Type 1), Focal (Type 2), Tuberculoma (Type 3), Miliary (Type 4) and Fibro-cavernous or cavity (Type 5). Figure 1 illustrates the five common types of TB in montage form of 3D lung CT images where red circles exemplify the diseased regions in each type.

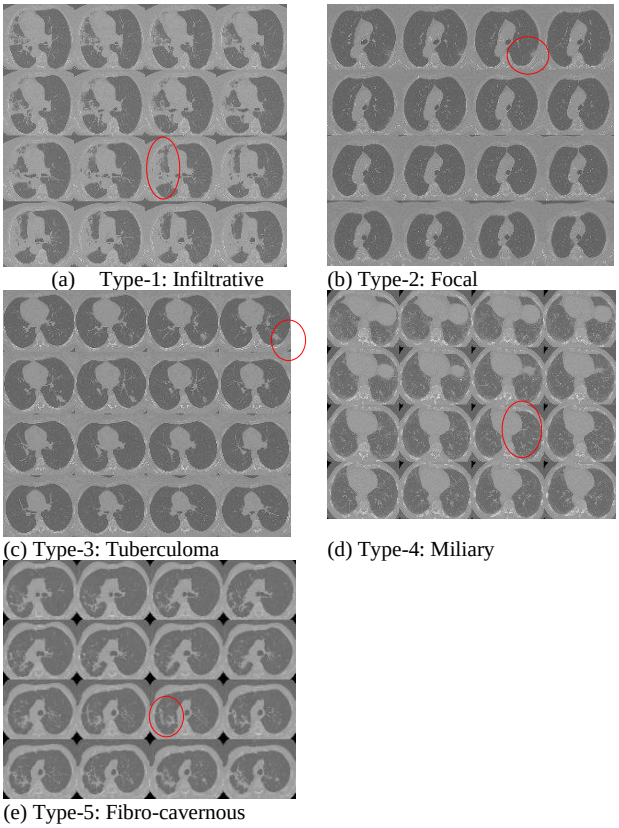

Figure 1. Five types of TB disease. From top to bottom (a) to (e): Infiltrative, Focal, Tuberculoma, Miliary, and Fibro-cavernous, where red circles refer to the diseased regions.

From Figure 1, it can be seen that for Types 4 (Miliary) and 5 (Cavity), the visual features are apparent with wide-spread dissemination of small spots (i.e. Mycobacterium tuberculosis**)** and dominating holes (cavities) respectively. However, for Types 1 to 3, visual features are not easily

distinguishable, in particular for Types 2 and 3 where diseased nodules only occur in a few slices of a lung volume. Therefore, state of the art deep learning based techniques will be applied to distinguish these subtle differences.

Deep learning models refer to a class of computing machines that can learn a hierarchy of features by building high-level attributes from low-level ones [3], thereby automating the process of feature construction. One of these models is the well-known convolutional neural network (CNN). While these models present potential for image analysis, they require large amounts of training data, which is difficult to meet in the medical domain. Hence data augmentations usually take place to enlarge datasets. For example, in [4], images are segmented into small patches for analysis of multi-drug resistance. That study however takes 3D CT volumes as a stack of 2D slices, which lacks consideration of depth information. Hence in this study, an enhanced deep learning architecture is developed for classification of TB types. In doing so, transfer learning is applied built on a residual network, ResNet [5], while taking into account of depth (3rd dimension) information of 3D CT images.

2. DATASETS

Data are collected from the competition organised by ImageCLEF2018 on Tuberculosis classification task (task#2) [6, 7] with 1008 training datasets from 677 subjects. Since the test datasets have unknown ground truth, these training data are studied in this paper. As listed in Table 1, these data are divided into training (~60%) and testing sets (~40%).

Table 1. The data numbers that are applied for training and testing.

Type	**1**	**2**	**3**	**4**	**5**	**Total**
train	200	170	100	70	70	610
Test	176	103	54	36	29	398
Total	**376**	**273**	**154**	**106**	**99**	**1008**

As demonstrated in Figure 2, each 3D dataset firstly undergoes a pre-processing stage to remove surrounding boundaries based on the provided lung masks. As a result, each volume has a dimension of 256 × 256 × depth z, with depth varying between 20 to 250 slices.

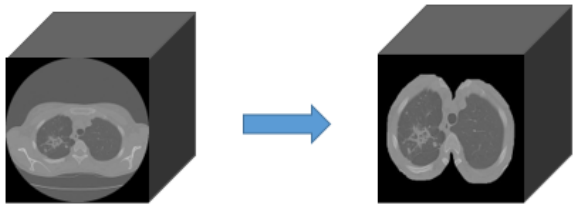

Figure 2. Data pre-processing to remove background information based on lung masks.

3. ENHANCED DEEP RESIDUAL LEARNING – DEPTH-RESNET

While a convolutional neural network (CNN) architecture can be constructed by stacking multiple layers of convolution and subsampling in an alternating fashion, to go deeper, the increased depth tends to make little contribution to the accuracy of a trained model. This is due to the well-known vanishing gradient obstacle, i.e. as the gradient is back-propagated to earlier layers, repeated multiplication may cause the gradient to become infinitely small. As a result, as the network becomes deeper, its performance gets saturated or even starts degrading rapidly.

Recently, deep residual networks (ResNet) [5, 8] introduce the notion of '*identity shortcut connection*' that bypasses one or more layers. A key advantage of residual units is that their skip connections allow direct signal propagation from the first to the last layer of the network, especially during the back propagation. This is due to the fact that gradients are propagated directly from the loss layer to any previous layer while skipping intermediate weight layers that have the potential to trigger vanishing or deterioration of the gradient signals.

In this work, inspired by [9, 10], an enhanced ResNet, depth-ResNet is introduced to be applied for classification of the five types of tuberculosis from CT lung images, which is built on the ResNet-50 model and illustrated in Figure 3.

To take advantage of ResNet-50 using 3x3 filters to perform spatial convolution, the depth convolution also adopts 3 pixels, i.e. 1x1x3 between the current, front and back frames. Considering some TB types (e.g., Type 3 in Figure 1 (c)) have lesions only covering a few slices along the depth, a 4-frame block is chosen in this study. As illustrated in Figure 3, to minimise the classification errors, a global pooling layer followed by a 5-way fully connected layer, optimised using a Softmax approach is conducted.

In Figure 3, for each residual unit, the input feature map $x_l \in \mathbb{R}^{H\times W\times D\times C}$, where H, W, D are the spatial dimensions along the height, width, and depth directions for a 2D dataset and C the feature dimension. Such maps can be thought of as stacking 2D spatial maps of C dimensional features along the depth (z) dimension. Built upon the inception concept, the depth (z) convolution block operates on the dimensionality reduced input, $x_{l,z}$ with a bank of 3D filters, $W_{l,z}$. Biases $b \in \mathbb{R}^C$ are also applied with initial values of 0 as formulated in Eq. (1).

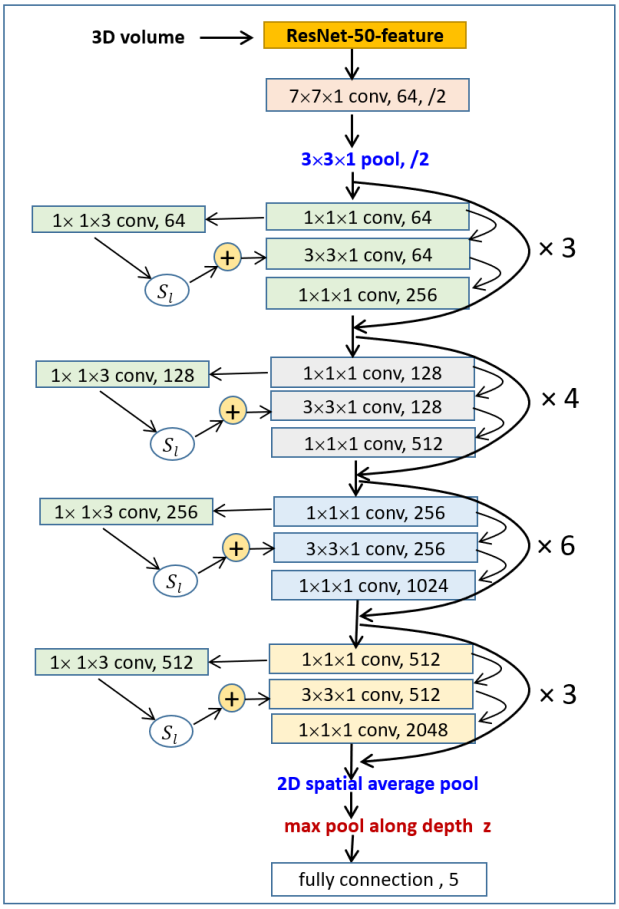

Figure 3. The depth-ResNet architecture applied in this paper, where ×N at each conv level refers to the block (e.g. conv5_x) repeats N (e.g. 3) times consecutively.

$$x_{l,z} = W_{l,z}x_{l,1} + b \qquad (1)$$

where $W_{l,z}|1 \le z \le 3$ are the convolutional filter kernels arranged as a matrix.

Hence the residual unit $\mathcal{F}$ is expressed in Eq. (2).

$$\mathcal{F} = f\left(W_{l,3}\left(S_l f\left(x_{l,z}\right) + f\left(W_{l,2} f\left(W_{l,1}x_{l,1}\right)\right)\right)\right) \qquad (2)$$

where S_l is affine scaling along the depth direction with a bias between 0 and 0.01. This scaling is adaptive to facilitate generalisation of performance and will be learnt during the training of the network. The convolution at each convolution layer along depth (z) direction ($x_{l,z}$) takes place between 4 neighbouring slices or feature maps, i.e. front, current, and back, with a randomly chosen stride (between 1 and 7 in this study). This feature is then added to the block with a scaling factor as a component of the residual unit. The pooling involves two stages. The *avg-pool* occurs for 2D spatial global average pooling whereas *max-pool* is conducted along the z direction performing global max pooling upon those feature maps.

Figure 4 elaborates the depth receptive field of a single neuron. In Figure 4, the convolution at each convolution layer along the depth (z) direction ($x_{l,z}$) takes place between 3 neighbouring slices or feature maps, i.e. front, current, and back, with randomly chosen stride (between 1 and 7 in this study). This feature is then added to the block with a scaling factor as a component of the residual unit.

The system is implemented in Matlab with the MatConvNet [11] toolbox by following standard ConvNet training procedures [12]. Upon training, 4 slices are chosen from each volume with randomly selected stride between 1 and 7 from 5 categories with a batch of 128 (=32 blocks). At testing time, each dataset undertakes the same pre-processing procedure to generate 256×256×depth volume. Then the trained depth-ResNet model (Figure 3) selects 4 slices at equal depth space and propagates these slices through the trained model to produce a single prediction for this volume with type labeled between 1 and 5.

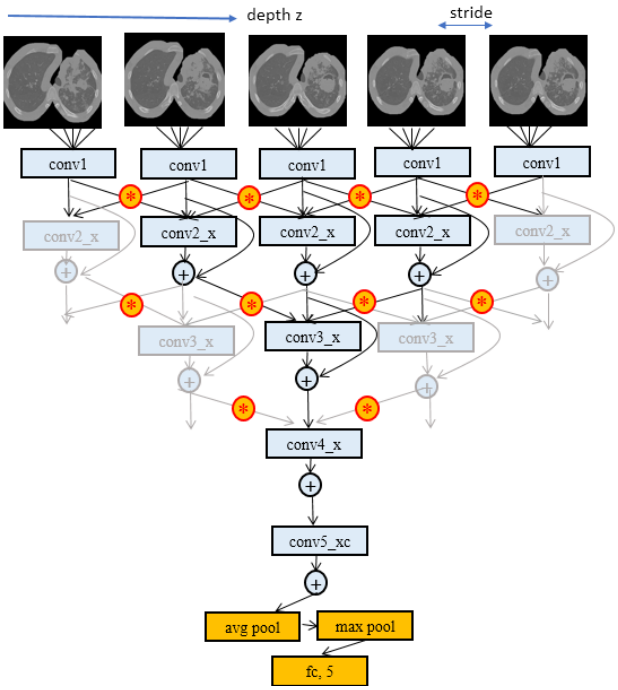

Figure 4. A block in the depth-ResNet that is applied in the paper.

4. RESULTS

Table 2 gives the processing results for depth-ResNet whereas Table 3 lists results obtained from the original model of ResNet by which the network processes each slice individually rather than as a 4-slice block. On average, depth-ResNet performs better than ResNet with a 71.60% accuracy

rate as opposed to 68.59%. While larger datasets do improve the classification rate with 81.81% for Type 1 (200 training data) and 80.58% for Type 2 (N=170), the accuracy also depends on the characteristics of TB types. For example, Type 5 has the smallest number of training data sets (N=70), the accuracy is better than that for Type 3 where N=100.

Table 2. Confusion matrix for 4-slice results.

Type	1	2	3	4	5	Avg
1	**144**		5	3	5	
2	25	**83**	31	12	1	
3		5	**15**			
4	5	1		**21**	1	
5	2	14	3		**22**	
Accuracy (%)	81.81	80.58	27.77	58.33	66.66	**71.60**

Table 3. Confusion matrix for ResNet results.

Type	1	2	3	4	5	Avg
1	**135**	12	7	6	4	
2	25	**85**	32	10		
3	3	4	**11**			
4	11	2	1	**19**	2	
5	2		3	1	**23**	
Accuracy (%)	76.70	82.52	20.37	52.77	84.61	**68.59**

Table 4. Sensitivity and Specificity for the approaches that applied where 'T' refers to 'Type', 'Se' refers to 'Sensitivity' and 'Sp' to 'Specificity'.

Method	T	1	2	3	4	5	Avg (%)
Resnet	Se	76.76	82.52	20.37	52.77	79.31	62.33
	Sp	88.44	81.49	98.00	95.76	98.40	92.41
Depth-Resnet (8-slice)	Se	81.81	74.75	27.77	50.0	75.86	62.03
	Sp	84.41	85.01	98.00	96.79	97.36	92.31
Depth-Resnet (4-slice)	Se	81.81	80.58	27.77	58.33	75.86	**64.87**
	Sp	94.46	81.04	98.56	98.10	95.10	**93.45**

5. CONCLUSION AND FUTURE DIRECTIONS

In comparison with the ImageCLEF Tuberculosis Competition [13] where the top accuracy rate was 42.27% (applying testing datasets), the developed depth-ResNet in this work (71.60%) appears to be a better way forward to process TB data by taking depth information into consideration. In the competition, another deep learning based approach applied an ensemble of 3D CNN [14], which achieved 35.33% accuracy of classification. Given the small size of training data, conventional texture-based graph models are also developed [15], which has achieved better results with 38.49% accuracy (2nd place).

In this study, while the overall classification appears to be reasonable, with a 71.60% accuracy rate, the result for Type 3 only sustains 27.77%. With regard to diseased regions, Type-3 Tuberculoma has the smallest number of slices whereas for Type-4 Milliary, the disease covers most of the slices along the depth. Even though Type 4 has the smallest number of volume of training data (N=70, the same as that for Type 5), the training sub-volumes with required type-information are much larger than that for Type 3. Hence, in the future, in addition to enlargement of datasets, incorporation of texture models will be attempted.

Since the testing datasets only include 8 subject samples in this type, the remaining work is to evaluate the results from real testing datasets when the ground truth is obtained. In order to obtain higher accuracy, it is recommended that medical knowledge should be embedded. Additionally, 3D segments should also be included to further enhance the characteristics that 3D datasets entail.

In this study, another interesting comparison has been made, which is to test the datasets collected from the ImageCLEF 2017 competition [16], which might come from different sources. Table 5 gives the results for the three approaches, ResNet, depth-ResNet-8-slice and depth-ResNet-4-slice.

Table 5. Test results from applying datasets from the ImageCLEF 2017 TB competition based on the models that are trained using the 2018 datasets.

Type	1	2	3	4	5	Avg (%)
ResNet	85.71	53.33	10.00	55.00	65.00	55.4
Depth-ResNet-8	89.29	39.07	15.00	52.50	70.00	54.0
Depth-ResNet-4	90.71	49.16	19.00	51.25	63.55	**56.8**

In the 2017 TB competition, the best result was achieved using the ResNet approach with averaged accuracy rate of 40.33% [17]. While the results in Table 5 are not as good as those in Table 4, the differences might attribute to the discrepancies between different imaging scanners from where those data are acquired. Again, the biggest challenge is to detect Type 3 Tuberculoma.

In the future, clinicians' knowledge will be incorporated to improve the classification accuracy, especially for Type 3 Tuberculoma.

6. REFERENCES

[1] WHO. Tuberculosis, Fact Sheet, March 2017.http://www.who.int/mediacentre/factsheets/fs104/en/. Retrieved in June 2017.

[2] BBC. Parts of London have higher TB rates than Iraq or Rwanda. http://www.bbc.co.uk/news/uk-england-london-34637968. Retrieved in June 2017.

[3] Y. LeCun, Y. Bengio, G. Hinton, Deep Learning, Nature. 521: 436-444, 2015.

[4] X. Gao, Y. Quan , Prediction of multi-drug resistant TB from CT pulmonary Images based on deep learning techniques, *Molecular Pharmaceutics*, 15(10) : 4326-4335, 2018.

[5] K. He, X. Zhang, S. Ren, J. Sun , Deep Residual Learning for Image Recognition, IEEE *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[6] L. Cappellato, N. Ferro., J. Nie, L. Soulier Eds., CLEF 2018 Working Notes, Working Notes of CLEF 2018 – Conference and Labs of the Evaluation Forum, CEUR-WS, Eds., 2018.

[7] ImageCLEFtuberculosis, http://www.imageclef.org/2018/tuberculosis. Retrieved in May, 2018.

[8] K. He, X. Zhang, S. Ren, J. Sun, Identity Mappings in Deep Residual Networks, *European Conference on Computer Vision (ECCV)* (2016).

[9] Feichtenhofer C, Pinz A, Wildes R, Temporal Residual Networks for Dynamic Scene Recognition, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[10] Gao XW, James-Reynolds C, Currie E, Analysis of Tuberculosis Severity Levels From CT Pulmonary Images Based on Enhanced Residual Deep Learning Architecture, NeuroComputing, 392:233-244 (2020).

[11] MatConvNet: http://www.vlfeat.org/matconvnet/. Retrieved in May (2018).

[12] A. Krizhevsky, I. Sutskever, G. Hinton. Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems 2012.* NIPS 2012 (2012).

[13] https://www.imageclef.org/2018/tuberculosis.

[14] A. Ishay, O. Marques, ImageCLEF 2018 Tuberculosis Task: Ensemble of 3D CNNs with Multiple Inputs for Tuberculosis Type Classification, In: CLEF2018 Working Notes. CEUR Workshop Proceedings, Avignon, France, CEUR-WS.org http://ceur-ws.org (September 10-14 2018).

[15] Y. Dicente Cid, H. Müller, Texture-based Graph Model of the Lungs for Drug Resistance Detection, Tuberculosis Type Classification, and Severity Scoring: Participation in ImageCLEF 2018 Tuberculosis Task, ceur-ws.org, In: CLEF2018 Working Notes. CEUR Workshop Proceedings, Avignon, France, CEUR-WS.org, http://ceur-ws.org (September 10-14 2018).

[16] Y. Dicente Cid, A. Kalinovsky, V. Liauchuk, V. Kovalev V., H. Müller, Overview of ImageCLEFtuberculosis 2017, Predicting Tuberculosis Type and Drug Resistances, in CLEF 2017 Labs Working Notes of CEUR Workshop Proceedings, CEUR-WS.org (http://ceur-ws.org), Dublin, Ireland, September 11-14, 2017.

[17] https://www.imageclef.org/2017/tuberculosis.