

Deep Learning based TB Severity Prediction

Ujjwel Balwal¹, Srinivasa Arun Yeragudipati¹, Bhuvana Jayaraman¹[0000-0002-9328-6989], and Mirnalinee Thanga Nadar Thanga Thai¹[0000-0001-6403-3520]

Department of CSE, SSN College of Engineering, Chennai, India
{ujjwel17179,srinivasaarun17166}@cse.ssn.edu.in
{bhuvanaj,mirnalineett}@ssn.edu.in

Abstract. Computer Aided Diagnosis (CAD) of diseases has undergone large developments with the application of deep learning algorithms to detect the presence of diseases. This paper presents an approach for predicting the presence of tuberculosis, caverns and pleurisy in a set of 3D CT scans of the chests of patients, which is the key task of the ImageCLEF 2020 Tuberculosis challenge. We used the masks provided by the ImageCLEF organizers to segment the 3D CT images, made 2D projections of the segmented 3D images, and augmented them in order to balance the images in the dataset. An AlexNet based model is used to predict the probability of the presence of tuberculosis, caverns and pleurisy from these 2D projections. We achieved the eighth place out of all the teams who made a submission in this task, achieving a mean Area Under the Curve (AUC) score of 0.601 and a minimum AUC score of 0.432. An analysis of the results obtained by the authors following this approach presented, exploring the role of the model's complexity in reduction of the desired performance.

Keywords: Deep Learning · Tuberculosis · Computer Tomography · Projections · AlexNet

1 Introduction

Tuberculosis (TB) is caused by bacteria (*Mycobacterium tuberculosis*) that most often affect the lungs of human beings [1], though can also affect other parts of the body. Conventionally, Lung TB diagnosis is done by analysing chest X-rays (CXR) and/or microbiological confirmation (looking for bacterium *Mycobacterium tuberculosis*, MTB) using various techniques [2]. In recent years, developments in Computer-Aided Diagnosis (CAD) have started gaining traction and has made huge contributions to the detection and diagnosis of Tuberculosis and analysis of 3D Computed Tomography (CT) images is a vital step in diagnosing TB. Techniques using heuristic knowledge extracted from the bacilli bacteria's

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CLEF 2020, 22-25 September 2020, Thessaloniki, Greece.

shape and colour have shown promising results [3] [4] [5], and proprietary technology around automated screening is also being developed by major healthcare companies [6] [7]. According to a report in 2013, around 3 million cases of TB went undiagnosed, mainly because of undertrained staff, inaccurate tests, lack of equipment [8]. Thus, the availability of Digital CXR, CTs with automated computer-aided interpretation is much needed to curb the potentially lethal disease, particularly in low resource high-burden settings.

This approach of analyzing 3D CT images of lungs for Tuberculosis consists of projecting the 3D image into 2D on the three planes - XY, YZ and XZ respectively. Following this, an AlexNet based model is used to predict the probability of a particular lung affected by TB, the probability of the presence of Caverns, and the probability of the presence of Pleurisy, respectively.

2 Task and Dataset

The main task for ImageCLEF 2020 tuberculosis [9] is to determine the probability of a patient suffering from tuberculosis by determining the probabilities for the following criteria: LeftLungAffected, RightLungAffected, CavernsLeft, CavernsRight, PleurisyLeft, PleurisyRight.

The working notes of the participating teams are to be published in the Proceedings of the 11th International Conference of the CLEF Association (CLEF 2020) [11]

The provided dataset, ImageCLEFmed Tuberculosis 2020, comprises of chest Computed Tomography (CT) scans of 403 TB patients. Out of these samples, 283 are designated for training and remaining 120 for testing. The CT images are digitized as a set of 2D slices and the distance between each of these 2D slices can vary between 0.5 - 5mm in any axis, depending upon the resolution of the 3D image. In our case, these slices are stored in the compressed NIfTI format.

3 Methodology

3.1 Data Preprocessing and Dataset Creation

The data preprocessing task is performed similar to the method used by Vitali Liauchuk [10] in the submission of the previous year's ImageCLEF Medical Tuberculosis task. For our submissions, we use the first mask [12] for predicting LungAffected, Caverns and the second mask [13] for detecting Pleurisy. The 3D NIfTI image is compressed into a pseudo-RGB image where the first (red) channel contains mean values, second (green) channel contains maximum values and the third (blue) channel comprises standard deviation values. For each lung, these pseudo-RGB images are generated for three planes - XY, YZ and XZ. After the projection process, we cut the images in half, separating the two

lungs. Thus, each of the 3D image is mapped to six 2D projections: XY Left, XY Right, YZ Left, YZ Right, XZ Left, XZ Right.

Using random selection on generated images, we take a ratio of 3:1 for training and validation sets, respectively. We observed that the data is skewed, especially with pleurisy where the number of unaffected samples outweigh the affected ones. We augment the images using random flipping and rotation to balance the dataset wherever possible.

3.2 Proposed Model Architecture

The 2D projections of the lung images does the heavy-lifting of extracting the important features required to mark the lesion or deformity in the lung. These features can easily be detected by a simple neural network. It would thus reduce the problem to detection of these lesions in the 2D projections. After pre-processing and data augmentation, we observed that recent state-of-the-art image classification neural networks overfit quickly because of their inherent complex feature extraction. Hence we choose a relatively simple AlexNet-based solution to train for prediction.

AlexNet [14] is a CNN with stacked convolutional layers. It consists of 11×11 , 5×5 , 3×3 , convolutions, max pooling, dropout, data augmentation, ReLU activations and SGD with momentum. It attaches ReLU activations after every convolutional and fully-connected layer.

Keeping the underlying AlexNet structure, we changed the number of out channels for five ConvBlocks to 64, 128, 256, 512, 512 respectively, each one of the ConvBlock is then followed by ReLU and MaxPooling of stride 2.

The basic block of our architecture, ConvBlock, is a simple 2D convolutional layer, followed by a ReLU layer and a MaxPool2D layer. The simple 2D convolutional layer has a kernel size of 2 and stride of 2, and the MaxPool2D layer has a kernel size of 3 and stride of 1. We added Dropout, linear, ReLU and again Linear layer after the five ConvBlocks, in order to build a suitable model for Binary classification, which is then changed into probability distribution.

The architecture of the proposed model used for this task is shown in Fig 1. The input to this model is a pseudo-RGB image which consists of three channels. The output is the probability distribution of the target variable that signifies the probability of lung being affected. We use PyTorch [16] to construct and train the model. Starting from left, the number in parentheses indicates the number of output channels for each of the ConvBlocks. Following ConvBlocks is the dropout layer, which has an activation probability of 0.5. The following Linear layer is a fully connected layer, the number of nodes present in whom are represented by the bracket enclosed numbers. To introduce non linearity among the two Linear layers, we use ReLU activation function. SoftMax layer follows the final Linear

layer, which converts the output of the model into a probability distribution.

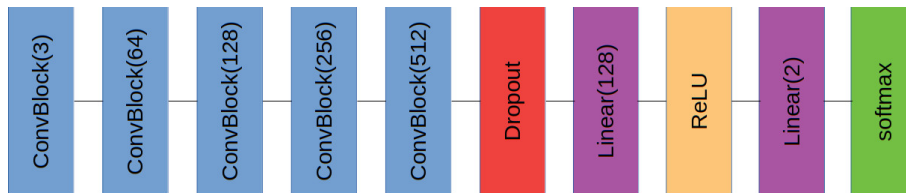


Fig. 1. Left to Right: The proposed model architecture. The number in parentheses indicates the number of output channels for each of the ConvBlock, and the number of nodes for each of the Linear Layers

3.3 Training

We trained separate models for Lung, Caverns and Pleurisy respectively. We observed that the patterns associated with the disease do not change with respect to the side of the lung, whether left or right. Therefore, we use left and right affected lung samples from the dataset collectively, to train one model. Separate models were trained for all 3 different 2D projections, thus for each subtask say, LungAffected, we train one model for XY-projection, one for YZ-projection and one for ZX-projection. We approach the training as a classification task. We used Kaiming initialization [17] to initialize the weights of the network and trained the model using a mini-batch size of 4. We used Categorical cross entropy loss and a learning rate of 10^{-5} while training. We used Adam optimizer with weight decay of 0.0005. The loss and accuracy saturated after 5 to 7 epochs for all the training sets. In the end, we have a total of 9 trained models, XY, YZ, ZX - projection model for each of Lung, Caverns and Pleurisy detection subtask.

We performed test time augmentation while predicting the probability values. The XY, YZ and ZX projections were taken and given as input to the model separately. The trained model does not differentiate in left and right lung, so we pass both projections one after the other to the same model and store the prediction values separately. The resulting outputs were taken and given as input to a SoftMax layer, which converts them into a probability distribution. This provides us with the probabilities for the predictions. While predicting the values individually for XY, YZ and ZX projections, the prediction scores for LungAffected and Caverns were almost similar, but the scores of Pleurisy values for the XY and YZ axes were low while the ZX-axis score was substantially higher. Hence, we used only the ZX-projection for Pleurisy and mean of all the projections for others.

4 Experimental Results

We did two submissions for the task. The primary difference between the two models was the number of channels and inclusion of test time augmentation, which gave a significant improvement to the scores, bringing the mean AUC to 0.601 shown in Table 1. We mentioned the approach and methodology in section 3. With our best prediction, we achieved the eighth place out of the nine teams who made a submission and the results are shown in Table 2. Our best submission, Run 2 (JBTTM) achieved a mean AUC score of 0.601 and min AUC score of 0.432, Run 1 achieved better min AUC score of 0.471 but poor Mean AUC score of 0.484. The top ranking team SenticLab.UAIC scored 0.924 and 0.885 on Mean and Min AUC respectively.

The method of using AlexNet on 2D projections suggested promising results. We detected TB lesions with a substantial certainty and performance in case of caverns is good. However, it did not perform as expected in detecting pleurisy, partly because it is a non-localized phenomenon like a TB lesion. We can also attribute the reason for a lower score to the large number of channels in the network which added unnecessary complexity to the model and caused it to overfit that resulted in poor performance of our model, especially in detecting pleurisy.

Table 1. Performance of the two runs submitted by our team. Run 2 represents best run out of the two submitted runs

Submission	Mean AUC	Min AUC
Run 1	0.484	0.471
Run 2	0.601	0.432

Table 2. The best participants' runs submitted in this task [15], for each of the 9 teams in the final leaderboards

Rank	Group Name	Mean AUC	Min AUC
1	SenticLab.UAIC	0.924	0.885
2	SDVA-UCSD	0.875	0.811
3	Chejiao	0.791	0.682
4	CompElecEngCu	0.767	0.733
5	KDE Lab	0.753	0.698
6	FAST_NU_DS	0.705	0.644
7	Uaic2020	0.659	0.562
8	JBTTM	0.601	0.432
9	Sztaki_dsd	0.595	0.546

5 Conclusion

In this work, we experimented with an AlexNet Based Model to predict the probability of a lung affected by TB, the probability of the presence of Caverns, and the probability of the presence of Pleurisy respectively. Data preprocessing and augmentations are done as described in the previous sections to prepare the images for the deep neural network. The performance of the models submitted are measured using mean AUC and minimum AUC. Our second submission run has achieved 0.602 and 0.432 for the specified measures. The performance when compared with the other submissions of this task has shown reasonable yet improvable results. Our team, JBTTM, has achieved eighth place out of the nine teams who have submitted their runs. We observed that the increased model complexity led to overfitting and thereby pulling down the model performance. A smaller and simpler model, along with proper regularization techniques, could be used in order to achieve a better result.

References

1. WHO Fact Sheet: 24/03/2020: Tuberculosis. World Health Organization, <https://www.who.int/news-room/fact-sheets/detail/tuberculosis>.
2. Suleiman, K., Lessem, E.: An Activist's Guide To Tuberculosis Drugs. Treatment Action Group (2017).
3. Forero, M., Cristobal, G., Álvarez-Borrego, J.: Automatic identification techniques of tuberculosis bacteria. In: Proceedings of SPIE - The International Society for Optical Engineering, Vol. 5203 (2003).
4. Veropoulos, K., Campbell, C., Learmonth, G., Knight, B., Simpson, J.: The automatic identification of tubercle bacilli using image processing and neural computing techniques. In: Proceedings of the eighth international conference on artificial neural networks, Vol. 2 (1998)
5. Zaidi, S.M.A., Habib, S.S., Van Ginneken, B.: Evaluation of the diagnostic accuracy of Computer-Aided Detection of tuberculosis on Chest radiography among private sector patients in Pakistan. *Sci Rep* 8, 12339 (2018). <https://doi.org/10.1038/s41598-018-30810-1>
6. Faiz Ahmad, K., Pande, T., Tessema, B., Song, R., Benedetti, A., Pai, M., Lönnroth, M., M. Denking, C.: Computer-aided reading of tuberculosis chest radiography: moving the research agenda forward to inform policy. *European Respiratory Journal* Jul 2017, 50 (1) 1700953; DOI: 10.1183/13993003.00953-2017
7. Murphy, K., Habib, S.S., Zaidi, S.M.A.: Computer aided detection of tuberculosis on chest radiographs: An evaluation of the CAD4TB v6 system. *Sci Rep* 10, 5492 (2020). <https://doi.org/10.1038/s41598-020-62148-y>
8. Stop TB Partnership - Fact Sheet: The Missing 3 Million, <http://www.stoptb.org/assets/documents/resources/factsheets/Stop%20TB%20infographic%20Missing%203%20Million.pdf>.
9. Kozlovski, S., Liauchuk, V., Dicente C., Yashin, Tarasau, A., Kovalev, V., Muller, H.: Overview of ImageCLEFtuberculosis 2020 - Automatic CT-based Report Generation and Tuberculosis Severity Assessment. CLEF working notes, CEUR (2020).
10. Liauchuk, V.: Projection-based CT Image Analysis for TB Severity Scoring and CT Report Generation. In: CLEF2017 Working Notes. CEUR Workshop Proceedings, Dublin, Ireland (2019).

11. 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, 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. Proceedings of the 11th International Conference of the CLEF Association (CLEF 2020), Thessaloniki, Greece, LNCS Lecture Notes in Computer Science, 12260, Springer (September 22-25, 2020).
12. Dicente Cid, Y., Jimenez-del-Toro, O., Depeursinge, A., Müller, H.: Efficient and fully automatic segmentation of the lungs in CT volumes, CEUR Workshop Proceedings, Vol. 1390 (2015).
13. Liauchuk, V., Kovalev, V. ImageCLEF 2017: Supervoxels and co-occurrence for tuberculosis CT image classification. In: CLEF2017 Working Notes. CEUR Workshop Proceedings, Dublin, Ireland (2017).
14. Krizhevsky, A., Sutskever, I., Hinton Geoffrey, E.: ImageNet Classification with Deep Convolutional Neural Networks. In: Advances in Neural Information Processing Systems, Vol. 25(2) (2012).
15. ImageClef - 2020: AICrowd - ImageCLEF 2020 Tuberculosis - CT report, <https://www.aicrowd.com/challenges/imageclef-2020-tuberculosis-ct-report/leaderboards>.
16. Torch Contributors: PyTorch Documentation, <https://pytorch.org/docs/stable/index.html>.
17. Kaiming, H., Zhang, X., Shaoqing, R., Jian S.: Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: CoRR, (2015)