TB Detection using CXR

Abhilash Kuhikar
Indiana University Bloomington
School of Informatics and Computing
akuhikar@iu.edu

Karen Sanchez Trejo
Indiana University Bloomington
School of Informatics and Computing
karsanc@iu.edu

Shyam Narasimhan
Indiana University Bloomington
School of Informatics and Computing
shynaras@iu.edu

Abstract

In this paper we present different techniques to classify chest x-ray (CXR) images as Tuberculosis (TB) infected or not. We analyzed CXR of 336 cases with manifestation of tuberculosis, and 326 normal cases. We did transfer learning, using VGGnet (a pretrained Convolutional Neural Network (CNN)) as feature extractor. Then we analyzed the generated heatmaps using Grad-CAM (Selvaraju et al. 2017) and found that they were not focusing on the areas where the disease is actually present. Next we built and trained our own CNN and found that the heatmap generated from this network did a slightly better job than VGG as far as the high activation areas of the CXR related to the disease were concerned. We then tried UNet which is based on the semantic segmentation. The UNet network is trained to predict the lung mask. We trained a classifier to learn from the latent features derived from convolutional layers of UNet so that it focuses more on the lung regions for classification. All the implementation was done in Pytorch.

Keywords : Tuberculosis, Convolutional Neural Network, Grad-CAM

Introduction

Tuberculosis (TB) is caused by bacteria (Mycobacterium tuberculosis) that most often affect the lungs. It is one of the leading causes of death worldwide. In 2015, more than 10 million people fell ill with TB and 1.8 million died from the disease. (World Health Organization 2018)

Chest X-rays (CXR) play a crucial role in TB diagnosis, especially in the case of pulmonary TB (PTB), which is one of the most common presentations of TB. Although CXRs do not provide ground truth for confirming TB, they still offer a high sensitivity method for detecting TB-related abnormalities in the lungs (cavities, lymphadenopathy, pleural effusion, etc.). In addition, since CXRs provide a low-cost, rapid examination even in remote settings, it has been recognized as a powerful screening test for TB, especially in areas and populations with higher disease burden. While the cost of acquiring a CXR has become much more affordable, the interpretation of CXR scans is currently limited by cost and access to trained radiologists. And many patients are diagnosed too late, being unable to treat their symptoms using conventional TB antibiotics (SemanticMD 2018). Hence the importance of creating models that can analyze a CXR. The aim of this project is to use different networks to classify the CXRs as TB infected or not.

There are two types of TB:

- Latent TB: A person can have TB bacteria but doesnt develop the disease. In this case there are not any abnormalities in the chest X-rays (CXR).
- Active TB: If the immune system cant stop the bacteria from growing then the person shows symptoms that can be seen in chest X-rays (CXR).

If a person is suspected to have TB disease, a TB skin test or TB blood test is conducted. If the test is positive, it means that the persons body is infected with TB bacteria. Then additional tests are needed to determine if the person has latent TB infection or TB disease, such as chest X-rays (CXR). (Centers for Disease Control and Prevention 2016)

Problem statement

Chest X-rays (CXR) are used to detect abnormalities. The radiological features show considerable variation, but in most cases they are characteristic enough to suggest the diagnosis. The most common features are: (Jaeger et al. 2013)

- Cavitation: Appears in 50% of the patients. Within the cavity there may be a small quantity of fluid, visualized as an airfluid level. (Fig 1)
- Lymphadenopathy: Hilar and mediastinal nodes are larger than usual. (Fig 1)
- Patchy, poorly defined segmented consolidation: in the apical and posterior segments of the upper lobes, and in the superior segment of the lower lobe. (Fig 1)
- Miliary tuberculosis: TB is spread through blood vessels and appears as multiple tiny nodules that are distributed uniformly.
- Extension to the pleura, resulting in pleural effusion: It is an accumulation of fluid in the pleural space which appears as a large white surface. (Fig 2)

These features could be subtle and not detectable for a person that does not have the expertise. Hence, the aim of this project is to build a Convolutional Neural Network (CNN) that classifies X-rays as TB positive or TB negative.

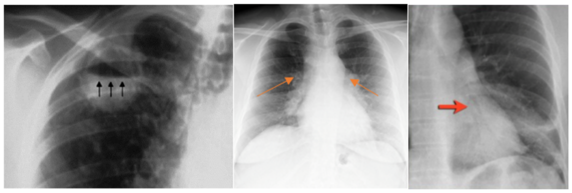

Figure 1: Cavity (Jaeger et al. 2013), lymphadenopathy in the hilar nodes (med-ed.virginia 2018), lobar consolidation (Radiology Assistant 2018)

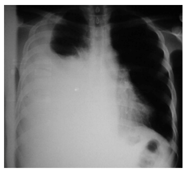

Figure 2: Large right sided pleural effusion (ScienceDirect Topics 2018)

Previous related work

The paper (Lakhani and Sundaram 2017) describes the use of deep convolutional networks for detecting the Tuberculosis using chest radiographs. Two different DCNNs, AlexNet and GoogLeNet, were used to classify the images as having manifestations of pulmonary TB or as healthy. Both untrained and pretrained networks on ImageNet were used, and augmentation with multiple preprocessing techniques. Ensembles were performed on the best-performing algorithms. They have generated heatmaps to analyse the activations on the chest Xray images. The heatmaps show that the network is focusing on parts of the image where the disease is present. The networks used here were deep networks and the data available with them after the augmentation was considerable enough given the depth of their networks.

There are several applications which are capable of detecting tuberculosis with an accuracy comparable to that of radiologists. (SemanticMD 2018)

Our approach

We had a very small amount of data available. Here we have used VGGnet and our own CNN from scratch model to train and test the data available with us. We then analyzed the heatmaps from VGGnet and our own CNN model. However we found that the heatmaps generated from VGGnet were not focusing on the areas where the disease is actually present. Hence we tried our own CNN and found that the heatmap generated from this network did a slightly better job than VGG as far as the high activation areas of the xray related to the disease was concerned. We then tried UNet which is based on the semantic segmentation. The UNet network will first predict the lung mask for a given CXR. We have used the latent features derived from the Convolution layers of this network for further classification of the CXR. This way, our classifier network built on top of latent features of UNet would focus on lung areas to detect the abnormalities related to TB.

Dataset and pre-processing

The dataset posted on Kaggle (`https://www.kaggle.com/kmader/pulmonary-chest-xray-abnormalities`) was provided by the National Library of Medicine, Maryland, USA in collaboration with Shenzhen No.3 Peoples Hospital, Guangdong Medical College, Shenzhen, China (Jaeger et al. 2014) (Candemir et al. 2014). It contains CXR of 336 cases with manifestation of tuberculosis, and 326 normal cases.

The images are in png format. Some of them only had one channel. So they had to be converted to 3 channels. This was done by repeating the values for the missing channels. Then the images were resized to 224×224 for the VGGNet and the CNN build and trained from scratch. For the UNet the images were resized to 256×256.

To train the models, 70% of the dataset was used, leaving 30% for testing.

Methods

Convolutional Neural Networks (CNN) are used for pattern recognition in images. The input of the network is an image of size $m \times m \times r$, where m is the height and width of the image and r is the number of channels.They consist of three types of layers:

- Convolutional layers: Every layer has k filters (named kernels) of size $n \times n \times q$, where n is smaller than m, giving as output k feature maps of size $m - n + 1$.
- Pooling layers: Each map is downsampled in this layer, reducing the number of features.
- Fully-connected layers: They perform the usual job of a multilayer neural network. In this case binary classification. (O'Shea and Nash 2015)

Transfer learning using VGGNet

There are CNN's with predefined architectures that have been already trained on a very large dataset. These CNN's are used as an initialization or as a fixed feature extractor. One of them is VGGNet (Fig 3). It consists of 13 convolutional layers with 3×3 filters and 3 fully connnected layers. It was trained on ImageNet dataset, which contains 1.2 million images with 1000 categories. (Simonyan and Zisserman 2014).

In this project, the last fully-connected layers of VGGNet were removed, treating the rest of the VGGNet as a fixed feature extractor. Then we trained a Neural Network with one hidden layer to get the binary classification considering the features from VGGNet as the input to the network.

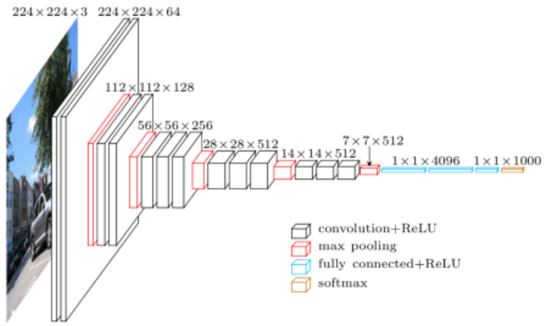

Figure 3: VGGNet architecture

CNN from scratch

Since the CXR dataset is different compared to the dataset that was used to train VGGNet, it is expected that pretrained VGGNet won't be the best feature extractor for the CXR dataset. That is why a CNN architecture was created from scratch to classify the data. The architecture was defined through experiments (Fig 4), adding layers according to the accuracy obtained. The final CNN that was trained has 3 convolutional layers (with 18 3×3 filters) and 3 maxpool layers. Since all the convolutional layers of this network were trained on the dataset, the filters in the convolutional layers are expected to learn the features specific to the CXR.

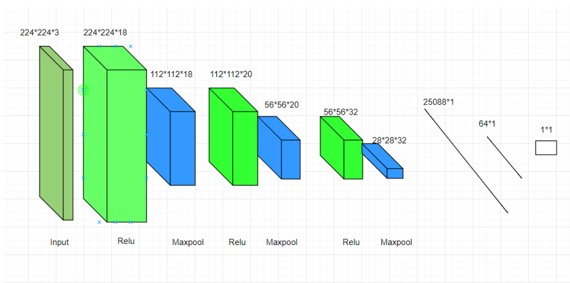

Figure 4: CNN architecture

UNet

We analyzed the heatmaps from VGGnet and our own CNN model. However we found that the VGG network was concentrating on the lower parts of the CXR which has little relevance to the affected area(lungs) in the case of TB. Hence we used UNet in the hope to extract better features for our classification task. UNet architecture consists of two parts:

- Conv: Series of convolutional and pooling layers
- Deconv: Series of deconvolutional and unpooling layers

The architecture (Fig 5) is generally used for the semantic segmentation task. We have used this to our advantage to train the entire architecture to extract the lung mask from the images. We then used the initial half (conv) of this trained model as the feature extractor for our classification problem. For classification, we have used the neural network with two hidden layers. The advantage here will be that given any CXR image, the network will only concentrate on the lung area(assuming it has learned the mask from the image) and hence will be robust to new test CXR images.

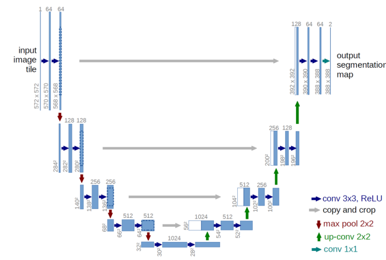

Figure 5: UNet architecture

Grad-CAM

Gradient-weighted Class Activation Mapping (Grad-CAM) is a way to visualize the regions of the image that are more important for predictions in the CNN. This method computes the gradient of the target classes with respect to the feature map of the last convolutional layer. It uses the last layer because it is the one that captures spatial information before it gets lost in the fully-connected layer. Then the gradients are backpropagated and global-average-pooled to obtain the neuron importance weights of the gradients. These weights are a partial linearization of the CNN that capture the importance of each pixel for the target class. (Selvaraju et al. 2017)

In this project we used the implementation from (Jacob Gildenblat 2018).

Training

The networks were trained on GPU - GeForce GTX 1060 6BG and i7 processor with 16 GB RAM. The VGGNet was trained with 14 epochs(giving 0 train error at 13th epoch) for half an hour. The CNN from scratch model was trained with 40 epochs(giving 0 train error at 39th epoch) for about 7 minutes. The Unet model was trained with 20 epochs for about an hour(0.2 train error at the last epoch).

The **loss function** used is Binary Cross Entropy Loss. For training Unet model for pixel wise segmentation, we have taken BCE pixel-wise. We used the next optimizers:

- VGG Stochastic Gradient Descent (SGD), with Learning Rate as 0.001 and momentum 0.9
- OwnCNN Stochastic Gradient Descent, Learning Rate 0.006, Momentum 0.9, LR Decay by 0.1 factor every epoch
- Unet Pixel - wise segmentation SGD, with LR = 0.000001 and Momentum = 0.9
- Classifier on Unet Features SGD, LR = 0.001 and Momentum = 0.9

Results

The results are expressed as accuracy (corrected classified cases divided by the total number of cases), recall (number of true positives divided by the total number of actual positive cases), precision (number of true positives divided by the total number of cases classified as positive). The ROC curve was also calculated, it shows the true positive rate (or recall) in the y axis, against the false positive rate (number of false positives divided by the total number of actual negative cases) at various threshold settings. The area under the ROC curve is a measure of how well a classifier can distinguish between two diagnostic groups, having perfect classification when it is 100%. All of the results measures were calculated in the test sample of the data.

The CNN that was trained using VGGNet as feature extractor had the next results: Accuracy = 89.39%, Recall = 86.79%, Precision = 92.90%, AUROC = 95.2%.

		Classifier		
		Negative	Positive	Total
Truth	Negative	85	7	92
	Positive	14	92	106
	Total	99	99	198

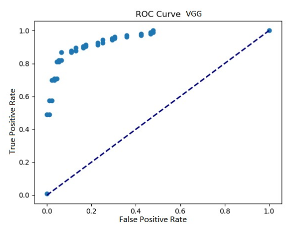

Figure 6: ROC curve for VGGNet

The CNN that was build and trained from scratch had the next results: Accuracy = 81.00%, Recall = 85.85%, Precision = 80.00%, AUROC = 90.00%.

		Classifier		
		Negative	Positive	Total
Truth	Negative	70	22	92
	Positive	15	91	106
	Total	85	113	198

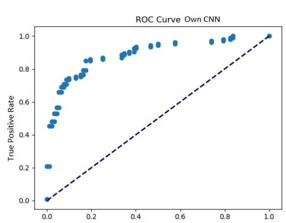

Figure 7: ROC curve for CNN from scratch

The results for both of the CNN's are similar, that is why a visualization is important to understand what the networks are doing. The heat maps (Fig 8) show regions of the image that are more important for predictions in the CNN. We can see that VGGNet is giving more importance to areas below the lungs, where there are not lung abnormalities caused by TB. On the other hand, the CNN build and trained from scratch is activating parts of the lungs that could be abnormalities. Although it is not perfect, since it is also giving importance to some parts of the arms.

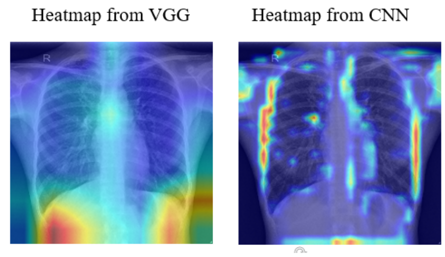

Figure 8: Heat maps (generated with Grad-CAM) of a true positive case

After analyzing the heatmaps, we decided to try UNet to see if we can get better results. With the small number of layers, the Unet detected the mask with the f1 score value as 0.45, when keeping 0.5 as probability threshold. We believe that increasing the layers in UNet would boost the detection of mask. For classifier on top of UNet features, here is what we got: Accuracy = 85.30%, Recall = 82.20%, Precision = 88.70%, AUROC = 91.01%.

		Classifier		
		Negative	Positive	Total
Truth	Negative	81	10	91
	Positive	19	88	107
	Total	100	98	198

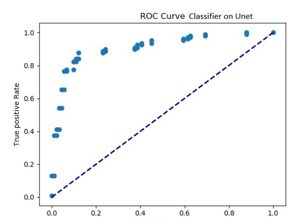

Figure 9: ROC curve for UNet

We also tried all the models on unseen Montgomery images, which had lots of CXRs tilted and shifted. The VGG Net and our Own CNN performed with accuracy of about 40%, but the classifier on top of UNet gave the accuracy of about 50%. We believe that with higher number of layers in UNet, we would be able to detect the lung masks better and achieve robust results.

Conclusions

The CNN that was trained using VGGNet as feature extractor and the CNN build and trained from scratch have similar performance. In general, it can be said that they have high accuracy, recall and precision. Nevertheless, we have to take into account that the purpose of the models is to give a diagnosis. The recall is the ability of the classifier to detect the patients that actually have TB. A classifier with recall of 85% detects 85% of the patients that have TB, but 15% go undetected. This could be fatal, a person that has TB has 15% of being classify as No TB, which can lead to the wrong treatment and the disease keep evolving. Looking at the heat maps, we can realize that VGGNet is activating parts of the image that are not related to abnormalities caused by TB. The CNN build and trained from scratch is doing a better job, since it is actually activating important regions of the lungs. However, it also gives importance to parts not related to TB like the arms. So we can still improve the model, making it concentrate only in the lung area.

Future Work

We want to try generating the heatmaps from the UNet network. We hope that these heatmaps will convey more relevant information than the other two networks used. To make the networks concentrate on the lung regions only, we can use an Attention Model. With guidance of medical expert, we can do pixel-wise annotation or bounding box annotations of lung abnormalities to boost the accuracy and make the classification more relevant in radiologists perspective. Finally, we can generate clinical readings based on the visual features (Image captioning can be a good approach).

References

[Candemir et al. 2014] Candemir, S.; Jaeger, S.; Palaniappan, K.; Musco, J. P.; Singh, R. K.; Xue, Z.; Karargyris, A.; Antani, S.; Thoma, G.; and McDonald, C. J. 2014. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE transactions on medical imaging* 33(2):577–590.

[Centers for Disease Control and Prevention 2016] Centers for Disease Control and Prevention. 2016. Diagnosing latent tb infection disease. [Online; accessed 16-November-2018].

[Jacob Gildenblat 2018] Jacob Gildenblat. 2018. pytorch-grad-cam. [Online; accessed 29-November-2018].

[Jaeger et al. 2013] Jaeger, S.; Karargyris, A.; Candemir, S.; Siegelman, J.; Folio, L.; Antani, S.; and Thoma, G. 2013. Automatic screening for tuberculosis in chest radiographs: a survey. *Quantitative imaging in medicine and surgery* 3(2):89.

[Jaeger et al. 2014] Jaeger, S.; Karargyris, A.; Candemir, S.; Folio, L.; Siegelman, J.; Callaghan, F. M.; Xue, Z.; Palaniappan, K.; Singh, R. K.; Antani, S. K.; et al. 2014. Automatic tuberculosis screening using chest radiographs. *IEEE Trans. Med. Imaging* 33(2):233–245.

[Lakhani and Sundaram 2017] Lakhani, P., and Sundaram, B. 2017. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 284(2):574–582.

[med-ed.virginia 2018] med-ed.virginia. 2018. Hilar adenopathy. [Online; accessed 2-December-2018].

[O'Shea and Nash 2015] O'Shea, K., and Nash, R. 2015. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.

[Radiology Assistant 2018] Radiology Assistant. 2018. Chest x-ray - lung disease. [Online; accessed 2-December-2018].

[ScienceDirect Topics 2018] ScienceDirect Topics. 2018. Tuberculous empyema - an overview. [Online; accessed 2-December-2018].

[Selvaraju et al. 2017] Selvaraju, R. R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D.; et al. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. *ICCV* 618–626.

[SemanticMD 2018] SemanticMD. 2018. Semanticmd ai for tb. [Online; accessed 16-November-2018].

[Simonyan and Zisserman 2014] Simonyan, K., and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[World Health Organization 2018] World Health Organization. 2018. Tuberculosis. [Online; accessed 16-November-2018].