11-785 Final report: Lung Disease detection from Chest X-ray images and Radiology reports

Daniel Marew
dmarew@andrew.cmu.edu

Nebiyou Yismaw
nyismaw@andrew.cmu.edu

Abstract

We present different deep learning models that can recognize the presence or absence of lung diseases from chest X-ray images and radiology reports. We use two publicly available datasets. One of which contains chest X-ray images and their corresponding disease labels and another one which contains x-ray images and their associated radiology reports. As a baseline, we trained a convolutional neural network(CNN) binary classifier that classifies X-ray images as healthy or unhealthy. From this model, we produce heat maps to visualize the discriminative image regions that led the model to classify the X-ray images as unhealthy, which can be of great assistance to radiologists. We will then build a text analyzer model based on long short-term memory (LSTMs) that classifies radiology reports (every X-ray image has one) as healthy/unhealthy. We will then combine the two binary classifiers to get better results. Finally, we will train an image captioning model that uses attention to generate captions that contain information about the disease types from the X-ray images.

1 Introduction

Detecting different lung diseases from chest X-rays is a challenging task that relies on the availability of expert radiologists. In todays health care environment, Radiologists are required to examine tremendous amounts of chest X-ray images Jackson (2015) which might cause misinterpretation and misdiagnosis. In this work, we present a computer-aided diagnosis (CAD) system that can automatically detect lung diseases from chest X-rays which we hope will mitigate this problem.

Our system relies on three different models. The first one is a 121-layer convolutional neural network similar to the one used in Rajpurkar et al. (2017) that inputs a chest X-ray image and outputs the probability of presence of lung disease along with a heat map that shows the discriminative image regions that led the model to classify the X-ray images as abnormal. This model will be useful in scenarios where only X-ray images are available.

In the absence of X-ray images, we can utilize the additional information we get from the radiology reports to determine the presence or absences of diseases. That is exactly what our second model is going to do. It is a many to one long short term memory (LSTM) model that outputs the probability that a radiology report is healthy or unhealthy.

When both X-rays and radiology reports are available, however, we can combine the outputs of the above two models to make more accurate predictions.

Finally, we will train a third image captioning model based on attention, similar to Xu et al. (2015), to generate captions that contain relevant information about the disease types from the X-ray images.

2 METHODOLOGY

We used three different approaches to detect abnormalities in chest X-ray images and their reports. The first approach includes an image classifier than takes in chest X-ray images and outputs probabilities of normal and abnormal labels. The second approach involves a text classifier that takes in radiology reports, analyze them and output probabilities. After the building the two models, we combined the outputs in order to get improved results. The additional features we have in our project include generating heat-maps of the chest X-ray images and captions that contain relevant disease information.

2.1 BASELINE OR INITIAL EXPERIMENT

The task of determining the presence or absence of a lung disease from X-ray images is a binary classification problem, where the input is a frontal-view chest X-ray image X and the output is a binary label $y \in \{0, 1\}$ indicating the absence or presence of lung disease (Rajpurkar et al. (2017)). Hence as a baseline technique, we trained a binary classifier on the chest X-ray images. Our based line model is a deep convolutional neural network based on Huang et al. (2016) with the final fully connected layer removed and replaced with a binary projection layer. A Pre-trained Densenet-121 Huang et al. (2016) model was fine-tuned as a basis for our binary classifier. The final fully connected layer was removed and replaced by a linear layer with binary outputs. Weights from a Pre-trained model were loaded and the model was fine-tuned end-to-end using a default Adam optimizer. Before training, some pre-processings were performed on the original image data. These pre-processings include data augmentation and transformation. The first step was to extract RGB data from the image. After doing so, the original images were cropped and re-sized from a size of 1024*1024 to a size of 224*224 and normalized. To improve efficiency, the training data was augmented by applying random transformations on it. The model was tested on an unseen data which had 200 normal and around 250 abnormal test images.

Metric	Score
Accuracy	0.670
Recall	0.665
Precision	0.621
F1	0.654

Table 1: Baseline model scores

2.2 IMPROVED IMAGE CLASSIFICATION

The baseline image classification was improved by using a different CNN architecture based on He et al. (2015). The paper presented a residual learning framework to ease the training of networks that are substantially deeper than other deep networks.

2.3 RADIOLOGY REPORT ANALYSIS

In the absence of X-rays, we can utilize the additional information we get from the radiology reports to determine the presence or absences of diseases. In order to do that, we traind a many to one long short term memory (LSTM) model that outputs the probability that a radiology report is normal or abnormal. we will train it on a publicly available radiology dataset which contains chest x-ray images and their associated reports published on the Web as part of Openi (2018).

In cases where both X-ray images and radiology reports are available, we combined the outputs of the above two models (fig: 1)to get better prediction accuracy.

2.4 HEATMAP

In order to make the baseline model more transparent, a visualization technique which was proposed in Selvaraju et al. (2016) was used. The approach uses the class-specific gradient information flowing into the final convolutional layer of a CNN to produce a coarse localization map of the important

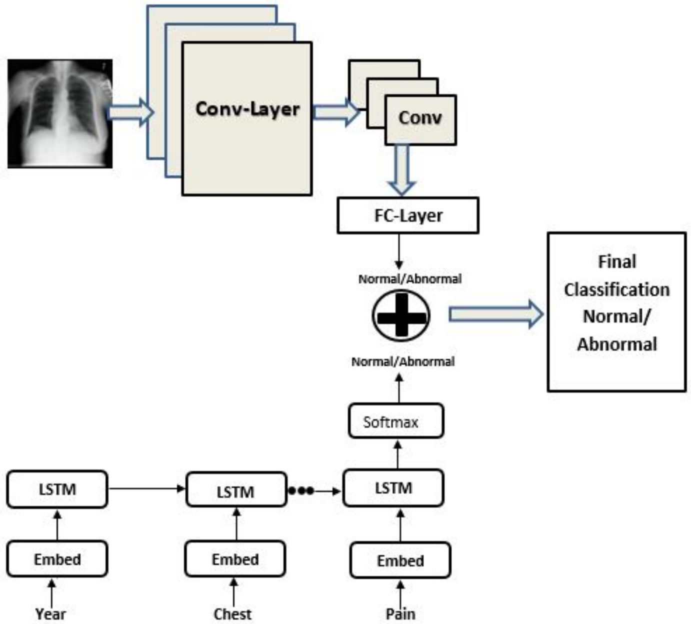

Figure 1: Combined classification

regions in the image. In order to obtain the class-discriminative localization map, Szegedy et al. (2014) computes gradient of the output with respect to feature maps. These gradients are global-average pooled to obtain weights. These weights are linearly combined with the feature maps and passed through a ReLU activation to generate heat maps. The final image produced is a heat map that localizes the area in the Chest X-ray which are most indicatives of abnormality.

Figure 2: Heat map

2.5 IMAGE CAPTIONING WITH ATTENTION

The final milestone of our project is going to be an attention-based model that automatically learns to describe the contexts of disease like location, severity, and the affected organs in addition to determining the presence and absence of disease. The attention model will have two parts. An encoder, a pre-trained convolutional neural network (VGGNet-19) and a decoder which is a long short-term memory (LSTM) network that produces a caption by generating one word at every time step conditioned on a context vector, the previous hidden state and the previously generated words Xu et al. (2015). By visualizing the attention at every time step we get to see which regions of the X-ray were focused on when a certain word (location, severity, organ etc) is generated.

3 RESULTS

The classification accuracy, with a 90% confidence interval, and the corresponding F1-scores of the different models are reported in Table 2. The receiver operator characteristic curve of the different models is shown in Figure 3.

Model	Accuracy	F1-score
Image Classifier	77±2.63%	0.827
Text Classifier	86.8±2.01%	0.897
Combined Model	87.5±2.04%	0.903

Table 2: Final Results

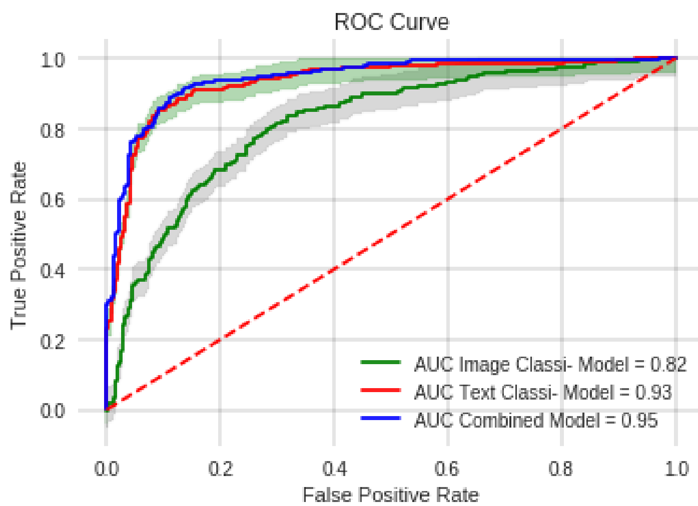

Figure 3: ROC of all the models

The results obtained from the image captioning model was not very accurate. The model over-fit the training data after few iterations of training. Due to time and different constants, we weren't able to fix the issue with the attention model.

4 RELATED WORK

Medical text classification using convolutional neural networks at a sentence level was proposed in Hughes et al. (2017). The work showed that it was possible to use CNNs to represent the semantics of clinical text enabling semantic classification at a sentence level. In Rajpurkar et al. (2017) an algorithm that can detect pneumonia from chest X-rays at a level exceeding practicing radiologists was developed. It was also showed that a simple extension of the algorithm was able to detect multiple diseases. In Wang et al. (2017) a new database, namely "ChestX-ray8", with text mined labels was presented. It demonstrated that commonly occurring thoracic diseases can be detected and even spatially-located via a unified weakly-supervised multi-label image classification and disease localization framework. A publicly available radiology dataset of chest X-rays and their reports, and image annotations were used to mine disease names and to train CNNs Shin et al. (2016). RNNs are then trained to describe the contexts of a detected disease, based on the deep CNN features. Significantly improved image annotation results were obtained using the recurrent neural cascade model by taking the joint image/text contexts into account. Qayyum et al. (2017) proposed a framework of deep learning for content based medical image retrieval system by using deep CNNs that was trained for classification of medical images. An intermodal dataset that contains twenty four classes and five modalities were used to train the network.

5 Future Work

Few extra steps that can be taken in the future that will improve the usability of the project include:

- Working with larger datasets, like the NIH Chest X-ray Dataset
- Solving the issue with the attention model
- Extracting disease names
- Building APIs

6 Data & Technical Requirements

We use two publicly available datasets. The first ChestX-ray14 which contains over 100,000 frontal-view X-ray images with 14 diseases. From this dataset, we extracted 20000 images 50% of which are healthy. The second radiology dataset we decided to use contains chest x-ray images and their associated reports published on the Web as part of Openi (2018).It has 3,955 radiology reports from the Indiana Network for Patient Care and 7,470 associated chest x-rays.

We would like to let the staff know that we have been forced to slightly deviate from our original proposal because of issues related to our the dataset. We expected to get a dataset which has 100,000 Chest X-ray images with their radiology reports and 10,000 of which have healthy/unhealthy labels as described in the project ideas document. Unfortunately, the dataset we weren't able to get X-ray images which had both labels and reports.

References

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. URL `http://arxiv.org/abs/1512.03385`.

Gao Huang, Zhuang Liu, and Kilian Q. Weinberger. Densely connected convolutional networks. *CoRR*, abs/1608.06993, 2016. URL `http://arxiv.org/abs/1608.06993`.

Mark Hughes, Irene Li, Spyros Kotoulas, and Toyotaro Suzumura. Medical text classification using convolutional neural networks. *CoRR*, abs/1704.06841, 2017. URL `http://arxiv.org/abs/1704.06841`.

W. L. Jackson. In radiology, turnaround time is king. June 2015.

Openi. An open access biomedical search engine, 2018. URL `https://openi.nlm.nih.gov`.

Adnan Qayyum, Syed Muhammad Anwar, Muhammad Awais, and Muhammad Majid. Medical image retrieval using deep convolutional neural network. *CoRR*, abs/1703.08472, 2017. URL `http://arxiv.org/abs/1703.08472`.

Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, Matthew P. Lungren, and Andrew Y. Ng. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *CoRR*, abs/1711.05225, 2017. URL `http://arxiv.org/abs/1711.05225`.

Ramprasaath R. Selvaraju, Abhishek Das, Ramakrishna Vedantam, Michael Cogswell, Devi Parikh, and Dhruv Batra. Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization. *CoRR*, abs/1610.02391, 2016. URL `http://arxiv.org/abs/1610.02391`.

Hoo-Chang Shin, Kirk Roberts, Le Lu, Dina Demner-Fushman, Jianhua Yao, and Ronald M. Summers. Learning to read chest x-rays: Recurrent neural cascade model for automated image annotation. *CoRR*, abs/1603.08486, 2016. URL `http://arxiv.org/abs/1603.08486`.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *CoRR*, abs/1409.4842, 2014. URL `http://arxiv.org/abs/1409.4842`.

Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M. Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *CoRR*, abs/1705.02315, 2017. URL http://arxiv.org/abs/1705.02315.

Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron C. Courville, Ruslan Salakhutdinov, Richard S. Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. *CoRR*, abs/1502.03044, 2015. URL http://arxiv.org/abs/1502.03044.