PNEUMONIA DETECTION USING CHEST X-RAYS

Potnuru Anusha, PVSL Hari Chandana, Sai Saketh Aluru, K Sai Surya Teja

Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

1. INTRODUCTION

Chest X-rays are currently the best available method for diagnosing pneumonia [1], playing a crucial role in clinical care and epidemiological studies [2]. Pneumonia was the 4th leading cause of death in the world in 2016, resulting in 3.0 million deaths worldwide. It was responsible for more than 1 million hospitalisations and 50,000 deaths per year in the US [3]. However, detecting pneumonia in chest X-rays is a challenging task that relies on the availability of expert radiologists. The appearance of pneumonia in X-ray images is often vague, can overlap with other diagnoses, and can mimic many other benign abnormalities. These discrepancies cause considerable variability among radiologists in the diagnosis of pneumonia [4].
In this work, we explore the task of automatically detecting pneumonia from chest X-rays using deep neural networks.

1.1. Background and Motivation

Chest radiography (chest X-ray or CXR) is an economical and easy-to-use medical imaging and diagnostic technique. The technique is the most commonly used diagnostic tool in medical practice and has an important role in the diagnosis of the lung disease. Well-trained radiologists use chest X-rays to detect illnesses, such as pneumonia, tuberculosis, interstitial lung disease, and early lung cancer.

The great advantages of chest X-rays include their low cost and easy operation. Even in underdeveloped areas, modern digital radiography (DR) machines are very affordable. Therefore, chest radiographs are widely used in the detection and diagnosis of the lung diseases, such as pulmonary nodules, tuberculosis, and interstitial lung disease. Chest radiography contains a large amount of information about a patient's health. However, correctly interpreting the information is always a major challenge for the doctor. The overlapping of the tissue structures in the chest X-ray greatly increases the complexity of the interpretation. For example, detection is challenging when the contrast between the lesion and the surrounding tissue is very low or when the lesion overlaps the ribs or large pulmonary blood vessels. Even for an experienced doctor, it is sometimes not easy to distinguish between similar lesions or to find very obscure nodules. Therefore, the examination of the lung disease in chest X-ray will cause a certain degree of missed detection. The wide application of chest X-rays and the complexity of reading them make computer-aided detection (CAD) systems a hot research topic since the system can help doctors to detect suspicious lesions that are easily missed, thus improving the accuracy of their detection. With automation at the level of experts, we hope that this technology can improve healthcare delivery and increase access to medical imaging expertise in parts of the world where access to skilled radiologists is limited.

Fig. 1: Chest X-ray of a normal, pneumonia

2. RELATED WORK

Automated diagnoses enabled by chest radiographs have received growing interests. These algorithms are increasingly being used for conducting lung nodule detection and pulmonary tuberculosis classification [5]. The performance of several convolutional models on diverse abnormalities relying on the publicly available OpenI dataset [6] found that the same deep convolutional network architecture does not perform well across all abnormalities [7], ensemble models significantly improved classification accuracy when compared with single model, and finally, deep learning method improved accuracy when compared to rule-based methods. Experiments using multiple disease classification datasets were also performed, and the statistical dependency between different disease labels was studied [8] to arrive at more precise predictions, thereby outperforming other techniques on given 13 images selected from 14 classes [9]. Algorithms for mining and predicting labels emanating from radiology images as well as reports have been studied [10], but the image labels were generally constrained to disease tags, thus lacking contextual information. Detection of diseases from X-ray

	Pneumonia	Normal
Train	3106	1079
Val	777	270
Test	390	234

Table 1: Dataset distributions.

images was examined in [11], classifications on image views from chest X-ray, and body parts segmentation from chest X-ray images and computed tomography was performed in [12].

ChexNet[13] is a recent work on various disease detection including pneumonia on a dataset of ChestX-ray14, currently the largest publicly available chest X- ray dataset, containing over 100,000 frontal view X-ray images with 14 disease labels. Many of the images in the dataset contained more than one disease mentions. ChexNet used a 121 layered DenseNet convolution network to train on this huge dataset and obtained state of the art performance. They also compare their results of 420 images with real expert radiologists, and observed that they were able to outperform the radiologists, proving the usefulness of having computer aided diagnosis systems. We try to follow the approach of this ChexNet model in our project and use the 121 layered DenseNet as one of the models that we experiment on. Unfortunately, we couldn't make use of the large dataset, and instead use another publicly available smaller X-ray image dataset as described below.

3. DATASET DESCRIPTION

We make use of the dataset of validated Chest X-Ray images described and analyzed in "Deep learning-based classification and referral of treatable human diseases"[14] for the training and evaluation of our models. The authors of the dataset have split it into a training set and a testing set, making sure that the both the sets contained independent patient records.

Table 1 contains the distribution of number of samples in train and test datasets. The dataset contains a total of 5,856 grayscale X-ray images, along with the corresponding label indicating whether the patient was diagnosed with Pneumonia or not. We further split the training portion of the dataset in a ratio of 80% - 20% in a stratified fashion. The 80% is used for training and the remaining 20% for cross validation of our models. The test set provided by previously mentioned dataset is used for evaluating the performance and those results are reported here.

The images in the dataset were of varied dimensions, ranging from around 700 x 400 pixels, to 1700x1300 pixels. No information regarding the patient details except the final diagnosis result were available.

4. MODEL

4.1. Problem Formulation

The pneumonia detection task is a binary classification problem, where the input is a frontal-view chest X- ray image X and the output is a binary label $y\epsilon\{0,1\}$ indicating the absence or presence of pneumonia respectively.
For a single example in the training set, we optimize the binary cross entropy loss as

$$L(X,y) = -w_{+}.ylogp(Y=1|X) - w_{-}.(1-y)logp(Y=0|X) \quad (1)$$

where $p(Y=i|X)$ is the probability that the network assigns to the label i, $w_{+} = |N|/(|P|+|N|)$, and $w_{-} = |P|/(|P|+|N|)$ with $|P|$ and $|N|$ the number of positive cases and negative cases of pneumonia in the training set respectively.

4.2. Preprocessing

Since the images in the dataset were of varied dimensions, we scaled them down to a dimension of 150x150 pixels, keeping in mind the computational limitations at our disposal. We also normalize the pixel values of the images to values between 0-1 by dividing each intensity value by 255.

4.3. Evaluation

For the evaluation of the model, we compute four metrics: accuracy, precision, recall and F1 score.
Accuracy: What proportion of the samples are predicted correctly?
Precision: What proportion of positive identifications was actually correct?
Recall: What proportion of actual positives was identified correctly?
F1 Score: It is the harmonic mean of precision and recall. It is a single score that balances both the concerns of precision and recall in one number. It takes account of both false positives and false negatives. Considering the context of the problem we are working on, we believe that recall is a more important metric here than precision.

5. EXPERIMENTS

We explored different convolution neural network architectures for our task of predicting Pneumonia from X-Ray images. We trained two custom CNN architectures which we name Model 1 and Model 2, and three popular Deep convolution network architectures, namely VGGNet, ResNet and DenseNet. All the models are trained with Adam optimizer, with an initial learning rate of $8e-5$, and the learning rate is reduced by a factor of 10 every time the validation loss plateaus. We make use of weighted binary cross entropy loss for the training of these models.

5.1. Model Architecture 1

We make use of 3x3 convolution layers in this model. The model architecture consists of convolution layers of filter sizes 16, 16, 32, 32, 64, 64, 96, 96, 128, 128 respectively with Relu activation after each of these layers. Max pooling layers of kernel size 2x2 are applied after every 2 convolution layers. The output obtained from the final convolution layer is flattened and passed to Dense layers of 64 followed by another Dense layer of size 2 to get final logits.

5.2. Model Architecture 2

This model is slight modification to the above model. It contains two additional convolution layers, i.e, layers with filter sizes 16, 16, 32, 32, 64, 64, 96, 96, 128, 128, 128, 128 with ReLU activation. Instead of fixing the filter dimension to 3x3, the filter dimension of 3 layers at time were taken as hyperparameters and the best result was obtained for the dimension values [(7x7),(5x5),(3x3),(3x3)]. Hence the first 3 layers have 7x7 filters, next three layers have 5x5 filters, and the rest 3x3 filters. The output obtained from the final convolution layer is flattened and passed to Dense layer of 64, followed by another dense layer of 2 to get the final logits.

5.3. VGGNet

VGGNet is a 16 layered convolution network architecture proposed by VGG group, Oxford [15]. This model makes use of the VGGNet, pretrained on ImageNet dataset. The output from the final convolution layer of VGGNet is flattened and passed through two fully connected layers of 1024 neurons and 2 neurons respectively, with a dropout layer in between to obtain the probability scores of the given input image.

5.4. DenseNet

DenseNet is a 121 layered, Deep convolution Network proposed by Huang et al.[16]. DenseNets improve flow of information and gradients through the network, making the optimization of very deep networks tractable. We replace the final fully connected layer to one that has only 2 neurons and train the model end to end.

5.5. ResNet

ResNet [17] stands for Residual Network. The architecture consists of skip connections, which are known to be very effective in the task of image classification. Here we make use of 50 layered ResNet architecture for our task, since the dataset is small and the original 150 layered ResNet architecture is likely to overfit the dataset.

6. RESULTS

We report the results that we obtained on the previously mentioned model architectures in Table 2. The evaluation metrics of accuracy, precision, recall, and F1 score are mentioned in the table.

6.1. Analysis

From the results in Table 2, we can observe that all the models achieve good recall scores on the dataset, with the Deeper architectures achieving better results. Out of all the 5 architectures we have examined, we found that DenseNet gives the best results in terms of all the metrics. Hence, the further experiments of producing the Class Activate Maps as detailed below, are done for this model. We also provide the epoch wise model loss and accuracy on the training and validation test sets in Fig.2

7. MODEL INTERPRETATIONS

To interpret the network architectures, we also produce heatmaps to visualise the areas of the image most indicative of the disease using class activation mappings [18]. To generate the CAMs,we feed an image into the fully trained network and extract the feature maps that are output by the final convolutional layer . Let f_k be the kth feature map and let $w_{c,k}$ be the weight in the final classification layer for feature map k leading to prediction $\hat{y}$. We obtain a map M_c of the most salient features used in classifying the image as having pneumonia by taking the weighted sum of the feature maps using their associated weights. Formally,

$$M_c = \sum_k w_{c,k} f_k \tag{2}$$

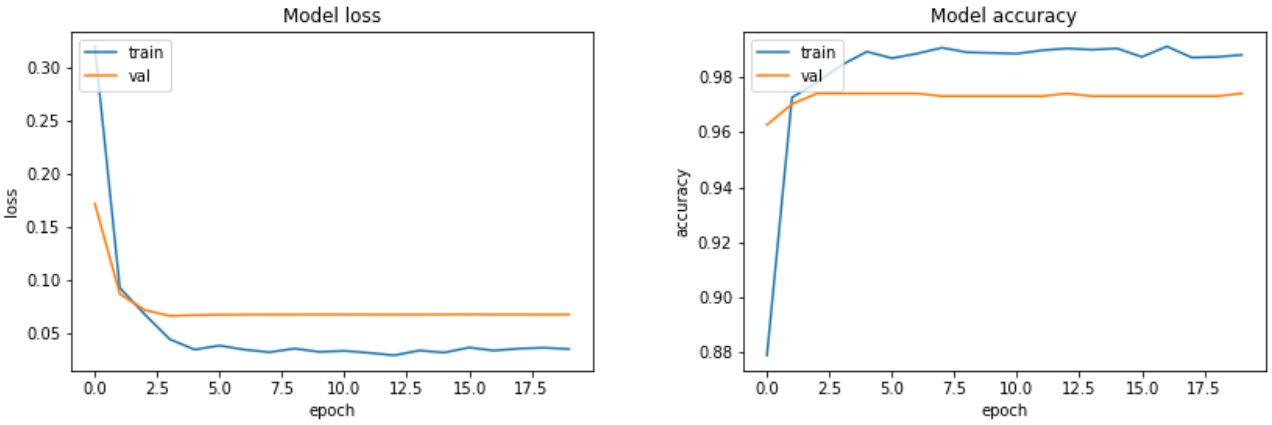
We use the DenseNet model that gave the best results in our case for this experiment. We identify the most important features used by the model in its prediction of the pathology by upscaling the map M_c to the dimensions of the image and overlaying the image. Fig.3 shows two such class activation maps for the X-Ray image of a pneumonia patient, along with the corresponding original image. We observe that the model not only predicts the correct label, but focuses mainly on the lower parts of the lungs of the patient to arrive at the conclusion.

8. LIMITATIONS

We identify two limitations in the task of Pneumonia classification that we explored here. Firstly, we note that only frontal radiographs were presented to the model during the training and evaluation phase, but it has been shown that upto 15% of

Table 2: Comparison of results using different models.

Model	Accuracy	Precision	Recall	F1-Score
Model 1	0.76	0.74	0.96	0.84
Model 2	0.795	0.79	0.93	0.855
Vgg	0.78	0.74	0.98	0.84
Resnet	0.84	0.80	0.99	0.88
Densenet	0.84	0.80	0.99	0.89



(a) Plot of loss vs epoch for DenseNet

(b) Plot of accuracy vs epoch for DenseNet

Fig. 2: Plots of loss and accuracy with epochs for baseline and DenseNet models.

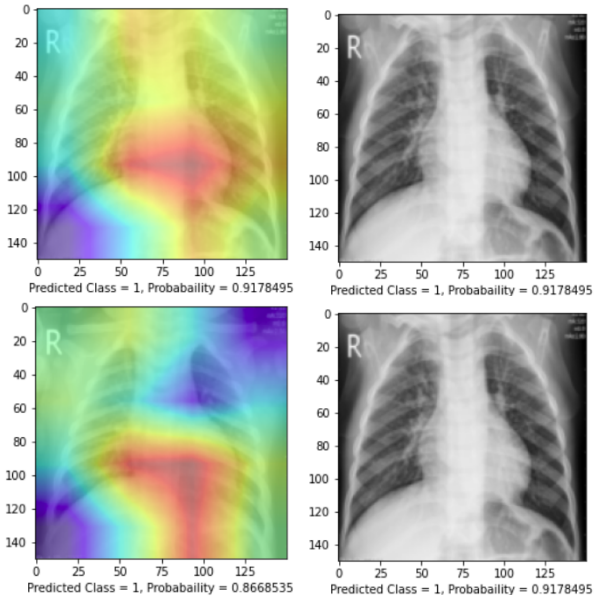


Fig. 3: Class Activation Maps of pneumonia positive cases.

the accurate diagnoses require the lateral view as well [19]. Secondly, the model didn't have any access to the patient's previous medical history, which has been shown to decrease radiologist diagnostic performance in interpreting chest radiographs. For e.g. given a pulmonary abnormality with a history of fever and cough, pneumonia would be appropriate.

9. CONCLUSION AND FUTURE WORK

In this project, we have explored few deep learning models for the task of classifying frontal X-Ray images to classify them as having Pneumonia or not. We observe that models like DenseNet and ResNet perform quite well in this task. In regards to the above specified limitations, the model could be further improved by providing the additional information like patient history or lateral X-rays and Computer aided Tomography (CAT) scans that are often useful in medical diagnosis. This approach of using automated diagnosis systems could even be applied to many similar diseases that make use of Chest X-rays such as Cardiomegaly, Effusion, Infiltration, etc. With automation at the level of experts,we hope that this technology can improve health care delivery and increase access to medical imaging expertise in parts of the world where access to skilled radiologists is limited.

10. REFERENCES

[1] World Health Organization. Pneumonia Vaccine Trial Investigators' Group and World Health Organization, "Standardization of interpretation of chest radiographs for the diagnosis of pneumonia in children / world health organization pneumonia vaccine trial investigators' group," 2001.

[2] Thomas Cherian, E. Kim Mulholland, John B Carlin, Harald Ostensen, Ruhul Amin, Margaret de Campo, David Greenberg, Rosanna Lagos, Marilla Lucero, Shabir A Madhi, Katherine L O'Brien, Steven Obaro, Mark K Steinhoff, and WHO Radiology Working Group, "Standardized interpretation of paediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies / thomas cherian .... [et al.]," *Bulletin of the World Health Organization : the International Journal of Public Health 2005 ; 83(5) : 353-359*, pp. Summaries in English, French, Spanish and Arabic, 2005.

[3] "Cdc, 2017," `https://www.cdc.gov/features/pneumonia/index.html`.

[4] Mark I Neuman, Edward Y Lee, Sarah Bixby, Stephanie Diperna, Jeffrey Hellinger, Richard Markowitz, Sabah Servaes, Michael C Monuteaux, and Samir S Shah, "Variability in the interpretation of chest radiographs for the diagnosis of pneumonia in children," *Journal of hospital medicine*, vol. 7, no. 4, pp. 294–298, 2012.

[5] Paras Lakhani and Baskaran Sundaram, "Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, 2017.

[6] Dina Demner-Fushman, Marc D Kohli, Marc B Rosenman, Sonya E Shooshan, Laritza Rodriguez, Sameer Antani, George R Thoma, and Clement J McDonald, "Preparing a collection of radiology examinations for distribution and retrieval," *Journal of the American Medical Informatics Association*, vol. 23, no. 2, pp. 304–310, 2016.

[7] Mohammad Tariqul Islam, Md Abdul Aowal, Ahmed Tahseen Minhaz, and Khalid Ashraf, "Abnormality detection and localization in chest x-rays using deep convolutional neural networks," *arXiv preprint arXiv:1705.09850*, 2017.

[8] Li Yao, Eric Poblenz, Dmitry Dagunts, Ben Covington, Devon Bernard, and Kevin Lyman, "Learning to diagnose from scratch by exploiting dependencies among labels," *arXiv preprint arXiv:1710.10501*, 2017.

[9] 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," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2097–2106.

[10] Hoo-Chang Shin, Le Lu, Lauren Kim, Ari Seff, Jianhua Yao, and Ronald M Summers, "Interleaved text/image deep mining on a very large-scale radiology database," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1090–1099.

[11] Uri Avni, Hayit Greenspan, Eli Konen, Michal Sharon, and Jacob Goldberger, "X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words," *IEEE Transactions on Medical Imaging*, vol. 30, no. 3, pp. 733–746, 2010.

[12] Simon Hermann, "Evaluation of scan-line optimization for 3d medical image registration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3073–3080.

[13] Kaylie Zhu Brandon Yang Hershel Mehta Tony Duan Daisy Ding Aarti Bagul Curtis Langlotz Katie Shpanskaya Matthew P. Lungren Andrew Y. Ng Pranav Rajpurkar, Jeremy Irvin, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," .

[14] Kang; Goldbaum Michael (2018) Kermany, Daniel; Zhang, ""labeled optical coherence tomography (oct) and chest x-ray images for classification, mendeley data, v2"," `http://dx.doi.org/10.17632/rscbjbr9sj.2`.

[15] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015.

[16] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger, "Densely connected convolutional networks," *CoRR*, vol. abs/1608.06993, 2016.

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.

[18] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba, "Learning deep features for discriminative localization," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[19] Suhail Raoof, David Feigin, Arthur Sung, Sabiha Raoof, Lavanya Irugulpati, and Edward C. Rosenow, "Interpretation of plain chest roentgenogram," Feb. 2012.