MULTILABEL CLASSIFICATION OF THORAX DISEASES WITH NIH CHEST X RAYS USING TRANSFER LEARNING

Sayan Guha
9 July,2019
sguha3436@gmail.com

1 Introduction

Chest X-ray exams are one of the most frequent and cost-effective medical imaging examinations available.A substantial number of chest x-rays are used to diagnose a plethora of conditions. These diagnoses are still primarily done by radiologists manually poring over each scan, with no automated assistance.In this project I aim to use deep learning methods to predict thorax disease categories using chest x-rays.The problem can be cast as a multiclass multilabel image classification problem with 14 different labels.

2 Data

The NIH Chest X ray Dataset is the largest publicly available dataset.It comprises of 112,120 X-ray images with disease labels from 30,805 unique patients.To create these labels, the authors used Natural Language Processing techniques to text-mine disease classifications from the associated radiological reports.The labels are expected to be greater than 90 % accurate and suitable for weakly-supervised learning.The image dimensions are 1024× 1024 pixels each of which originally has corresponding information on patient age, gender, ID, and number of followup visits to the hospital.

3 Problem Formulation

The actual labels can be one or many of 14 common thorax disease types, which include Atelectasis, Cardiomegaly, Consolidation, Edema, Effusion, Emphysema,Fibrosis, Hernia, Infiltration, Mass, Nodule, Pleural Thickening, Pneumonia, and Pneumothorax. Hence this is a example of a multilabel classification problem in which the input is a frontal-view chest X-ray image and the output is a 14 dimensional binary vector y ϵ $\{\ 0, 1\}$ indicating the presence or absence of the above listed 14 diseases.

4 Methodology

4.1 Data Preprocessing

The dataset contains a total of 112,120 images,out of which 60,361 images belong to the No finding category.These images are removed from training and testing purposes.The dataset is then split into 80 % training and 20% test set resulting in 41287 training images and 10352 test images.The images are scaled down from 1024× 1024 pixels to 224× 224 pixels and are normalized on the mean and standard deviation of images in theImageNet training set before inputting them to the neural network.

4.2 Network Architectures

Transfer Learning is used for classification of diseases.Pre built models on the Imagenet Dataset such as ResNet50,InceptionV3 and DenseNet Architectures were used for this purpose.A comparison of the architectures used and the results which they give are presented in this report.

The first transfer learning base model used is ResNet.They were created originally for the ImageNet challenge to allow training of deeper nets while minimizing the difficulty of transforming the data through so many layers.ResNet-50 is a modified version of the first 152layer net, but shares the same architectural characteristics,with direct transfer of activations to the next layer to ensuregood features learned early in the network are not skewedor lost in subsequent layers.

The second base model used was InceptionNet Version3.The salient features included adding batch normalization to the auxiliary classifiers which did not contribute much to the training processes and acted as regularizers.Also smart factorization techniques like representing a 5×5 convolution by two 3×3 convolution and a n× n convolution by a 1×n followed by a n×1 convolution which made the convolutions less expensive and also reduced the number of parameters which further helped in reducing overfitting were used.Also representational bottleneck was reduced based on the intuition that neural networks perform better when convolutions didnt alter the dimensions of the input drastically.

The third base model is the DenseNet121.DenseNets improve flow of information and gradients through the network, making the optimization of very deep networks tractable.It is similar to ResNets except that at layer L ResNets usually have 1 skip connection from the previous layer but DenseNets have L-1 skip connections thus making gradient flow easier.It also removes the vanishing gradient problem,strengthen feature propagation, encourage feature reuse, and substantially reduces the number of parameters.

4.3 Training

For training purpose each image is scaled down to 224× 224 pixels and are normalized on the mean and standard deviation of images in the ImageNet training set before inputting them to the respective network.The fully connected layers of ResNet,InceptionNet V3 and DenseNet are removed and are replaced by a global average pooling layer.The final layer is a fully connected layer with 14 nodes having "sigmoid" nonlinearity applied to it and each node specifies the probability of the presence of the 14 listed diseases.

The loss function used is the unweighted binary cross entropy loss

L(X,y)=$\sum_{c=1}^{14}[-y_c logp(Y_c = 1|X) - (1 - y_c)logp(Y_c = 0|X)]$

where $p(Y_c = 1|X)$ is the predicted probability that the image contains the pathology c and $p(Y_c = 0|X)$is the predicted probability that the image does not contain the pathology c.

The weights of the networks are initialized with weights from the model pretrained on ImageNet.Only the weights of the final layers are trained using Adam with standard parameters(β1= 0.9 and β2= 0.999).An initial learning rate of 0.001 is used.

The batch size used for training was 32.The step size for each epoch was 100 and overall 400 epochs were used for training each network.

5 Results

The metric used for comparison are the AUROC Scores for each disease which are listed in Table 1.Also Figures 1-3 shows the plot of the ROC curves for each disease and using the different transfer learning architectures for the test set which consists of 10352 images.

Pathology	Resnet50	InceptionV3	DenseNet121
Atlectasis	0.6738	0.7307	0.7375
Cardiomegaly	0.8104	0.8717	0.8806
Consolidation	0.6738	0.6856	0.6875
Edema	0.8041	0.8327	0.8260
Effusion	0.7692	0.8068	0.8096
Emphyesma	0.7083	0.7624	0.7669
Fibrosis	0.7241	0.7354	0.7389
Hernia	0.7364	0.8307	0.8566
Infiltration	0.6476	0.6686	0.6590
Mass	0.7426	0.7462	0.7735
Nodule	0.6697	0.6669	0.6781
PleuralThickening	0.6942	0.6847	0.6896
Pneumonia	0.6251	0.6609	0.6142
Pneumothorax	0.7041	0.7880	0.7726

Table 1: AUROC SCORES FOR EACH PATHOLOGY

The outputs for each image are probabilities indicating the presence of the 14 listed pathologies for each image.To assign labels to each image thresholding is used.The level of thresholding is empirically determined and varies from 0.05 to 0.3 for different pathology labels.Table 2 summarizes the various statistics for each pathology and for different learning architecture.For evaluation purposes the accuracy is defined as the ratio of total number of true positives retrieved by the deep learning system to the total number of images which contain that particular pathology in the actual test set.

6 DISCUSSIONS

From the above AUROC scores it seems the DenseNet Architecture with pre-trained weights on the Imagenet seems to be the most suitable for classification of most of the thorax diseases.However InceptionNet architecture performs better in detecting Edema,Infiltration and Pneumothorax while ResNet50 achieves higher AU-

Pathology	Total Images	Training	Testing	TP(DenseNet121)	Accuracy	TP(InceptionNetV3)	Accuracy	TP(ResNet50)	Accuracy
Atlectasis	11559	9239	2320	1951	84.09	2201	94.87	1941	83.63
Cardiomegaly	2776	2266	510	364	71.37	225	44.11	152	29.80
Consolidation	4667	3726	941	670	71.2	583	61.95	737	78.32
Edema	2303	1809	494	250	50.60	210	42.50	273	55.26
Effusion	13317	10646	2671	2407	90.11	2532	94.79	2649	99.17
Emphyesma	2516	2027	489	131	26.78	319	65.23	248	50.71
Fibrosis	1686	1354	332	157	47.28	207	62.34	194	58.43
Hernia	227	189	38	16	42.10	1	2.63	0	0
Infiltration	19894	15895	3999	3823	95.59	2622	65.56	3326	83.17
Mass	5782	4581	1201	645	53.7	934	77.76	652	54.28
Nodule	6331	5124	1207	751	62.22	836	69.26	616	51.03
PleuralThickening	3385	2683	702	225	32.05	307	43.73	26	31.17
Pneumonia	1431	451	280	88	31.42	51	18.21	67	23.92
Pneumothorax	5302	4244	1058	476	44.99	663	62.66	783	74.00

Table 2: ACCURACY PERCENTAGE AND OTHER STATISTICS FOR EACH PATHOLOGY

ROC scores for PleuralThickening.The fact that DenseNet architectures encourage feature reuse and has more skip connections for flow of gradients has helped the system to get better results.

An important point to note is that the dataset is highly imbalanced.The number of images having labels as Infiltration is the highest while Hernia's count only stands at 227.It can be seen from Table 2 that classes such as Atelectasis,Effussion and Infiltration which are significantly well represented have higher accuracy scores returned by the transfer learning architecture than the ones which are significantly under represented like Hernia.

From the ROC figures we can easily verify that that all the classifiers have score >0.5 which means the classifiers can easily discriminate between the positive and negative classes.

From Table2 if we compute the mean accuracy of the 3 learning architectures we can easily see that DenseNet outperforms the other two method.

7 CONCLUSION

Early diagnosis and treatment of thorax pathologies is critical to preventing complications including death. With approximately 2billion procedures per year, chest X-rays are the most common imaging examination tool used in practice,critical for screening, diagnosis, and management of a variety of diseases. However, two thirds of the global population lacks access to radiology diagnostics, according to an estimate by the World Health Organization.Leveraging fine-tuned Deep Convolutional Neural Network architectures for automatic classification of chest x-ray images at an expert level might help in alleviating this problem.With the development of automatic multiclass multilabel algorithms for detection of thorax diseases it can be hoped that Deep Learning technology can improve healthcare delivery and increase access to medical imaging expertise in parts of the world where access to skilled radiologists is limited.

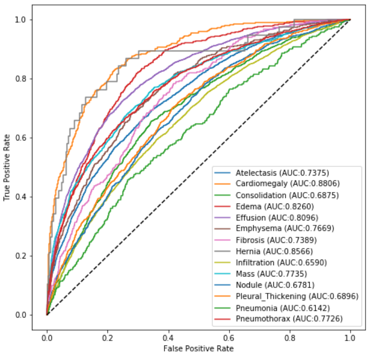

Figure 1: AUROC FOR DENSENET

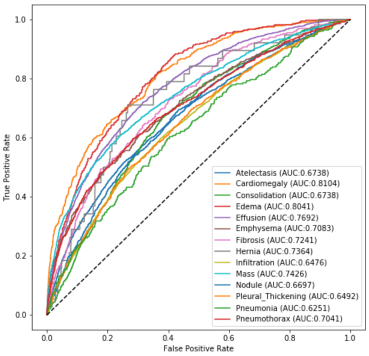

Figure 2: AUROC FOR RESNET

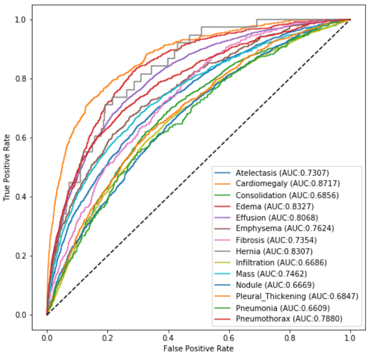

Figure 3: AUROC FOR INCEPTION NET