ADAPTIVE DATA AUGMENTATION TRAINING BASED ATTENTION REGULARIZED DENSENET FOR DIAGNOSIS OF THORACIC DISEASES

Roshan Shetty*

Research Scholar, School of Electronics & Communication Engineering, Reva University
Bengaluru, Karnataka-560064, India
roshan564956@gmail.com

Prasad Narasimha Sarapadi

Professor, School of Electronics & Communication Engineering, Reva University
Bengaluru, Karnataka-560064, India
prasadsn@reva.edu.in

Abstract

Recently researchers using deep learning in particularly CNN have demonstrated breakthrough performance in a variety of medical image analysis tasks and Chest radiography is one among them. The Chest X-Ray dataset called Chest X-Ray14 was used to develop existing Deep Learning models to classify 14 multiple thoracic diseases in real-time. The proposed model is based on two major observation of existing models. First, many existing methods used entire Chest X-Ray images for training purposes since resizing of images results in relatively low-resolution images, at the same time systems are made to focus equally on the entire image instead of more on abnormal regions. Second, this dataset is highly imbalanced disease class distribution, thus the learned feature doesn't have enough discriminatory powers to do effective classification. To address these issues proposed model comes with a separate Grad-CAM-based attention mechanism to guide Deep model such as DenseNet-121 to focus more on regions of pathological abnormalities and improve classification performance on resized images. To deal with data imbalance, an adaptive data augmentation technique is introduced so that system improves its performance on minority class diseases labels. Our experiments demonstrated, that the proposed model outperformed several other state-of-art deep learning models in terms of our achieved AUC scores.

***Keywords*: Thoracic diseases; Deep CNN; DenseNet; Adaptive Data Augmentation; Chest X-Ray14.**

1. Introduction

Thoracic diseases are associated with cavity of thorax such as lungs, heart, esophagus, chest wall, and diaphragm in human body, causing serious health problems. Chest X-Ray is one of the frequent procedure while diagnosing these diseases. Accurate interpretation of X-ray images by radiologists is critical for diagnosis of these diseases. Deep learning with CNNs performance on medical images is increasing day by day. They offers several benefits to radiologists in terms of increased workflow, decreased workload, better performance and improved health care to the patients [Yasaka and Abe (2018)]. Deep learning models to identify one or two thoracic diseases motivated an approach to develop models that can identify multiple disease in a real time [Rajpurkar *et al.* (2017)]. However meaningful applications can be created from Deep Neural Network (DNN) only if we feed enough data to data hungry DNN. For Chest X-Ray first largest public data in digital form is Open-I [13], which contains 395 radiology reports and 7470 associated images. Major breakthrough happened when National Institute of Health (NIH) released Chest X-ray8 dataset with 8 pathological labels as shown in Fig. 1, [Wang *et al.* (2017)]. Soon this database is extended to 14 labels and named as Chest X-Ray14 dataset. Many DNN approaches on this dataset utilized entire image for training purpose with very high computational complexity [Liu *et al.* (2019)]. Also resizing of images resulted in low resolution images with loss of details. In fact these details are very much essential for diagnosis, especially pathologies with very small lesion area like Nodule. Most DNNs usually fails to focus more on abnormal regions to make their diagnosis decisions.

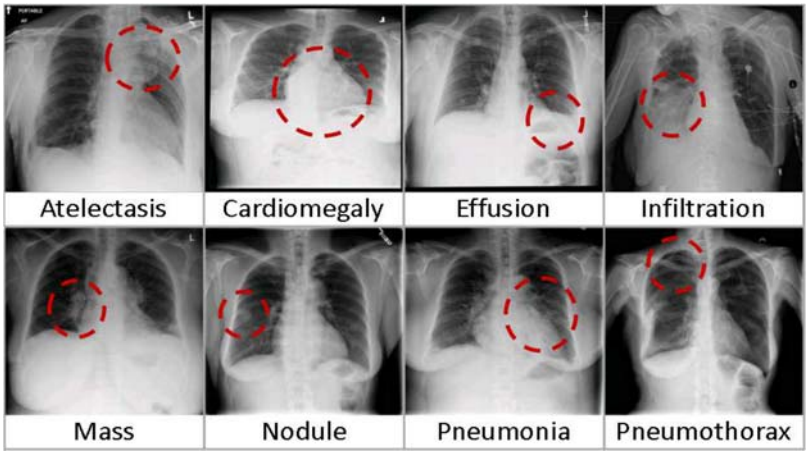

Fig.1. Chest X-ray8 dataset with 8 common thoracic disease labels

Hence visual mechanism will assist DNNs to pay more attention on uncommon regions of resized X-Ray and localize them effectively [Wang *et al.* (2020 a)]. Chest X-Ray14 dataset is very sensitive to misclassification due to imbalanced 14 pathology distribution. For example thoracic disease class called infiltration is the largest with 85.57 times larger than smaller class called Hernia [Yang *et al.* (2020)]. The Fig.2. shows imbalance nature of ChestX-Ray14 dataset. Due to imbalance DNN models may not be trained adequately and fails to acquire enough discriminatory powers to do effective classification.

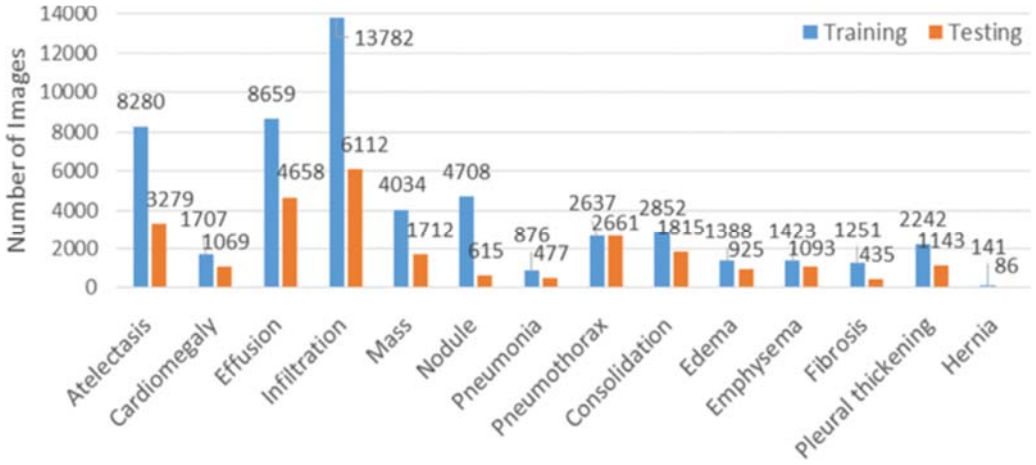

Fig. 2. Imbalanced disease class distribution in ChestX-Ray14 dataset.

In order to solve such an issues, we propose Adaptive Data Augmentation training based Attention Regularized DenseNet for effective classification of multiple thoracic disease. This proposed model consists of an attention mechanism, which aims to produce class discriminative maps of abnormal regions on image and enhance classification performance. Attention mechanism is achieved with the help of separate attention block. It consists of Grad-CAM [Selvaraju *et al.* (2017)] as a core element in between convolution layers. The aggregate predictions of DenseNet block and attention block are taken to classify thoracic diseases on each resized lower resolution input images. In this work, to tackle the imbalance nature of dataset, we propose Adaptive Data Augmentation training technique. Specifically, we monitor the performance of a model during each training iteration. Whenever the model shows relatively poor performance on a disease class that class will be emphasized more during next iteration.

The notable contribution of the proposed method are :- (1) Attention block with Grad-CAM as a core element in between stacked convolutional layers. They convert extracted features from DenseNet into attention maps. These attention maps highlight disease specific abnormal regions in an image, and (2) In order to tackle disease class imbalance, Adaptive Data Augmentation training technique is utilized. Last but not least, the proposed method has the advantage of providing the best representation of images and has the capacity to handle large imbalanced dataset such as Chest X-Ray14 data.

2. Related Work

To process images, deep learning architecture known as CNNs are used widely. The success of CNN encouraged AI researchers to use it on the diagnosis of thoracic diseases also. Designing a model for Chest X-Ray related

diagnosis is very formidable due to complexity of thoracic diseases and also the less quantity of radiographs in digital form. Classification of the detected diseases is also a difficult task, because some radiographs may contain multiple disease also. The recent researches involved in developing the deep learning models for thorax diseases are reviewed in this section.

The [Wang *et al.* (2017)] proposed thorax diseases classification framework with 4 separate pre-trained CNN models for the initialization of deep network. Global pooling and prediction layers are used for disease localization through heatmap. However they used binary relevance approach to predict labels even though there are more than one pathology in some radiographs. The [Yao *et al.* (2018)] proposed a model to exploit statistical dependencies between all 14 pathologies and classification improvement is observed. The [Gundel *et al*. (2018)] developed Location-aware Dense Network (DNetLoc) for thoracic disease detection. The DNetLoc shows improved performance, when a location information of pathologies are used explicitly. The DNetLoc method showed limitations in an official split of images, due to the high disease imbalance in a dataset. The [Rajpurkar *et al.* (2017)] used Densenet-121 and batch normalization with improved accuracy on Chest X-Ray14 dataset. They used CAM generated heatmap to interpret network predictions. Their per-class AUC score on pneumonia was reached to radiologist level prediction. A study by [Gunn *et al.* (2018)] developed attention guided CNN framework (AG-CNN). Here attention mechanism is achieved by using binary mask via thresh holding the feature maps on same dataset. Their CAM generated heatmap produced some unreliable classification and failed to indicate some of the abnormal regions. The [Wang *et al*. (2020 a)] developed Thorax-Net model where the ResNet architecture was used as a backbone network supported by Gradient-weight Class Activation Maps (GRAD-CAM) instead of CAM. The uniqueness of Grad-CAM is, attention is produced by adding self-generated gradient signal with extracted feature maps from ResNet. Performance of this model was effected by imbalance nature of dataset Chest X-Ray14.

3. Materials and Method

In the year 2017, the Chest X-Ray8 database was released by NIH (National Institutes of Health) with 8 pathology labels mined from associated radiology reports through NLP techniques. In the same year, this database was extended to 14 thoracic diseases and made public as Chest X-Ray14 dataset [Wang *et al.* (2017)]. It contains 112,120 frontal view images with 15 labels. Out of 15, there are 14 different thoracic disease labels plus 'No Finding'. There are 51,708 images containing more than one disease label and 60,412 are normal cases belongs to 'No Finding' class. This dataset poses a multiple thoracic disease classification problem. The dataset is available in both Patient-wise official split and Image-wise random-split. All images are available in Portable Network Graphics (PNG) format with 1024×1024 rescaled size.

3.1. *Problem and Objective*

For each input, classifier returns a 14 D vector probabilities of thoracic diseases. Note that if all probabilities are Zero, the image X belongs to the 15th class of 'No Finding'. Objective is to provide a good attention mechanism to identify and localize lesion regions in an image and guide DenseNet to pay more attention on those regions rather than focusing on entire resized low resolution images. Further proposed model $M(X)$ performance should not be effected by imbalance nature of dataset and produce more accurate predictions for minority disease class also.

3.2. Architecture *of a proposed model*

This section explains the proposed method along with architecture as shown in Fig. 3.

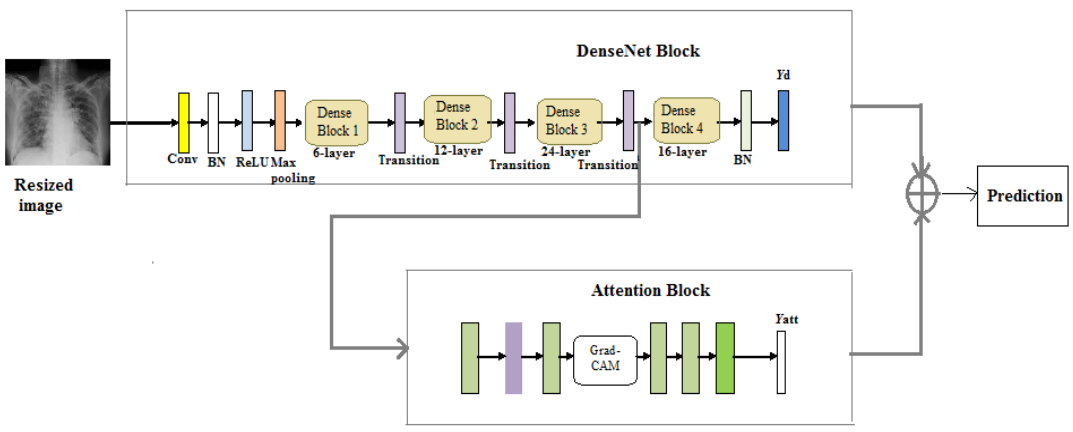

Fig.3. Architecture of a proposed model

3.2.1. Feature Extractor and Classifier

Motivated by the results obtained by pre-trained CNN models on medical image interpretation, we selected CNN pre-trained model such as DenseNet-121 as both feature extractor and classifier. All pre-trained CNNs are trained with the ImageNet database which contains images other than X-Rays also. This prevents random initialization of the parameters in a new model. The DenseNet has been recognized as having a stronger capacity on various computer vision tasks than other pre-trained CNN models, such as ResNet and VGG. DenseNet consists of series of dense blocks followed by a transition layer. The dense block consists of series of densely connected units where each unit consists of two convolution layers preceded by Batch Normalization (BN) and ReLu activation. The transition layer performs down sampling of generated feature maps from a dense block. Prior to first dense block a sequence of Convolution, Batch Normalization, and ReLu followed by max-pooling are used.

To adopt DenseNet to our problem of 14 disease classification, we kept only 14 neurons in the last fully connected layer and replaced their activation with sigmoid function. Label prediction vector $Y_{d.}$ is used to indicate the probability of 14 diseases. If all vectors are zero, it represents 'No finding' or 15th class.

3.2.2 Attention mechanism

Separate attention block is proposed here to highlight pathological abnormalities by analyzing learned feature maps from DenseNet. This block is further aimed to improve the classification performance of the proposed model. The Gradient-weighted Class Activation Mappings (Grad-CAM) is used as a core element here. It takes gradient of output class into final convolution layers and produces attention maps to highlight abnormal regions. Attention mechanism is very effective to localize disease like nodule with very small lesion area in resized images. Input to Grad-CAM is, learned feature maps from DenseBlock 3. The reason is deeper layers of Deep CNN have richer visual and semantic information. The feature maps are made to travel through 1×1, 3×3, and 1×1 convolution layers followed by ReLU function. This process further refines feature maps before feeding into Grad-CAM. The class-specific attention map $\bar{A}_c$ for each disease class *c* as shown below in "Eq. (1)". Which is a weighted combination of a *k-th* feature map from third convolution layer denoted as $\tilde{A}_k$. Where α_{ck} is a class-specific weight followed by ReLU function in order to remove the negative picture elements.

$$\bar{A}_c = ReLU\left(\sum_k \alpha_{ck}\ \tilde{A}_k\right) \quad (1)$$

The last three convolutional layers with 1×1, 1×1, and 14×14 kernels convert class discriminative attention maps into a classification result. To adopt attention block to our problem of 14 disease classification, we kept only 14 neurons in the last fully connected layer and replaced their activation with sigmoid function. Label prediction vector $Y_{att.}$ is used indicate the probability of 14 diseases. Finally, both Y_{att} and Y_d are combined and average is taken to produce final diagnosis predictions of the proposed model.

3.3. *Training and Testing*

The parameters or weights of neural networks are optimized through a training process. In this proposed training, key factor is handling of the imbalanced dataset. Due to imbalance, learned features of a network may not have enough discriminatory powers to do accurate classifications [Wang *et al.* (2020a)]. Before training we resized original PNG formatted images to 224×224, and normalized them to [-1, 1] range. The proposed training is carried out on training, validation sets of both Patient-wise official split, and Image-wise random-split of Chest X-Ray14 dataset separately.

The Adaptive Data Augmentation training consists of following steps on training dataset:

- It consists of total *T* iterations, for *k*-th disease class. Setup an augmentation ratio $\delta_k >= 1$, which is initially 1.
- For every *T*, *k*-th class image is augmented by δ_k times larger.
- Later this augmented dataset is shuffled and split into equal-sized batches before feeding into the model. The model optimization is carried out with stochastic gradient descent (SGD) algorithm.

The loss function used here is unweighted binary cross-entropy as shown in "Eq. 2". Where p^T is probabilistic prediction during a single training sample of (X, y^T).

$$L(X, y^T; p^T) = \sum_{k=1}^{K}(-y_k \log p_k - (1 - y_k)\, log(1 - p_k)) \quad (2)$$

The Adaptive augmentation training consists of following steps on validation dataset:

- The validation set is used to authenticate trained model. Here e_k, is denoted as a performance parameter of the trained model on the *k-th* validation data.

- Based on the ranking position of generated e_k, set up π_k. For example, if $e_1 = 0.86$, $e_2 = 0.62$, *and* $e_3 = 0.9$, then set $\pi_1 = 2$, $\pi_2 = 3$, and $\pi_3 = 1$ since $e_3 > e_1 > e_2$.
- Efficiency of a model on *k-th* class is measured by π_k value. Large π_k value indicate poor performance of a model on *k-th* disease class. So we need focus more training of *k-th* disease class.
- Revise augmentation ratio δ_k by a factor of π_k as shown in "Eq. 3".

$$\delta_k = \frac{\pi_k - 1}{K-1}(\Delta - 1) + 1 \qquad (3)$$

- Where the parameter $\Delta > 1$ is the maximum possible augmentation ratio. This is how minority classes will get more iteration and solves class imbalance issue.

We adopted the mini-batch stochastic gradient descent algorithm with a batch size of 32, 64, 128, and 256. The used learning rate is 0.001, and the maximum iteration number of 500.

4. Results and Discussion

4.1. *Experimental Setup and Evaluation Metrics*

The proposed method of thoracic disease detection is used Intel i7 processor system configurations, 500GB of hard disk and 8GB of RAM, GPU configuration is 22GB NVIDIA RTX 2080Ti. Time taken train the model is 22 hours and implemented using PyTorch.

The performance of the proposed method is evaluated using following metric: Area Under Curve (AUC). It measures entire two-dimensional area under the Receiver Operating Characteristic (ROC) curve. AUC is always in the range of 0 to 1. The ROC is plotted by creating True Positive Rate (TPR) against False Positive Rate (FPR) at various threshold values.

4.2. *Experiment with Patient-wise official split of dataset*

Table 1 shows result and performance comparison in terms of AUC scores of the proposed method with existing works [Yao *et al.* (2018)], [Gundel *et al.* (2018)] , and [Wang *et al.* (2020 a)] by using a Patient-wise official split.

Table. 1. Performance comparison with AUC scores of proposed method with existing methods in terms of Patient-wise official split.

Thoracic diseases	**[Yao *et al.* (2018)]**	**[Gundel *et al.* (2018)]**	**[Wang *et al.* (2020a)]**	**Ours**
Atelectasis	0.733	0.767	0.7505	**0.951**
Cardiomegaly	0.856	0.883	0.8710	**0.941**
Effusion	0.806	0.828	0.8181	**0.892**
Mass	0.718	0.821	0.7994	**0.898**
Pneumonia	0.684	0.738	0.6938	**0.862**
Consolidation	0.711	0.745	0.7415	**0.821**
Edema	0.806	0.835	0.8354	**0.952**
Emphysema	0.842	0.895	0.8428	0.879
Fibrosis	0.743	0.818	0.8040	0.802
Infiltration	0.673	0.709	0.6815	**0.811**
Nodule	0.777	0.758	0.7147	**0.917**
Pleural Thickening	0.724	0.761	0.7463	**0.931**
Pneumothorx	0.805	0.846	0.8254	0.836
Hernia	0.775	0.896	0.9022	**0.952**
Average	0.761	0.807	0.7876	**0.8985**

This split includes 80% of images as training set, 20% as a testing set. We used 10% images of training set for validation purpose. The training set includes 50,500 normal and 36,024 diseased images; similarly testing set includes 9912 normal and 15,684 diseased images. The uniqueness of this split is images of the same patient must appear either in training or in the test set but not in both. The method [Gundel *et al.* (2018)] used external dataset PLCO along with Chest X-Ray14 data. Our proposed method outperformed all three existing models without being trained by external data. The higher average AUC scores of 0.8985 (89.85%) is achieved for the detection of 14 thoracic diseases compared to three existing methods as shown in Table 1. In 11 out of total 14 cases we got high per-class AUC scores, indicated with bold black in Table 1.

4.3. *Experiment with Image-Wise Random split*

Here complete data is randomly split into 70% as training, 10% as validation, and 20% as testing set. When Image-wise random split of dataset is utilized, the radiograph of similar subject appears simultaneously in both training and testing set. The proposed method showed effective performance using this split than existing methods by achieving an average AUC of 0.9307 (93.07%). Table 2. Shows in 10 out of total 14 cases we got high per-class AUC scores, indicated with bold black.

Table 2. Performance comparison with AUC scores of proposed method with existing methods in terms of Image-Wise Random split.

Thoracic diseases	**[Wang *et al.* (2017)]**	**[Rajpurkar *et al.* (2017)]**	**[Wang *et al.* (2020a)]**	**Ours**
Atelectasis	0.716	0.821	0.856	**0.943**
Cardiomegaly	0.807	0.905	0.957	0.932
Effusion	0.784	0.883	0.919	0.911
Mass	0.706	0.862	0.905	**0.943**
Pneumonia	0.633	0.763	0.869	**0.938**
Consolidation	0.708	0.794	0.870	**0.923**
Edema	0.835	0.893	0.943	**0.963**
Emphysema	0.815	0.926	0.959	0.921
Fibrosis	0.769	0.804	0.889	**0.914**
Infiltration	0.609	0.720	0.776	**0.922**
Nodule	0.671	0.777	0.832	**0.924**
Pleural Thickening	0.708	0.814	0.883	0.859
Pneumothorax	0.806	0.893	0.941	**0.964**
Hernia	0.767	0.939	0.951	**0.973**
Average	0.738	0.842	0.896	**0.9307**

In both split, our method achieved higher average AUC values compared to other competitors. The major reason is the incorporation of the attention mechanism with the deep learning model such as DenseNet-121. It guides feature extractor to focus more on abnormal regions rather than focusing equally on the entire image. Also learned features from DenseNet with Adaptive Data Augmentation training has enough discriminatory powers to do the effective classification of pathologies.

4.4. *Contribution of Adaptive Data Augmentation Training and Attention Block*

We used Adaptive Data Augmentation training which belongs to oversampling training technique where key idea is, poorly performed minority disease images are augmented more. The loss function used is unweighted binary cross-entropy as shown in "Eq (2)". Our results reveal that effect of imbalanced disease distribution on our model is very less.

Two other training techniques are under-sampling and penalize miss-classification. Under-sampling utilizes

less training data than available. This is not suitable for this dataset, since the dataset itself is small and imbalanced. In third technique, during training process a model is penalized with additional cost for doing classification mistakes [Wang *et al.* (2020 a)]. The weighted cross-entropy or standard binary cross-entropy loss functions are used here. In this difficulty is manually fixing of weights for different misclassifications.

Comparison of our method with the two existing methods:

1) Thorax-Net model [Wang *et al*. (2020 a)] used weighted binary cross entropy loss function to penalize misclassifications to overcome class imbalance issue. ResNet is used as backbone network with attention mechanism. The slight improvement is observed in minority disease classification such as pneumonia, edema, and hernia classifications. It is highlighted with bold black in Table 3. Performance of this model is highly effected by imbalance nature of dataset. Our method outperformed their AUC score on above diseases with good margin as shown Table 3 with bold black. Also our method achieves best AUC performance in 13 out of the 14 cases compared to [Wang *et al*. (2020 a)].

2) Adaptive sampling, unweigted binary cross entropy loss function is used to overcome class imbalance issue of Chest X-Ray14 dataset in a model **[**Yang *et al.* (2020)]. DenseNet-121 is used as backbone network without any attention mechanism. They achieved moderately good AUC scores on minority disease classes such as edema, emphysema and hernia. But pathologies with very small lesion area, such as nodule and mass their scores are average shown in bold black in Table 3. The underlying reason is due to fact that model failed to adaptively concentrate on pathologically small abnormal regions in an image. Our attention regulated DenseNet showed good results on mass and nodule in terms of AUC scores due attention mechanism as shown in bold black Table 3. Also our method got high per-class AUC in11 out of the 14 compared to [Yang *et al.* (2020)].

Table 3. Performance comparison of proposed method with existing methods under Patient-wise official split, in terms of training techniques used.

Thoracic diseases	**[Yang *et al.* (2020)]**	**[Wang *et al.* (2020 a)]**	**Ours**
Atelectasis	0.814	0.750	0.951
Cardiomegaly	0.899	0.871	0.941
Effusion	0.873	0.818	0.892
Mass	**0.840**	0.799	**0.898**
Pneumonia	0.662	**0.693**	**0.862**
Consolidation	0.789	0.741	0.821
Edema	0.874	**0.835**	**0.952**
Emphysema	0.924	0.842	0.879
Fibrosis	0.809	0.804	0.802
Infiltration	0.701	0.681	0.811
Nodule	**0.775**	0.714	**0.917**
Pleural Thickening	0.772	0.746	0.931
Pneumothorax	0.865	0.825	0.836
Hernia	0.923	**0.902**	**0.952**
Average	0.821	0.787	**0.8985**

To measure the importance of visual mechanisms in our model, we conducted further experiments. We removed the attention block and recorded predictions only from the DenseNet block. We didn't change any hyper

parameter settings, training technique, and fine-tuned DenseNet-121without any attention block. Table 4 Shows AUC values of our model with, and without attention block. The proposed model with attention block showed better results with average per-class AUC increased from 0.747 (74.7%) to 0.8985 (89.85%). Therefore, the attention mechanism is very critical to improve classification performance of the models. Fig. 4. shows learned attention heatmap of pathologies Atelectasis (left) and Nodule (right).

Table 4. AUC Scores of proposed method with and without attention branch methods

Thoracic diseases	Our Model Without Attention Block	Our Model With Attention Block
Atelectasis	0.788	**0.951**
Cardiomegaly	0.811	**0.941**
Effusion	0.773	**0.892**
Mass	0.741	**0.898**
Pneumonia	0.672	**0.862**
Consolidation	0.687	**0.821**
Edema	0.811	**0.952**
Emphysema	0.789	**0.879**
Fibrosis	0.759	**0.802**
Infiltration	0.612	**0.811**
Nodule	0.705	**0.917**
Pleural Thickening	0.762	**0.931**
Pneumothorax	0.765	**0.836**
Hernia	0.823	**0.952**
Average	**0.747**	**0.8985**

Fig. 4. Learned attention heatmap of pathologies Atelectasis (left) and Nodule (right).

4.5. Limitations

The proposed method has a distinct advantage over other existing models in learning disease-discriminative features and also learning from the imbalanced dataset. However, some issues need to be considered. They are correlation among labels and utilizing them as a prior knowledge in a model. The reason is some of the X-Ray images contain multiple pathology labels also. There are some X-Ray images contain manual bounding boxes over abnormal regions, which are not considered here for disease localization purpose. Since we used unweighted loss function during adaptive data augmentation training, so any misclassification is equally unfortunate in our

model.

5. Conclusion

In this work, the effort has been made to develop a deep learning-based advanced thoracic disease classifier. The adaptive data augmentation training is proposed to overcome the data imbalance problem of popular Chest X-Ray14 dataset. A Grad-CAM based separate attention block is added to the DenseNet block to improve its classification performance on resized images. Only 14 neurons are kept at the final layers of both the blocks and aggregate is taken to produce final predictions of the classifier. The proposed method has the advantage of providing the best representation of images and the capacity to handle large imbalanced dataset. We compared our model with other models using patient-wise official split and image-wise random split of dataset. Our model obtained a higher average AUC of 89.85% and 93.07% in both split respectively. In future, the thoracic disease detection performance can be enhanced by powering attention modules to exploit the relationship between disease labels, disease location, and scales of learned feature maps. It is required to produce a more diverse Chest X-Ray dataset in the future with more pathology subjects, and less imbalanced disease class distribution.

Acknowledgments

Authors appreciate the support of MedPac System Solutions LLP, Bengaluru during this work. We are grateful to staffs, and infrastructure provided by school of Electronics and Communication Engineering, REVA University Bengaluru, India.

References.

[1] Allaouzi, I. and Ahmed, M.B. (2019): "A novel approach for multi-label chest X-ray classification of common thorax diseases". *IEEE Access*, *7*, pp.64279-64288.
[2] Chamveha, I. et al (2020): "Local Adaptation Improves Accuracy of Deep Learning Model for Automated X-Ray Thoracic Disease Detection: A Thai Study". *arXiv preprint arXiv:2004.10975*.
[3] Chen, B. *et al.* (2019): "DualCheXNet: dual asymmetric feature learning for thoracic disease classification in chest X-rays". *Biomedical Signal Processing and Control*, *53*, p.101554.
[4] Guan, Q. *et al* (2018): "Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification," *arXiv*: 1801.09927.
[5] Gundel, S. *et al* (2018): "Learning to recognize abnormalities in chest x-rays with location-aware dense networks". In *Iberoamerican Congress on Pattern Recognition* (pp. 757-765). Springer, Cham.
[6] He, H. and Garcia, E. (2009): "Learning from imbalanced data". TKDE, 362 21(9):1263–1284.
[7] He, K. *et al* (2016): "Deep residual learning for image recognition" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 770_778.
[8] Ho, T. K. K. and Gwak, J.(2020): "Utilizing Knowledge Distillation in Deep Learning for Classification of Chest X-Ray Abnormalities". *IEEE Access*, *8*, pp.160749-160761.
[9] Huang, G *et al.* (2016): "Densely connected convolutional networks" in *Proc. IEEE Conf. Comput. Vis.Pattern Recognit. (CVPR)*, Jul. 2016, pp. 4700_4708. [Online]. Available: https://arxiv.org/abs/1608.06993.
[10] Irvin, J. *et al.* (2019): "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison." [Online]. Available: https://arxiv.org/abs/1901.07031.
[11] Li, X. *et al*. (2019) "Deep learning-enabled system for rapid pneumothorax screening on chest CT". *European journal of radiology*, *120*, p.108692.
[12] Liu, H. *et al*. (2019): "SDFN: Segmentation-based deep fusion network for thoracic disease classification in chest X-ray images", Computerized Medical Imaging and Graphics, Vol. 75, pp. 66-73.
[13] Open-I: An open access biomedical search engine. https://openi.nlm.nih.gov
[14] Rajpurkar, P. *et al*. (2017): "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning". *arXiv*: 1711.05225 [cs, stat]. *arXiv: 1711.05225.*
[15] Selvaraju, R.R. *et al* (2017): "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 618–626.
[16] Simonyan, K and Zisserman, A. "Very deep convolutional networks for large-scale image recognition" in *Proc. ICLR*, 2015, pp. 1_14. [Online].Available: https://arxiv.org/abs/1409.1556.
[17] Wang, H. *et al* (2020 a): "Thorax-Net: An Attention Regularized Deep Neural Network for Classification of Thoracic Diseases on Chest Radiography". *IEEE journal of biomedical and health informatics*, *24*(2), pp.475-485.
[18] Wang, H. *et al* (2020 b): "Triple attention learning for classification of 14 thoracic diseases using chest radiography", Medical Image Analysis, Vol. 67, pp. 101846, 2020.
[19] Wang, X. *et al.* (2017): 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* (pp. 2097-2106).
[20] Yang, Y.-Y. *et al* (2020): "Detecting thoracic diseases via representation learning with adaptive sampling". *Neurocomputing*. doi:10.1016/ j.neucom.2019.06.113
[21] Yao, L. et al (2017): "Learning to diagnose from scratch by exploiting dependencies among labels", *arXiv preprint arXiv*: 1710.10501.
[22] Yao, L. *et al*. (2018): "Weakly supervised medical diagnosis and localization from multiple resolutions", arXiv preprint arXiv: 1803.07703.
[23] Yasaka, K. and Abe, O. (2018): "Deep learning and artificial intelligence in radiology: Current applications and future directions. PLoS Med15 (11)" e1002707. https://doi.org/10.1371.

Author's Profile

Roshan Shetty: Research Scholar in School of ECE at REVA University, Bengaluru, India. Graduation and Post-graduation from VTU Belagavi. Total 11 years in teaching experience in Engineering Field. My areas of Interest: AI and Embedded Systems.

Prasad Narasimha Sarappadi: Professor in School of ECE at REVA University, India since 2012.Total experience is 22 Years, completed graduation from Mangalore University, Post-graduation from VTU and Doctorate from Jain University. More than 80 Journals/Conferences in profile and presently guiding 8 research scholars. Area of Interest is AI, Embedded Systems and Signal Processing