Identifying pneumonia in chest X-rays: A deep learning approach

Amit Kumar Jaiswal [a], Prayag Tiwari [b], Sachin Kumar [c], Deepak Gupta [d], Ashish Khanna [d], Joel J.P.C. Rodrigues [e,f,g,*]

[a] Institute for Research in Applicable Computing, University of Bedfordshire, United Kingdom
[b] Department of Information Engineering, University of Padova, Padova, Italy
[c] Department of System Programming, South Ural State University, Chelyabinsk, Russia
[d] Maharaja Agrasen Institute of Technology, Delhi, India
[e] National Institute of Telecommunications (Inatel), Santa Rita do Sapucaí, MG, Brazil
[f] Instituto de Telecomunicações, Portugal
[g] Federal University of Piauí, Teresina, PI, Brazil

ABSTRACT

The rich collection of annotated datasets piloted the robustness of deep learning techniques to effectuate the implementation of diverse medical imaging tasks. Over 15% of deaths include children under age five are caused by pneumonia globally. In this study, we describe our deep learning based approach for the identification and localization of pneumonia in Chest X-rays (CXRs) images. Researchers usually employ CXRs for the diagnostic imaging study. Several factors such as positioning of the patient and depth of inspiration can change the appearance of the chest X-ray, complicating interpretation further. Our identification model (https://github.com/amitkumarj441/identify_pneumonia) is based on Mask-RCNN, a deep neural network which incorporates global and local features for pixel-wise segmentation. Our approach achieves robustness through critical modifications of the training process and a novel post-processing step which merges bounding boxes from multiple models. The proposed identification model achieves better performances evaluated on chest radiograph dataset which depict potential pneumonia causes.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

Pneumonia is one of the leading causes of death among children and old age people around the world. It is an infection caused by a virus, bacteria or other germs. Pneumonia results in inflammation in lungs which can be life threatening if not diagnosed in time. Chest X-ray is an important pneumonia diagnosis method worldwide. However, an expert knowledge and experience is required to read the X-ray images carefully. Therefore, the process of pneumonia detection by reading X-ray images can be time consuming and less accurate. The reason is that several other medical conditions i.e. lung cancer, excess fluid etc. can also show similar opacities in images. Therefore, accurate reading of images is highly desirable. The power of computing is world known and developing an identification model for finding pneumonia causes in clinical images can help in accurate and better understanding of the X-ray images.

X-ray image analysis is considered as tedious and crucial tasks for radiology experts. Therefore, researchers have proposed several computer algorithms to analyze X-ray images [1,2]. Also, several computer assisted diagnosis tools [3–5] have been developed to provide an insight of X-ray images. However, these tools are not able to provide sufficient information to support doctors in making decisions [6]. Machine learning is a promising approach in the field of artificial intelligence. A plenty of research works have been carried out to investigate the chest and lung diseases using machine learning. Vector quantization, regression neural networks has been used to investigate chest disease [7]. In another study [8], chronic pneumonia disease was analyzed and its diagnosis was implemented using neural networks. Another study [9] used chest radiographic images for the detection of lung diseases. They applied histogram equalization for image pre-processing and further feed forward neural network was used for classification. Although the above mentioned studies have performed efficiently, however lacks in terms of higher accuracy, computational time and error

---

* Corresponding author.

E-mail addresses: amitkumar.jaiswal@beds.ac.uk (A.K. Jaiswal), prayag.tiwari@unipd.it (P. Tiwari), deepakgupta@mait.ac.in (D. Gupta), ashishkhanna@mait.ac.in (A. Khanna), joeljr@ieee.org (J.J.P.C. Rodrigues).

rate. Deep learning has already been proved an effective approach in object detection and segmentation, image classification, natural language processing etc. Further, deep learning has also shown its potential in medical image analysis for object detection and segmentation such as radiology image analysis in order to study anatomical or pathological structures of human body [10–12]. Also, deep learning provided higher accuracy than traditional neural network architectures.

In the remainder of this article, we first review the literature related to pneumonia identification in chest X-ray images in Section 2 followed by proposed model architecture in Section 3 detailing algorithm and training steps in different stages. We have detailed our extensive analysis of RSNA dataset in Section 4 with image augmentation steps including the result from cleaned data, and evaluation metrics followed by evaluation result in Section 5 of our proposed model as well as ensembles of our model. Finally, we conclude our work in Section 6 along with future work.

2. Literature survey

Roth et al. [13] demonstrated the power of deep convolutional neural network (CNN) to detect the lymph node in clinical diagnostic task and obtained drastic results even in the presence of low contrast surrounding structures obtained from computer tomography. In another study, Shin et al. [14] addressed the problems of thoraco-abdominal lymph detection and interstitial lung disease classification using deep CNN. They developed different CNN architectures and obtained promising results with 85 percent sensitivity at three false positives per patient. Ronneburger et al. [15] developed a CNN approach with the use of data augmentation. They suggested that even trained on small samples of image data obtained from transmitted light microscopy; the developed model was able to capture high accuracy. Jamaludin et al. [16] applied CNN architecture to analyze the data obtained from spinal lumber magnetic resonance imaging (MRI). They developed an efficient CNN model to generate radiological grading of spinal lumber MRIs.

All these studies have performed well on radiological data except that the size of the data was restricted to few hundred samples of patients. Therefore, a detailed study is required to use the power of deep learning over thousand samples of patients to achieve the accurate and reliable predictions. Kallianos et al. [17] presented a state of art review stating the importance of artificial intelligence in chest X-ray image classification and analysis. Wang et al. [18] addressed this issue and prepared a new database ChestX-ray8 with 108,948 front view X-ray images of 32,717 unique patients. Each of the X-ray images could have multiple labels. They used deep convolutional neural networks to validate the results on this data and obtained promising results. They mentioned that chestX-ray8 database can be extended by including more disease classes and would be useful for other research studies.

Rajpurkar et al. [19] developed a 121 layer deep convolutional layer network chestX-ray14 dataset. This dataset is publically available with more than 0.1 million front view X-ray images with 14 disease labels. They mentioned that their algorithm is capable to predict all 14 disease categories with high efficiency. Irvin et al. [20] stated that large labeled dataset is the key to success for prediction and classification tasks. They presented a huge dataset that consists of 224,316 chest radiographic images of 65,240 patients. They named this dataset as CheXpert. Then they used convolutional neural networks to assign labels to them based on the probability assigned by model. Model used frontal and lateral radiographs to output the probabilities of each observation. Further, they released the dataset as a benchmark dataset. Besides the availability of a large dataset, it is highly desirable that every

object in the image should be detected carefully and segmentation of each instance should be done precisely. Therefore, a different approach is required to handle both instance segmentation and object detection. Such powerful methods are faster region based CNN (F-RCNN) [21] and FCN (Fully Convolutional Network) [22].

Moreover, F-RCNN can be extended with an additional branch for segmentation mask prediction on each region of interest along with existing branches for classification task. This extended network is called Mask R-CNN and it is better than F-RCNN in terms of efficiency and accuracy. Kaiming He et al. [23] presented Mask R-CNN approach for object instance segmentation. They compared their results with best models from COCO 2016 [24,25]. Luc et al. [26] extended their approach by introducing an instance level segmentation by predicting convolutional features.

3. Proposed architecture

In this section, we formulate and explore the problem pipeline followed by our model based on Mask-RCNN in detecting pneumonia symptoms from chest X-ray images.[1]

3.1. Problem settings

The problem consists of binary classification of chest X-rays on three different classes of lung opacities such as opacity, no opacity and not normal. The major issue is dissimilarity in quality X-rays in terms of brightness, resolution and position region of interest. To model such task, we describe our algorithm that can detect the visual signal for pneumonia in medical chest radiographs, and output either pneumonia positive or negative, and if positive it also returns predicted bounding boxes around lung opacities.

3.2. Modeling

In this section, we describe our modeling approach based on Mask-RCNN [23] which aims to identify lung opacity that are likely to depict pneumonia. Mask-RCNN is a deep neural network developed to solve instance segmentation in particular. Initially, we illustrate how we employed faster region based convolutional network [21] with pixelwise instance segmentation [27] for classification and localisation to build our model. We first input an image from chest X-ray sample data which goes through ROIAlign classifier extracting features from the input radiograph, and then F-RCNN model which then instantiated for pixelwise segmentation and makes a bounding box of the input image. It returns the predicted labels of image reported in Fig. 1. We use ResNet101 as a backbone detector in Mask-RCNN model and also compared it by replacing the detector with ResNet50 which is a convolutional network.

In perspective of pneumonia identification, Mask-RCNN model takes chest X-ray image as an input and predicts the bounding boxes of the image, label, mask including classes. It extends the algorithm of F-RCNN by adding a branch which induces binary mask predicting whether the given image pixel contributes to the given part of the object or not. Also, it is easy to train a Mask R-CNN and it adds a small overhead in terms of running time which is negligible. So, we may consider Mask R-CNN as an advanced faster R-CNN. We trained a RetinaNet [28] model which is a classic approach for object detection. However, the approach does not work well in all scenarios especially in the case on non-vertical/horizontal objects. With Mask R-CNN this issue can be resolved. Our Mask-RCNN based model gives more accurate pixelwise semantic segmentation than faster-RCNN for pneumonia

---

[1] Chest X-ray images are in DICOM format: https://en.wikipedia.org/wiki/DICOM
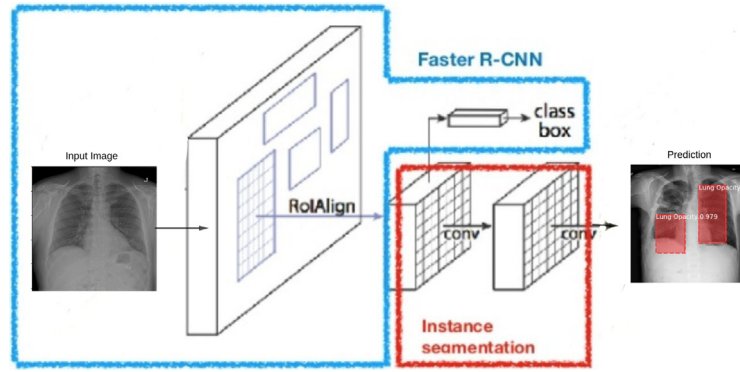
Fig. 1. Mask R-CNN based model for opacity identification and pixel-wise disease segmentation.

prone region in lungs around the rectangular bounding boxes. For instance, refer to the input image and prediction sample in Fig. 1.

3.3. Algorithm

Our algorithmic approach for identifying potential pneumonia causes is devised by Faster-RCNN [21]. We also tried several other object detection techniques such as You Look Only Once (YOLO3) [29] and U-Net [15] image detection architectures but it fails to produce better predictions, from our tests, we found that Mask-RCNN performing better in prediction tasks. We implemented the base network of Mask-RCNN pre-trained on COCO weights[2] using typical residual convolutional neural network (i.e., ResNet50 and ResNet101 [30]) for extracting features of actual human lungs and ROIAlign as a classifier and bounding box as a regressor. We performed pixel-wise segmentation of lung opacity selected by ROI classifier which helps scaling during inferencing and losses. We used multi-task loss [31] for training our model (identification and classification) and estimated hyperparameters based on a 10% stratified sample from the validation set of the training data. We employed stochastic gradient descent (SGD) with an initial learning rate of 0.00105 for training, the overall training time is 11.2 h for 20 epochs with batch size and image size of 16 and 512×512. Initially, we trained our model on positive images only followed by fine-tuning the model on all images, in which the foreground IoU threshold sets to 0.3. We also performed augmentation[3] during the training process. Our final prediction set is generated by post-processing followed by ensemble of two trained models on the entire training data. In every model, we generated unique predictions as our primary focus is on bounding boxes which "hits", which was post-processed, the parameter values are given in Table 1: The post-processing step is performed in three phases in which the very initial *Phase 0* applies non-maximum suppression[4] to fold prediction for 2 varied 5-fold assignments which takes three parameters as reported in Table 1 which identifies *maximum sustainable overlap, confidence threshold and minimum average confidence* followed by *Phase 1* which solicits classification probability to the output of phase 0 which consist of a set containing patient ID and their corresponding predictions. In *Phase 2*, we have the confidence threshold of our model which adds high confidence outcomes from phase 1 to the result of our identified set of pneumonia (from our identification model). Finally, in *Phase 3*, we convert the Mask-RCNN confidence to fitted class probability by taking maximum confidence for every patient and then merge it with class predictions. We aggregate

Table 1
List of parameters in post-processing stage.

Type of threshold	Value
Maximum overlap [Phase 0]	0.05
Confidence [Phase 0]	0.94
Minimum average confidence [Phase 0]	0.65
Class probability [Phase 1]	0.175
Confidence [Phase 2]	0.975
Below threshold confidence [Phase 3]	0.69

the confidence score for each and every bounding boxes before the post-processing stage using the following equation:

$$\hat{S}_c = \frac{1}{F}\sum_i S_{c,i} \tag{1}$$

where $\hat{S_{c,i}}$ is the confidence score for every ith bounding box for the cth set, F depicts a scaling factor and $\hat{S}_c$ is the ensemble of confidence score for the cth set. The bounding box from the ensembling of cth set is calculated by

$$\hat{P_{cl}} = median\{P_{cl}\} + \alpha.\sigma_{cl} \tag{2}$$

where P_{cl} denotes the group of pixel locations of the corner l (top-right, top-left, bottom-right, or bottom-left) of bounding boxes for each cth set, α depicts the scaling factor with value 0.1, σ_{cl} depicts the standard deviation of P_{cl}, and $\hat{P_{cl}}$ depicts pixel location of the ensembled bounding box for corner l. We discard those ensembled bounding boxes from the final prediction set which has a confidence score less than 0.25.

3.4. Training

We employ the RSNA pneumonia dataset[5] which is a subset of NIH CXR14 dataset [18]. For identification of pneumonia task, we have the following examples[6] or features from the stage 1 and stage 2 dataset: In Table 2, we have less examples in Stage 1 in comparison to Stage 2 dataset, though the exact number of lung opacity feature in Stage 1 dataset is 8964, which is the total number of ground truth bounding boxes, and 5659 is the total number of patients with pneumonia. Each of these patients has 1–4 ground-truth bounding boxes. Similarly, in Stage 2 dataset we have 6012 patient which is caused by pneumonia and a slightly greater number of lung opacity feature which is 9555 than Stage 1 dataset.

---

[2] https://github.com/matterport/Mask_RCNN/releases/download/v2.0/mask_rc-nn_coco.h5.

[3] The augmentation step discussed in later section.

[4] https://github.com/jrosebr1/imutils/blob/master/imutils/object_detection.py.

[5] The RSNA pneumonia dataset can be found at https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data.

[6] Here examples represent the patients' chest examination during data collection from expert team of radiographers

Table 2
Features of RSNA dataset.

# Features	Stage 1	Stage 2
Normal	8525	8851
Lung Opacity	5659/8964	6012/9555
Abnormal	11,500	11,821

Table 3
RSNA training and test image set.

# Images	Stage 1	Stage 2
Train Set	25684	26684
Test Set	1000	3000

Also, we give a brief overview on training and test data of RSNA dataset in these two stages reported in Table 3: The detailed information about the difference between number of images in Stage 1 and 2 is given in later section below.

We have examined the training data classifying the positive and negative features among patients' of different age reported in Fig. 2, whereas the various feature class among patients' of different age group reported in Fig. 3.

4. Experimental evaluation

4.1. Data preparation and augmentation

Dataset: We used a large publicly available chest radiographs dataset from RSNA[7] which annotated 30,000 exams from the original 112,000 chest X-ray dataset [18] to identify instances of potential pneumonia as a training set and STR[8] approximately generated consensus annotations for 4500 chest X-rays to be used as test data. The annotated collection contains participants ground truth which follows training our algorithm for evaluation. The sets containing 30,000 samples is actually made up of 15,000 samples with pneumonia related labels such as 'Pneumonia', 'Consolidation', and 'Infiltration', where 7500 samples are chosen randomly with 'No Findings' label, and another randomly selected 7500 samples without the pneumonia related labels and 'No Findings' label. They created a unique identifier for each of those 30,000 samples.

Annotation: Samples were annotated using a proprietary web based annotation system and permanently inaccessible to the other peoples' annotations. Every radiologists practitioners who took part in training initially executed on the similar set of 50 exemplary chest X-rays in a hidden manner, and then were visibly annotated to the other practitioners for the similar 50 chest X-rays for evaluations, as it enable for questions such as does an X-ray with healed rib fractures enumerate and no enumeration as 'Normal' and for preliminary calibration. The final sets of label comprises of as given in Table 4: There are 'Question' labels to suggest questions which will be answered by a chest radiologist practitioner. Overall coequal distribution of 30,000 human lungs is annotated by six radiologists experts to assess whether the collected images of lungs opacities equivocal for pneumonia with their analogical bounding box to set forth the status. Also, other twelve experts from STR collaborated in annotating fairly 4500 human lungs. Out of 4500 triple read conditions, we divided these chest X-rays into three sets containing 1500 human lungs in training set, 1000 in test set (initial stage) and rest 2000 in test set at final stage. However, the test sets is double checked by five radiologist practitioners including six other radiologists from the first group.

Primary consideration: We discuss the adjudicate during the data collection for such sophisticated task. A bounding box is assessed as isolated in multi-read case provided that it does not coincide with the bounding boxes of the other two readers i.e., these two readers fails to flag that particular area of the image as being unsure for pneumonia. Whenever the adjudicator concurs

that the isolated bounding box is valid then the box will endures a positive minority belief, in other cases it will be discarded. Initially, they assigned a confidence score to the bounding boxes. Also, a low confidence bounding boxes was discarded and high/intermediate boxes was aggregated into a group of appropriate pneumonia. Given a low probability based bounding box, they discard the box and check whether the labeling is abnormal or no lung opacity. The opposed bounding boxes was adjudicated by one of two thoracic radiology practitioners in multi-read cases which does not consent. Also, the practitioners found that the annotations of all three readers in adjudicated case is more than 15%. They used intersection for the rest of the bounding boxes in case of at least 50% coincide by one of the bounding boxes, this step has ample effect in discarding few pixel data for multiple readers including positive pixels. They used 1500 read cases out of the 4500 triple cases into the training set to average out few probable distinction among single and multi-read cases. Rest 3000 triple read cases allocated to the test set. The majority vote is used to distinguish weak labels. The radiologists followed the below requirement during data collection:

1. Bounding box (Lung Opacity): A patient's chest radiograph includes finding of fever and cough for potential signs of pneumonia.
2. They made few conjectures during the sabbatical of lateral radiography, serial examination and clinical information.
3. Based on Fleischner's [32] work, they considered every region which was more opaque than the neighbouring area.
4. They also excluded area such as nodule(s), evident mass(es), linear atelectasis and lobar collapse.

Data augmentation: We performed augmentation on lung opacities and images data with random scaling including shifting in coordinate space $((x_1, y_1), (x_2, y_2))$ as well as increasing/decreasing brightness and contrast including blurring with Gaussian blur under batches. Following these image augmentation, we found images after augmentation reported in Fig. 4.

Considering the outcome in Figs. 2 and 3 signifies status of patient class and labels from the X-ray images, as we see a highly imbalanced dataset. The imbalance of training and test dataset among too many negatives and too few positives generates a critical issue, as we want high recall but the model could predict all negatives to attain high accuracy and the recall significantly undergo. In this case, it is unsure whether or not the present imbalance is acceptable or not. We test whether balancing the class distribution would yield any improvement in this case. To do this, we have trained our model on two training data sets, one balanced and the other not. We then create the balanced dataset by augmenting more images to the negative (0) class. We discussed previously the augmentation steps which includes flipping, rotating, scaling, cropping, translating, and noise adding. Introducing the images in the current negative class can possibly create radically new feature that does not exist in the other class. For example, if we choose to flip every negative-class image, then we have in the negative class a set of images that have the right and left parts of the bodies switched while the other class does not have this feature. This is not desirable because, the network may learn unnecessary (and incorrect) features such as the image with the left part of the body being to a certain side is more likely to exhibit non-pneumonia. So

---

[7] Radiological Society of North America.
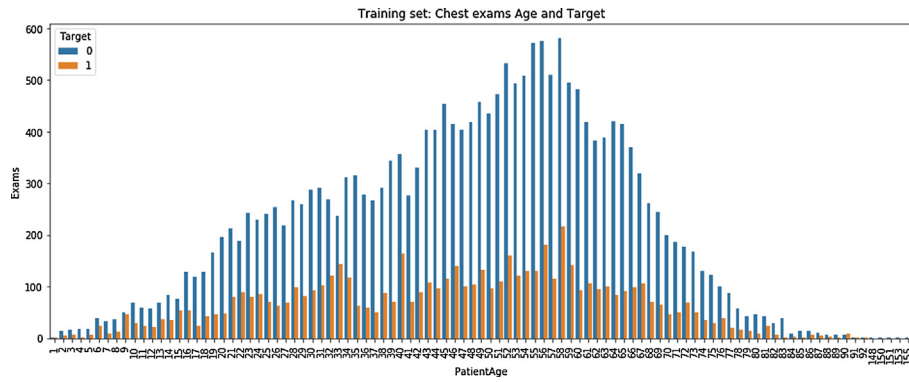[8] Society of Thoracic Radiology.

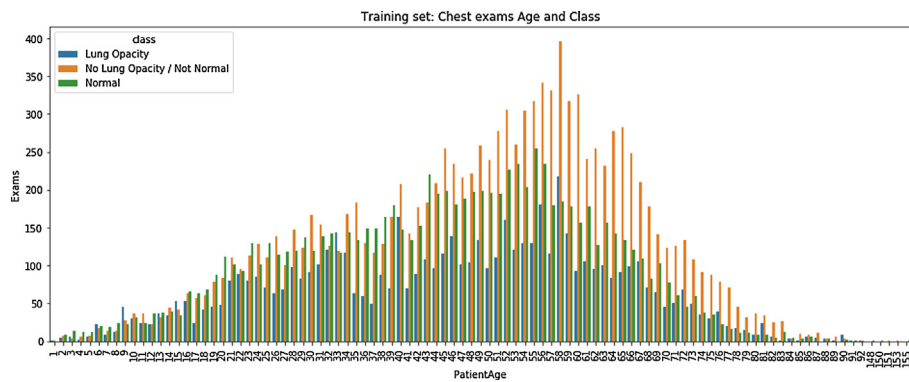**Fig. 2.** Positive and negative features among patients' of different age group.

Fig. 3. Feature class of patients' among different age group.

Table 4
List of labels.

Probability	Opacity	No opacity	Abnormal
High	Yes	No	No
Intermediate	Yes	No	No
Moderate	Yes	No	No

we scaled (cropping a little then resizing to the original size) the images.

We also classify the distribution of positional view features which is a radiographic view allied with the patient position given in training and test data as in Table 5:

Data cleaning: We have performed an extensive data cleaning on Stage 2 dataset and have explored the class probability ranking among males and females which is reported in Fig. 5 this shows that there are more chest X-ray images of males than females, both genders have the most classes of "No Lung Opacity/Not Normal", however other than this fact the men are more likely to have a class of "Lung Opacity" where as women are by proportion less likely. This clearly explains about the class probability ranking among the men and women.

4.2. Performance measures

We employ the mean of the intersection over union (IoU) of pairing ground truth bounding boxes and prediction at varied thresholds. The IoU can be computed from the paired threshold which is the region of the predicted bounding boxes and ground-truth bounding boxes as an evaluation metric for pneumonia identification task. It follows the below formula for IoU:
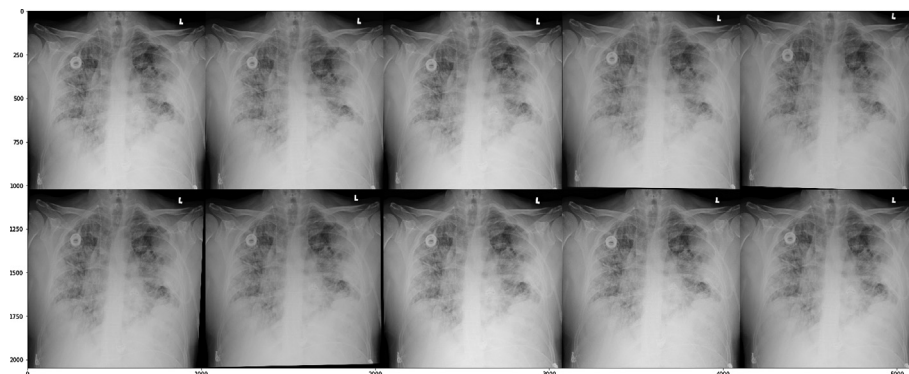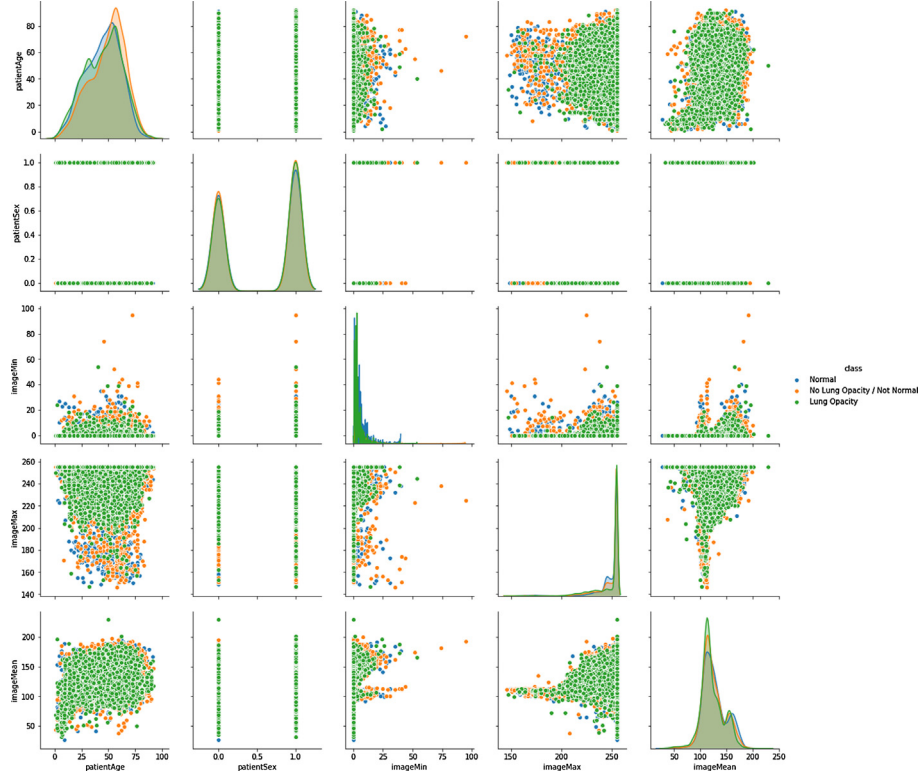

Fig. 4. Augmentation on chest X-ray images.

Table 5
Distribution of positional features in RSNA dataset.

Positional feature	Stage 1		Stage 2	
	Train Set	Test Set	Train Set	Test Set
Anterior/Posterior	20714	468	12173	1382
Posterior/Anterior	15161	532	14511	1618

Fig. 5. Distribution of patient age and sex.

$$IoU_{region}(B_{predicted}, B_{ground-truth}) = \frac{B_{predicted} \bigcap B_{ground-truth}}{B_{predicted} \bigcup B_{ground-truth}} \qquad (3)$$

The IoU determines a true-positive during pairing of predicted object with ground-truth object above the threshold which ranges from 0.4 to 0.75 at a step size of 0.05 to classify "misses" and "hits".[9] Pairing among predicted bounding boxes and ground-truth bounding boxes is assessed in descending order of the predictions and strictly injective which is based on their confidence levels.

Given any threshold value, the mean threshold value (MTV) over the outcomes for a particular threshold can be computed following the counts of true positives (c_{TP}), false negatives (c_{FN}), and false positives (c_{FP})

$$MTV(t) = \frac{c_{TP(t)}}{c_{TP(t)} + c_{FP(t)} + c_{FN(t)}} \qquad (4)$$

Also, we compute mean score (MS) for every image over all threshold values:

$$MS_i = \frac{1}{|Threshold|} \sum_t MTV(t) \qquad (5)$$

Therefore, we can compute the mean score for the dataset as follows:

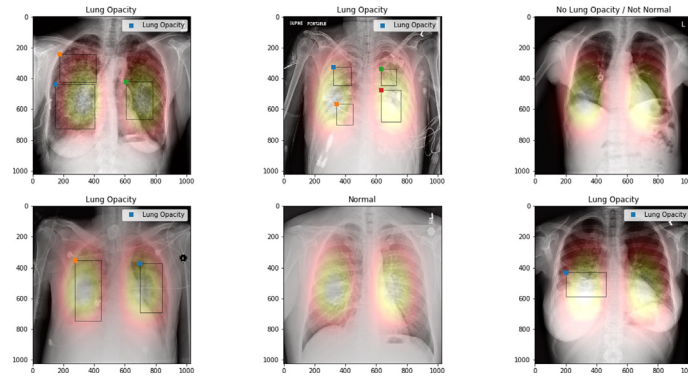$$MS_{dataset} = \frac{1}{|Image|} \sum_i^{Image} MS_i \qquad (6)$$

where *Image* in the dataset can be either a predicted bounding box or ground-truth bounding box.

5. Evaluation results

We report our prediction result in this section followed by results from ensemble model.

We perform ensembling in Stage 2 due to labelled dataset, whereas the dataset in Stage 1 was highly imbalanced. The variance in the dataset is due to radiologists are overlooked with reading high volumes of images every shift. We have discussed this in earlier section of this article. In Fig. 6, we overlay the probabilities of ground truth labels to check whether it is flipped or not. This also shows the successful predictions depicting inconsistency between ground-truth and prediction bounding boxes. We trained our proposed model in Stage 2 on NVIDIA Tesla P100 GPU and Tesla K80 in Stage 1, which also depicts that one needs an efficient computing resources to model such task on highly imbalanced dataset.

The prediction outcome of our model at given threshold is reported in Table 6, in which the best prediction set of bounding boxes and ground-truth boxes results in Stage 2. Also, the predicted sample set depicting pneumonia showing the position

---

[9] A predicted box hits when it reaches at a threshold of 0.5 provided its IoU with a ground-truth box is greater than 0.5.

Fig. 6. The results from stage 2 dataset. The probability overlaid on few images which includes all patient classes and labels. Green, orange, blue and red overlays shows predictions and ground truth labels, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6
Result: prediction at given threshold.

Model	Threshold	Stage 1	Stage 2
Mask-RCNN (ResNet50)	0.30	0.098189	0.183720
Mask-RCNN (ResNet101)	0.97	0.100155	**0.199352**

Table 7
Ensemble model results.

Model	Stage 2
Mask-RCNN (ResNet50 + ResNet101)	0.218051

(point) and the bounding box for each of the different image types is reported in Fig. 6. The position (point) and the bounding box for each of the different image types are reported in The training loss of our proposed model is reported in Fig. 7.

5.1. Ensembles

Our proposed approach, as illustrated in the beginning of this section, implies a typical ensemble model after post-processing step which is then employed to obtain prediction set of patients' having pneumonia. We ensembled our Mask-RCNN based model developed on ResNet50 and ResNet101 and the result is reported in Table 7.

6. Conclusion and future work

In this work, we have presented our approach for identifying pneumonia and understanding how the lung image size plays an important role for the model performance. We found that the distinction is quite subtle for images among presence or absence of pneumonia, large image can be more beneficial for deeper information. However, the computation cost also burden exponentially when dealing with large image. Our proposed architecture with regional context, such as Mask-RCNN, supplied extra context for generating accurate results. Also, using thresholds in background while training tuned our network to perform well in the this task.

With the usage of image augmentation, dropout and L2 regularization prevented the overfitting, but are obtained something weaker results on the training set with respect to the test. Our model can be improved by adding new layers, but this would introduce even more hyperparameters that should be adjusted. We intend to extend our model architecture in other areas of medical imaging with the usage of deep learning and computer vision techniques.
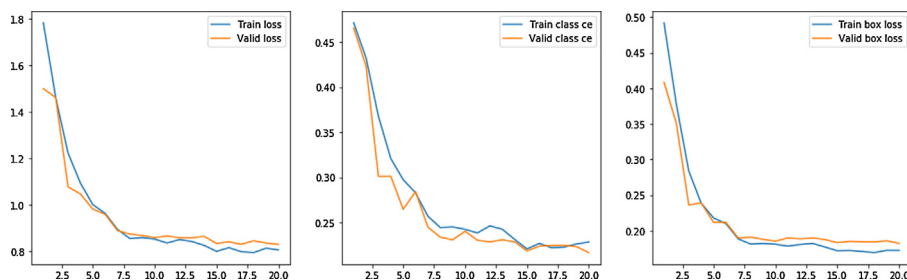
Fig. 7. Training loss: identification of pneumonia in Stage 2.

References

[1] U. Avni, H. Greenspan, E. Konen, M. Sharon, J. Goldberger, X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words, IEEE Trans. Med. Imaging 30 (3) (2011) 733–746.

[2] P. Pattrapisetwong, W. Chiracharit, Automatic lung segmentation in chest radiographs using shadow filter and multilevel thresholding, in: 2016 International Computer Science and Engineering Conference (ICSEC), IEEE, 2016, pp. 1–6.

[3] S. Katsuragawa, K. Doi, Computer-aided diagnosis in chest radiography, Comput. Med. Imaging Graph. 31 (4–5) (2007) 212–223.

[4] Q. Li, R.M. Nishikawa, Computer-aided Detection and Diagnosis in Medical Imaging, Taylor & Francis, 2015.

[5] C. Qin, D. Yao, Y. Shi, Z. Song, Computer-aided detection in chest radiography based on artificial intelligence: a survey, Biomed. Eng. Online 17 (1) (2018) 113.

[6] A.A. El-Solh, C.-B. Hsiao, S. Goodnough, J. Serghani, B.J. Grant, Predicting active pulmonary tuberculosis using an artificial neural network, Chest 116 (4) (1999) 968–973.

[7] O. Er, N. Yumusak, F. Temurtas, Chest diseases diagnosis using artificial neural networks, Expert Syst. Appl. 37 (12) (2010) 7648–7655.

[8] O. Er, C. Sertkaya, F. Temurtas, A.C. Tanrikulu, A comparative study on chronic obstructive pulmonary and pneumonia diseases diagnosis using neural networks and artificial immune system, J. Med. Syst. 33 (6) (2009) 485–492.

[9] S. Khobragade, A. Tiwari, C. Patil, V. Narke, Automatic detection of major lung diseases using chest radiographs and classification by feed-forward artificial neural network, in: 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), IEEE, 2016, pp. 1–5.

[10] G.G.R. Schramek, D. Stoevesandt, A. Reising, J.T. Kielstein, M. Hiss, H. Kielstein, Imaging in anatomy: a comparison of imaging techniques in embalmed human cadavers, BMC Med. Educ. 13 (1) (2013) 143.

[11] J. Li, Z. Liang, S. Wang, Z. Wang, X. Zhang, X. Hu, K. Wang, Q. He, J. Bai, Study on the pathological and biomedical characteristics of spinal cord injury by confocal raman microspectral imaging, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 210 (2019) 148–158.

[12] D.J. Winkel, T. Heye, T.J. Weikert, D.T. Boll, B. Stieltjes, Evaluation of an ai-based detection software for acute findings in abdominal computed tomography scans: toward an automated work list prioritization of routine ct examinations, Invest. Radiol. 54 (1) (2019) 55–59.

[13] H.R. Roth, L. Lu, A. Seff, K.M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, R.M. Summers, A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations, in: International Conference on Medical Image Computing and Computer-assisted Intervention, Springer, 2014, pp. 520–527.

[14] H.-C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R.M. Summers, Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imaging 35 (5) (2016), 1285–computing1298.

[15] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-assisted Intervention, Springer, 2015, pp. 234–241.

[16] A. Jamaludin, T. Kadir, A. Zisserman, Spinenet: automatically pinpointing classification evidence in spinal mris, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 166–175.

[17] K. Kallianos, J. Mongan, S. Antani, T. Henry, A. Taylor, J. Abuya, M. Kohli, How far have we come? artificial intelligence for chest radiograph interpretation, Clin. Radiol..

[18] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R.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.

[19] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, et al., Chexnet: radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv:1711.05225..

[20] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, et al., Chexpert: a large chest radiograph dataset with uncertainty labels and expert comparison. arXiv:1901.07031..

[21] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: towards real-time object detection with region proposal networks, Adv. Neural Inf. Process. Syst. (2015) 91–99.

[22] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.

[23] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2961–2969.

[24] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, et al., Speed/accuracy trade-offs for modern convolutional object detectors, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7310–7311.

[25] A. Shrivastava, R. Sukthankar, J. Malik, A. Gupta, Beyond skip connections: top-down modulation for object detection. arXiv:1612.06851..

[26] P. Luc, C. Couprie, Y. Lecun, J. Verbeek, Predicting future instance segmentation by forecasting convolutional features, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 584–599.

[27] A. Arnab, P.H. Torr, Pixelwise instance segmentation with a dynamically instantiated network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 441–450.

[28] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2980–2988.

[29] J. Redmon, A. Farhadi, Yolov3: an incremental improvement. arXiv:1804.02767..

[30] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[31] R. Girshick, Fast r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448.

[32] D.M. Hansell, A.A. Bankier, H. MacMahon, T.C. McLoud, N.L. Muller, J. Remy, Fleischner society: glossary of terms for thoracic imaging, Radiology 246 (3) (2008) 697–722.