Lung Disease Detection Using Deep Learning

Pranav Choudhary

Department of I.T., Maharaja Agrasen Institute of Technology, New Delhi, India

Abstract: Chest X-rays produce images of your heart, lungs, blood vessels, airways, and the bones of your chest. Chest X-rays can also reveal fluid in or around your lungs or air surrounding a lung. The X- ray images help the doctors/ radiologists to determine whether the patient is suffering from any chronic or acute lung disease such as lung nodule/mass, tuberculosis, aortic enlargement, cardiomegaly, pneumonia, pulmonary fibrosis etc. We use Convolution Neural Network (CNN) a deep learning technique which gained popularity due to its ability to learn mid and high level image representation. In this project various CNN models are used which detect different types of lung diseases through X-rays and if the lungs are in healthy shape, no findings are shown as a result . The performance of the project is tested on publicly available dataset.

KEYWORDS: Chest X-Ray, Deep Learning, Convolution Neural Network, Tuberculosis, aortic enlargement, cardiomegaly, pneumonia, pulmonary fibrosis.

DOI of the Article: https://doi.org/10.46501/IJMTST0706034

Available online at: http://www.ijmtst.com/vol7issue06.html

As per UGC guidelines an electronic bar code is provided to seure your paper

To Cite this Article: Pranav Choudhary. Lung Disease Detection Using Deep Learning. *International Journal for Modern Trends in Sceicen and Technology* 2021, *7*, 0706108, pp. 195-199. https://doi.org/10.46501/IJMTST0706034

Article Info. Received: 16 May 2021; Accepted: 6 June 2021; Published: 14 June 2021

INTRODUCTION

Lung illnesses, additionally called respiration diseases, are illnesses of the airlines and the opposite systems of the lungs [1]. Examples of lung ailment are pneumonia, tuberculosis and cardiomegaly, pulmonary fibrosis. According to the Forum of International Respiratory Societies [2], approximately 334 million humans be afflicted by asthma, and, every year, tuberculosis kills 1.4 million humans, 1.6 million humans die from lung cancer, even as pneumonia additionally kills tens of thousands and thousands of humans. The COVID-19 pandemic impacted the complete world [3], infecting tens of thousands and thousands of humans and burdening healthcare systems [4]. It is apparent that lung sicknesses are one of the main reasons for demise and incapacity in this world. Early detection plays a key role in increasing the chances of recovery and improving long-term survival rates [5,6]. Traditionally, lung disorder may be detected through pores and skin test, blood test, sputum pattern test [7], chest X-ray exam and computed tomography (CT) test exam [8]. Recently, deep learning has proven remarkable capacity while implemented on scientific photos for disorder detection, together with lung disorder.

Deep learning is a subpart of a far larger circle of relatives of device learning algorithms that makes use of many layers for function extraction and transformation. Data is handed via each layer and the output of one layer acts as an enter to the alternative layer, the primary layer is referred to as an input layer, and the remaining layer of the algorithm is referred to as an output layer. This technique is inspired by the structure and the function of the brain. In today's world it is applied in many fields such as natural language processing audio recognition, computer vision, speech recognition, social network filtering where its various architectures like deep neural networks, deep belief networks and recurrent neural networks have produced results which are comparable to that of human experts.

So, here we tend to propose a technique of detection of lung diseases by using Convolutional Neural Networks. Convolutional Neural Networks, additionally widely considered ConvNets, could be a Deep Learning approach that takes in an image in consideration as an input, assigns importance to numerous aspects of the image and it also has the power to differentiate one image from another simply. Temporal and spatial dependencies of an image are easily recognized by a Convolution Neural Networks through the employment of appropriate and applicable filters.

Structure of paper

The paper is organized as follows. Section 2 gives the information about related work of the project. Section 3 explains the methodology and dataset used. Section 4 presents and discusses the experimental results and finally Section 5 concludes the work done.

RELATED WORK

Computer Aided Diagnostic (CAD) systems have recently gained great success due to availability of large labelled datasets and advancement in supervised learning high-performing algorithms. With the help of Deep Convolutional Neural Networks the models can be trained to achieve expert level performance in detection of lung diseases. The performance of the CNN model depends on the high quality images from the dataset which is costly and time consuming to obtain but in the past few years many notable datasets for Chest X-Rays like ChestX-ray[8], ChestX-ray 14[10], Padchest[11], CheXpert3 , and MIMIC-CXR[12] . ChestX-ray14, an extended version of ChestX-ray8, was released by the US National Institutes of Health (NIH), containing over 112,000 CXR scans from more than 30,000 patients.

Most of existing CXR datasets depend on automated rule-based labelers that either use keyword matching (e.g. CheXpert[3] and NIH labelers[10]) or an NLP model (e.g. CheXbert[17]) to extract disease labels from free-text radiology reports. These tools can produce labels on a big scale but, at the same time, introduce an excessive rate of inconsistency, uncertainty, and errors[13], [18] . These noisy labels might also additionally cause the deviation of deep learning-primarily based totally on algorithms from stated performances while evaluated in a real-world setting[19] . In addition most of the researches are only report based, which does not identify the location of the disease in the lungs. There are few datasets which also provide the location of abnormality but they are either too small for training or not detailed enough.

The interpretation of x-rays is not only about image interpretation but also to localize the abnormality from the perspective of the radiologists. This also explains why use of CAD systems for X-rays is still very limited in clinical practices is still very limited

METHODOLOGY:

We developed a convolutional neural network to concurrently detect the presence of 14 different lung pathologies including pneumonia, pleural effusion, pulmonary masses, and nodules in frontal-view chest radiographs. The model was trained and internally validated on the dataset, with a held-out validation set with each of the original pathology labels. On this validation set, cardiothoracic specialist radiologists served as reference standard. We compared the model's discriminative performance on the validation set to the performance of 18 radiologists. We have used the 'timm' library which supports the model we use, resnet 18.

Resnet18 is a 18 layer deep convolutional neural network, it's pretrained network can classify images into 1000 object categories, therefore it learns rich feature representation for a wide range of images.

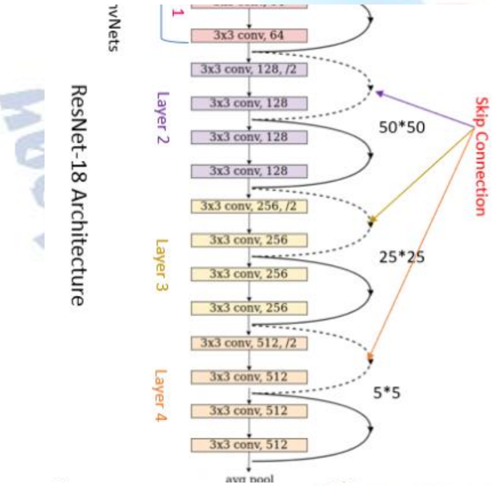


Figure 2: ResNet 18 Architecture

Dataset: All imaging information and also the corresponding ground truth labels for the training set only. The pictures were organized into 2 folders, one for training and a different one for testing. Every image features a unique, anonymous symbol that was encoded from the worth of the SOP Instance UID provided by the DICOM tag (0008,0018). The coding method was supported by the Python hashlib module . The medical specialists' native annotations of the training set were provided during a CSV file, annotations_train.csv. every row of the table represents a bounding box with the subsequent attributes: image ID (image_id), radiologist ID (rad_id), label's name (class_name), and bounding box coordinates (x_min, y_min, x_max, y_max). Here, rad_id encodes the identities of the seventeen radiology medical specialists, (x_min, y_min) are the coordinates of the box's higher left corner, and (x_max, y_max) are the coordinates of the lower-right corner. Meanwhile, the image-level labels were kept in several CSV files, image_labels_train.csv, with the subsequent fields: Image ID (image_id), radiologist ID (rad_ID), and international labels (labels). Specifically, every image ID goes with a vector of multiple labels equivalent to different pathologies, within which positive ones were encoded with "1" and negative ones were encoded with "0".

	image_id	class_name	class_id	rad_id	x_min	y_min	x_max	y_max
0	50a418190bc3fb1ef1633bf9678929b3	No finding	14	R11	NaN	NaN	NaN	NaN
1	21a10246a5ec7af151081d0cd6d65dc9	No finding	14	R7	NaN	NaN	NaN	NaN
2	9a5094b2563a1ef3ff50dc5c7ff71345	Cardiomegaly	3	R10	691.0	1375.0	1653.0	1831.0
3	051132a778e61a86eb147c7c6f564dfe	Aortic enlargement	0	R10	1264.0	743.0	1611.0	1019.0
4	063319de25ce7edb9b1c6b8881290140	No finding	14	R10	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...
67909	936fd5cff1c058d39817a08f58b72cae	No finding	14	R1	NaN	NaN	NaN	NaN
67910	ca7e72954550eeb610fe22bf0244b7fa	No finding	14	R1	NaN	NaN	NaN	NaN
67911	aa17d5312a0fb4a2939436abca7f9579	No finding	14	R8	NaN	NaN	NaN	NaN
67912	4b56bc6d22b192f075f13231419dfcc8	Cardiomegaly	3	R8	771.0	979.0	1680.0	1311.0
67913	5e272e3adbdaafb07a7e84a9e62b1a4c	No finding	14	R16	NaN	NaN	NaN	NaN

Figure 1: presentation of dataset

RESULTS:

We located that the model executed radiologist-level overall performance on eleven pathologies and did not acquire radiologist-level performance on three pathologies. The radiologists achieved statistically considerably better overall performance on cardiomegaly, calcification, and lung consolidation, model achieved better than radiologists in detecting atelectasis, there had been no statistically extensive variations in for the other 10 pathologies.

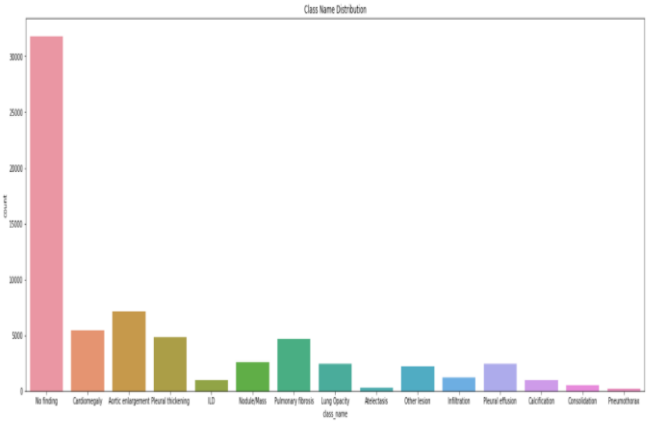


Figure 3: results predicted by radiologists

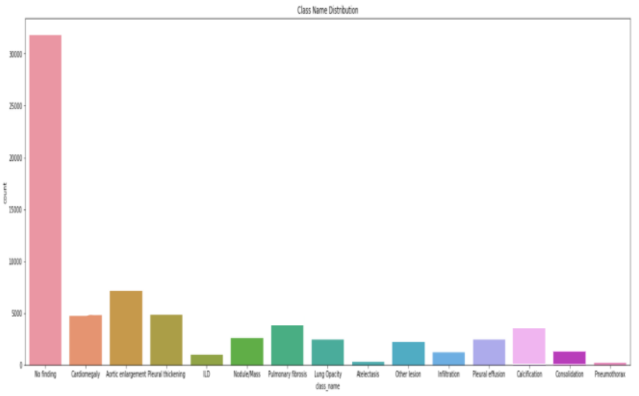
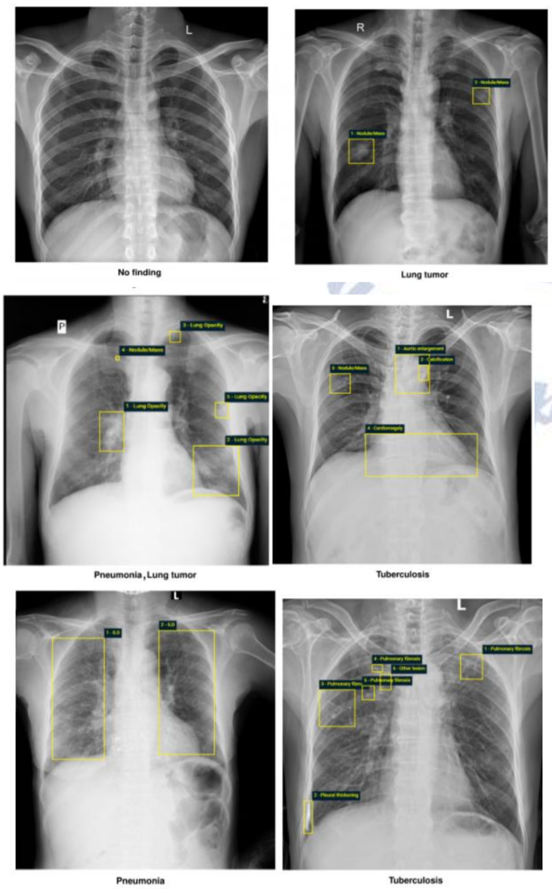


Figure 4: results predicted by model

Following are some resultant outputs along with labels assigned by radiologists



CONCLUSION:

From this study we have implemented Deep Learning using a Convolutional Neural networks model, ResNet 18 on the dataset of anterior chest x-rays to predict the lung disease and compared the results to that of experienced radiologists who have examined and labelled the x-rays in the dataset. Finally we have used deep learning architecture through convolutional neural networks to create computer aided diagnostic (CAD) systems.

ACKNOWLEDGMENT

I would like to acknowledge the support from my institute 'Maharaja Institute of Technology' and my mentors' and teachers' constant encouragement and guidance .

REFERENCES

1. Rajpurkar, P. et al. CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv preprint arXiv:1711.05225 (2017).
2. Rajpurkar, P. et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Medicine 15, e1002686, https://doi.org/10.1371/journal.pmed.1002686 (2018).
3. Irvin, J. et al. CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 590–597 (2019).
4. Majkowska, A. et al. Chest radiograph interpretation with deep learning models: Assessment with radiologistadjudicated reference standards and population-adjusted evaluation. Radiology 294, 421–431, https://doi.org/10.1148/ radiol.2019191293 (2020).
5. Rajpurkar, P. et al. CheXpedition: Investigating generalization challenges for translation of chest X-ray algorithms to the clinical setting. arXiv preprint arXiv:2002.11379 (2020).
6. Tang, Y.-X. et al. Automated abnormality classification of chest radiographs using deep convolutional neural networks. npj Digit. Medicine 3, 1–8, https://doi.org/10.1038/s41746-020-0273-z (2020).
7. Pham, H. H., Le, T. T., Tran, D. Q., Ngo, D. T. & Nguyen, H. Q. Interpreting chest X-rays via CNNs that exploit hierarchical disease dependencies and uncertainty labels. arXiv preprint arXiv:1911.06475 (2020).
8. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature 512, 436–444, https://doi.org/10.1038/nature14539 (2015).
9. Razzak, M. I., Naz, S. & Zaib, A. Deep learning for medical image processing: Overview, challenges and the

future. In Classification in BioApps, 323–350, https://doi.org/10.1007/978-3-319-65981-7_12 (Springer, 2018).

10. Wang, X. et al. ChestX-ray8: Hospital-scale chest X-ray database in Computer Vision and Pattern Recognition (CVPR), 2097–2106, https://doi.org/10.1109/CVPR.2017.369 (2017).
11. Bustos, A., Pertusa, A., Salinas, J.-M. & de la Iglesia-Vayá, M. Padchest: A large chest X-ray image dataset with multi-label annotated reports. arXiv preprint arXiv:1901.07441 (2019).
12. Johnson, A. E. et al. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. Sci. Data 6, 317, https://doi.org/10.1038/s41597-019-0322-0 (2019).
13. Oakden-Rayner, L. Exploring the ChestXray14 dataset: problems. https://lukeoakdenrayner.wordpress.com/2017/12/18/ the-chestxray14-dataset-problems/ (2017). (Online; accessed 04 May 2020). 7/10
14. Shiraishi, J. et al. Development of a digital image database for chest radiographs with and without a lung nodule: Am. J. Roentgenol. 174, 71–74, https://doi.org/10.2214/ajr.174.1.1740071 (2000).
15. Demner-Fushman, D. et al. Preparing a collection of radiology examinations for distribution and retrieval. J. Am. Med. Informatics Assoc. 23, 304–310, https://doi.org/10.1093/jamia/ocv080 (2016).
16. Jaeger, S. et al. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. Quant. Imaging Medicine Surg. 4, 475–477, https://dx.doi.org/10.3978%2Fj.issn.2223-4292.2014.11.20 (2014).
17. Smit, A. et al. CheXbert: Combining automatic labelers and expert annotations for accurate radiology report labeling using BERT. arXiv preprint arXiv:2004.09167 (2020).
18. Oakden-Rayner, L. Exploring large-scale public medical image datasets. Acad. Radiol. 27, 106 – 112, https://doi.org/10. 1016/j.acra.2019.10.006 (2020). Special Issue: Artificial Intelligence.
19. Nagendran, M. et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. BMJ 368, https://doi.org/10.1136/bmj.m689 (2020).
20. Vietnamese National Assembly. Regulation 40/2009/QH12 (Law on Medical Examination and Treatment). http://vbpl.vn/ hanoi/Pages/vbpqen-toanvan.aspx?ItemID=10482 (2009). (Online; accessed 11 December 2020).
21. Isola, S. & Al Khalili, Y. Protected Health Information (PHI). https://www.ncbi.nlm.nih.gov/books/NBK553131/ (2019).
22. US Department of Health and Human Services. Summary of the HIPAA privacy rule. https://www.hhs.gov/hipaa/ for-professionals/privacy/laws-regulations/index.html (2003).
23. European Parliament and Council of European Union. Regulation (EU) 2016/679 (General Data Protection Regulation). https://gdpr-info.eu/ (2016). (Online; accessed 11 December 2020).
24. Gao, X.W.; James-reynolds, C.; Currie, E. Analysis of tuberculosis severity levels from CT pulmonary images based on enhanced residual deep learning architecture. Neurocomputing 2019, 392, 233–244.
25. Gozes, O.; Frid, M.; Greenspan, H.; Patrick, D. Rapid AI Development Cycle for the Coronavirus (COVID-19) Pandemic: Patient Monitoring using Deep Learning CT Image Analysis Article. arXiv 2020, arXiv:2003.05037.