IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 6.252

Volume 2, Issue 1, November 2022

Chest Disease Detection and Classification

Prof. Ravi Rai Chaudhary[1], Ulhas Bhalerao[2], Amiy Singh[3], Aniket Shetye[4], Yash Shinde[5]

Professor, Department of Computer Engineering[1]

Students, Department of Computer Engineering[2,3,4,5]

SKN Sinhgad Insitute of Technology & Science, Lonavala, Maharashtra, India

Abstract: *Chest diseases & conditions such as Atelectasis, Cardiomegaly, Lung consolidation, Hernia, and Fibrosis becoming increasingly prevalent in the Asia-Pacific region. The Asia-Pacific Burden of Respiratory Diseases study examined the disease and economic burden of lung diseases across the Asia-Pacific and more specifically India. The objective is to use a deep learning model to diagnose pathologies from Chest X-Rays. ML approaches on CT and Xray images aided incorrectly in identifying lung diseases. Respiratory diseases range from mild and self-limiting, such as the common cold, influenza, and pharyngitis to life-threatening diseases such as bacterial pneumonia, pleural thickening, hernia, and severe acute respiratory syndromes, such as COVID-19. Authorities & Doctors will be able to deal with the effects more efficiently if such illnesses can be detected speedily and accurately with little human intervention in the future. In addition, various additional elements, such as environmental influences and commonalities among the most afflicted places, should be considered to slow the spread of lung diseases and precautions should be taken. Chest X-ray exam is one of the most frequent and cost-effective medical imaging examination. However clinical diagnosis of chest X-ray can be challenging, and sometimes believed to be harder than diagnosis via chest CT imaging. Even some promising work have been reported in the past, and especially in recent deep learning work on Tuberculosis (TB) classification. To achieve clinically relevant computer-aided detection and diagnosis (CAD) in real world medical sites on all data settings of chest X-rays is still very difficult, if not impossible when only several thousands of images are employed for study. This is evident from [2] where the performance deep neural networks for thorax disease recognition is severely limited by the availability of only 4143 frontal view images [3].*

Keywords: Lung disease, respiratory disease detection, Deep Learning, Disease Classification, Machine Learning

I. INTRODUCTION

In this database, we provide an enhanced version (with 6 more disease categories and more images as well) of the dataset used in the recent work [1] which is approximately 27 times of the number of frontal chest x-ray images in [3]. Our dataset is extracted from the clinical PACS database at National Institutes of Health Clinical Centre and consists of ~60% of all frontal chest x-rays in the hospital. Therefore, we expect this dataset is significantly more representative to the real patient population distributions and realistic clinical diagnosis challenges, than any previous chest x-ray datasets. Of course, the size of our dataset, in terms of the total numbers of images and thorax disease frequencies, would better facilitate deep neural network training [2]. Refer to [1] on the details of how the dataset is extracted and image labels are mined through natural language processing (NLP). Several successful machine learning algorithms have been developed in recent years and now the error rate has become very small with deep learning algorithms. Recently, machine learning, particularly in computer vision and speech recognition, almost approaches human perception level. Even if expert systems are used in practice in clinical settings, machine learning systems are still being used more experimentally today [4]. To diagnose or classify anything, patterns have to be identified. However, if the data we have is too large, it is hard to find these patterns. In addition, traditional methods cannot be used to find patterns or create mathematical models because gathered data is generally not linear [5].

In this study, our aim is to classify chest diseases, supported by the collected information. This information can include patient demographic info, preliminary queries, symptoms, respiratory organ operation take a look at results, biopsy results, X-ray results and final designation by chest physicians. Our results may embody connections between on the face of it unrelated information and conditions which can facilitate the sector of drugs with new insights [4].

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 6.252
Volume 2, Issue 1, November 2022

II. LITERATURE SURVEY

Inductive learning systems have been used in different medical fields such as oncology, liver pathology, prognosis of the survival in hepatitis, urology, diagnosis of thyroid diseases, rheumatology, diagnosing craniostenosis syndrome, dermatoglyphic diagnosis, cardiology, neuropsychology, gynecology, and perinatology. Automatically created diagnostic rules have increased the diagnostic correctness of specialist doctors [7]. Several studies have been reported that demonstrate the benefit of computerized lung disease analysis [8], [9], [10]. ML approaches on CT and X-ray images aided incorrectly in identifying COVID-19 [6].

III. METHODOLOGY

ChestX-ray dataset comprises 112,120 frontal-view X-ray images of 30,805 unique patients with the text-mined fourteen disease image labels (where each image can have multi-labels), mined from the associated radiological reports using natural language processing. Fourteen common thoracic pathologies include Atelectasis, Consolidation, Infiltration, Pneumothorax, Edema, Emphysema, Fibrosis, Effusion, Pneumonia, Pleural_thickening, Cardiomegaly, Nodule, Mass and Hernia, which is an extension of the 8 common disease patterns listed in our CVPR2017 paper. Note that original radiology reports (associated with these chest x-ray studies) are not meant to be publicly shared for many reasons. The text-mined disease labels are expected to have accuracy >90%.

3.1 Contents

1. 112,120 frontal-view chest X-ray PNG images in 1024*1024 resolution (under images folder)
2. Meta data for all images (Data_Entry_2017.csv): Image Index, Finding Labels, Follow-up #, Patient ID, Patient Age, Patient Gender, View Position, Original Image Size and Original Image Pixel Spacing.
3. Bounding boxes for ~1000 images (BBox_List_2017.csv):Image Index, Finding Label, Bbox[x, y, w, h]. [x y] are coordinates of each box's top left corner. [w h] represent the width and height of each box.
4. Two data split files (train_val_list.txt and test_list.txt) are provided. Images in the ChestX-ray dataset are divided into these two sets on the patient level. All studies from the same patient will only appear in either training/validation or testing set.

Distributions of 14 disease categories with co-occurrence statistics:

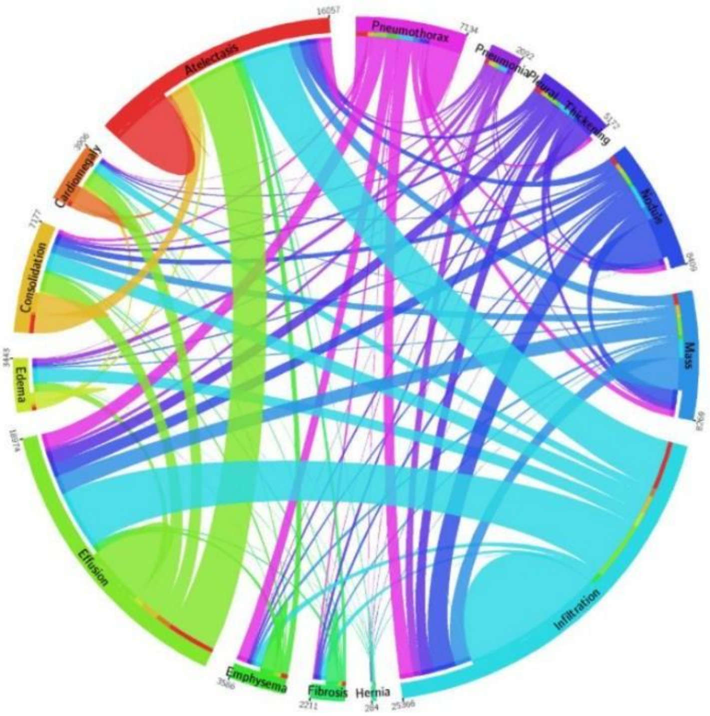


Figure 1: Distributions of 14 disease categories with co-occurrence statistics

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, November 2022
Impact Factor: 6.252

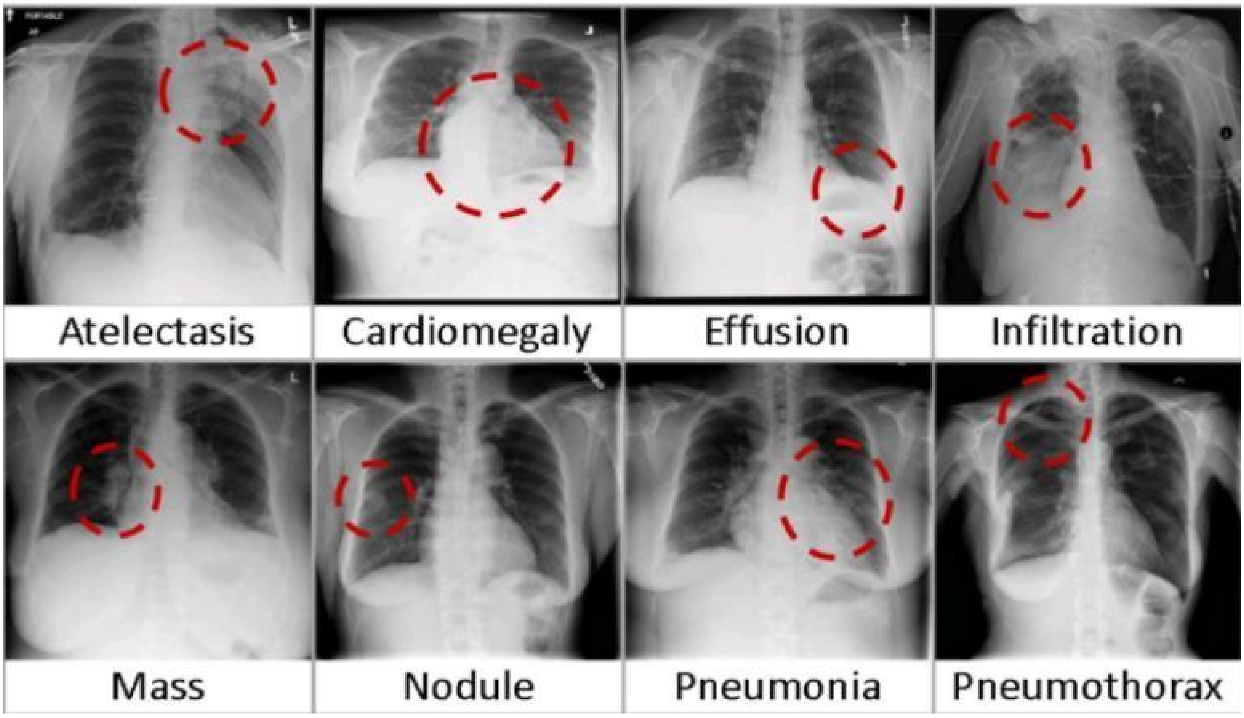
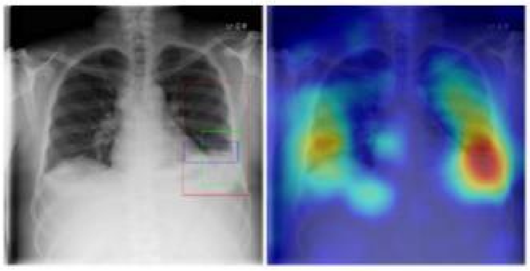


Figure 2: Eight visual examples of common thorax diseases

Radiology report	Keyword	Localization Result
findings include: 1. left basilar atelectasis/consolidation. 2. prominent hilum (mediastinal adenopathy). 3. left pic catheter (tip in atriocaval junction). 4. stable, normal appearing cardiomediastinal silhouette. impression: small right pleural effusion otherwise stable abnormal study including left basilar infiltrate/atelectasis, prominent hilum, and position of left pic catheter (tip atriocaval junction).	Effusion; Infiltration; Atelectasis	

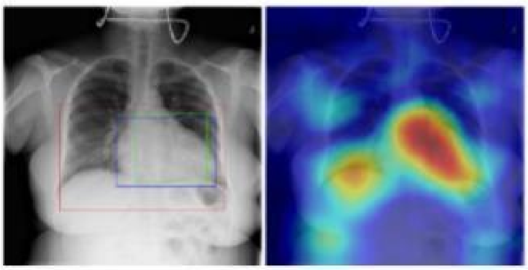
Radiology report	Keyword	Localization Result
findings include: 1. cardiomegaly (ct ratio of 17/30). 2. otherwise normal lungs and mediastinal contours. 3. no evidence of focal bone lesion. dictating	Cardiomegaly	

Figure 3: Two Samples of disease localization using weakly supervised deep neural networks

IV. CONCLUSION

In this study, our first goal was to build an ML model that can detect & classify lung diseases and patient information to a computer. Image datasets imported were used by this system in machine learning algorithms for purposes of automated analysis and diagnosis. Algorithms were run on multiple datasets to classify patients into:

1. Ill or healthy with x-ray images,
2. 14 diseases with text data,

Some classes are represented by just a few samples. Therefore, the classification accuracy drops as we have more classes. Also, the total number of samples affects the classification results. We have enough samples to classify 2 classes but for more accurate classification of more classes we need more samples. In 12 class classification of lung diseases. However, in deciding if the patient is healthy or sick, raw x-ray data can also be used as we found it to be highly accurate as well. In addition, this system will enable to record and store patient information, especially scans, to be shared with other physicians and to compare the new data recorded later to follow the prognosis of the patient. We believe that our method of diagnosis classification using patient data and respiratory sounds can lead the way for even more advanced computerized analysis techniques in the future..

V. ACKNOWLEDGMENT

The authors would like to acknowledge the support and guidance provided by management and guides of SKN Sinhgad Institute of Technology and Science, Lonavala for providing the necessary support and guidance in carrying out this work. This system works for classification of up to 14 chest diseases. This system provides accurate & valid results. This will help identify multiple lung diseases and ease workloads on doctors.

REFERENCES

[1]. Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, MohammadhadiBagheri, Ronald Summers, ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common ThoraxDiseases, IEEE CVPR, pp. 3462-3471,2017

[2]. Hoo-chang Shin, Kirk Roberts, Le Lu, Dina Demner-Fushman, Jianhua Yao, Ronald M. Summers, Learning to Read Chest X-Rays: Recurrent Neural CascadeModel for Automated Image Annotation, IEEE CVPR, pp. 2497-2506, 2016

[3]. Open-i: An open access biomedical search engine. https://openi.nlm.nih.gov

[4]. Murat Aykanat a,* , Özkan Kılıç a , Bahar Kurt,b , Sevgi Saryal c: Lung disease classification using machine learning algorithms DOI:10.18100/ijamec.799363

[5]. M. Aykanat, Ö. Kılıç, B. Kurt, and S. Saryal, "Classification of lung sounds using convolutional neural networks", EURASIP Journal on Image and Video Processing, vol. 65, pp. 1-9, 2017. DOI: 10.1186/s13640-017-0213-2.

[6]. Surbhi Gupta , Mohammad Shabaz , Sonali Vyas : Artificial intelligence and IoT based prediction of Covid-19 using chest X-ray images. PMID: 35783463 PMCID: PMC9233885 DOI: 10.1016/j.smhl.2022.100299Surbhi Gupta 1, Mohammad Shabaz 1, Sonali Vyas

[7]. I. Kononenko, "Inductive and Bayesian learning in medical diagnosis", Applied Artificial Intelligence, vol. 7, no. 4, pp. 317-337, 1993. DOI:10.1080/08839519308949993

[8]. B. Flietstra, N. Markuzon, A. Vyshedskiy, and R. Murphy, "Automated analysis of crackles in patients with interstitial pulmonary fibrosis", Pulm Med., vol. 2010, pp. 1-7, 2011.

[9]. H. Pasterkamp, S. S. Kraman, and G. R. Wodicka, "Respiratory sounds, advances beyond the stethoscope", Am J Respir Crit Care Med, vol. 156, pp. 974–987, 1997.

[10]. J. E. Earis, and B. M. G. Cheetham, "Current methods used for computerized respiratory sound analysis", Eur Respir Rev., vol. 10, no. 77, pp. 586–590, 2000.