Chest X-Ray based Lung Disease Detection using Convolutional Neural Network

Saravanan. M [a], Manoj Kumar [a], Dilsha Vijay [a], Gayathri. K [a] and Jenifar. A [a]

ª KPR Institute of Engineering and Technology, Coimbatore, India

Abstract

Humans suffer from a variety of health problems associated with their chests. There are several diseases associated with continual cardiomegaly, emphysema, fibrosis, pneumothorax, infiltration and other lung sickness. Diagnosing chest conditions as soon as possible is essential. As there are many methods, we analyze the problem of medical data scarcity in this paper using a set of datasets for detect and classify the lung diseases from chest radiograph images. We implemented convolutional neural networks methods to train the images. We collected the data set manually from different websites for 13 diseases with a set of nearly 1000 images. It helps the person to identify the diseases individually without the help of an expert. We got an accuracy of 97% using this algorithm, each disease accuracy is recorded individually.

Keywords

Lung disease detection, deep learning, CNN, Neural network, pre processing techniques

1. Introduction

There are various respiratory diseases that can affect the lungs. One of these is pneumonia, which kills about 1.6 million people annually. In addition to that tuberculosis, pneumothorax and countless others are a threat to human beings. It is estimated that lung diseases are responsible for the deaths of around 3 million people annually. Traditionally, an individual can be diagnosed with lung disease through various tests, such as a blood test and a chest X-ray examination. Pleural effusions (PE) are fluid buildups in the pleural cavity that are frequently a sign of a more serious illness such heart problems, pneumonia, or colon cancers. They've also been discovered to be prognostic indications, such as in the case of acute pancreatitis. Pneumothorax is a pleural illness that causes air to collect in the pleural space. Because air is less thick than lung parenchyma, the pneumothorax region will take on the structure of the lungs and lung cavity, occupying the upper portions of the lungs. Pulmonary fibrosis is a lung condition caused by scarring and damage to lung tissue. It's more difficult for your lungs to perform properly because of this thicker, rigid tissue. As your pulmonary fibrosis progresses, you will become increasingly breathless. When your airways or the little sacs at the end of them don't expand as they ought to when you breathe, you get atelectasis. A lung nodule is a tiny irregular spot that can be discovered during a chest CT scan. These scans are performed for a variety of purposes, including lung cancer screening and checking the lungs if you have symptoms. The majority of lung nodules detected on CT scans are not cancerous. They are more commonly caused by previous infections, scar tissue, or other factors. Cardiomegaly is a term used to describe the expansion of the heart, which is usually caused by a cardiac problem. Cardiomegaly can be caused by a number of disorders that impact how the heart works, including high blood pressure,

diabetes, and obesity. Shortness of breath is a symptom of emphysema, a lung disease. Persons who suffer from emphysema have compromised air sacs in the lungs. Over time, the inner sacs in the lungs weaken and tear, resulting in larger air spaces rather than small individual ones. Emphysema can go undetected for many years. Pleural thickening is a chronic condition wherein scar tissue thickens the pleural lung tissue, commonly known as pleura. For doctors, classifying chest X-ray abnormalities of these many kinds of lung diseases is a time-consuming operation; as a result, various algorithms have been proposed to effectively accomplish this work. Over the years, computer-aided diagnostic tools have been developed to capture significant information of X-rays to assist doctors in acquiring a thorough knowledge of the X-ray. On the other hand, such CAD system may not have reached a considerable degree of importance for making diagnoses in X-rays. As a result, their role has been confined to providing visualizing functionality to clinicians to aid in decision-making.

Patterns must always be identified in order to diagnose or categorize things. However, finding these connections can be difficult if the dataset we have is just too vast. Furthermore, since obtained data is rarely linear, conventional methods cannot be used to discover patterns or develop models. Many effective machine learning algorithms recently emerged, and deep learning techniques now have a low error rate. The images utilized to train this model were atelectasis, cardiomegaly, effusion, infiltration, nodule, pneumonia, pneumothorax, emphysema, fibrosis, pleural thickening, and no finding. This research presents a typology of practical applications for lung disorders as well as a market analysis on the subject. The remaining difficulties are also discussed, as well as prospective future directions.

2. Existing Work and System

The Existing System uses the K-NN algorithm and CNN but in most cases, they are using the CT scan images and detection of the early stages of the lung diseases. The existing System has its own advantages and disadvantages but most important disadvantages is that they are not trained enough to classify the real time images. To overcome this, we can use the deep learning techniques to increase the accuracy of the model to produce more precise output even when we use the real time dataset.
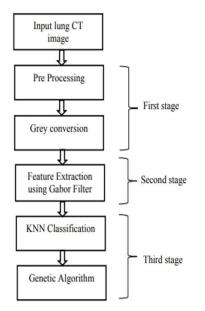

Figure 1: Architecture of existing system

3. Proposed System
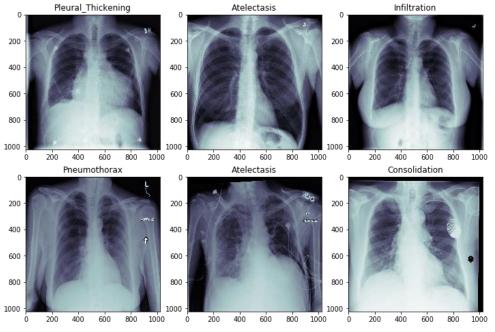
3.1. Data Collection

Figure 2: Sample Dataset

Data for the project was manually gathered from a variety of sources and cross-referenced with publicly available information. Because the initiative is centered on classifying different diseases. The datasets are sorted into folders and then trained separately. There are 900 images in all. Fig 1 shows the sample data. Fig 2 represents all the diseases and the amount of images taken for each category.
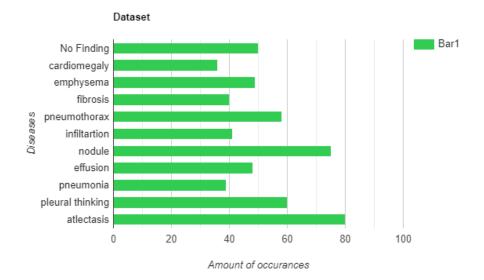

Figure 3: Dataset Visualization

3.2.   Pre-processing

By pre-processing the data, meaningful insights can be extracted from the data, thus improving the quality of the data. In Machine Learning, pre-processing refers to the process of preparing (cleaning and organizing) raw data for building and training Machine Learning algorithms. Here the data is processed in four steps. They are

- **Data quality assessment**

It is possible to receive data in a variety of formats when you collect data from different sources. you are likely to receive information in a variety of formats. For example, if we are collecting images in different websites then we need to change every image into single format.

- **Data cleaning**

As we have collected data from different sources, we have to remove unwanted information and and irrelevant data. It helps the data to run efficiently without any errors.

- **Data transformation**

We have already begun cleaning data; the data transformation will start changing the data into the proper format we have to download and use in other formats.

- **Data reduction**

As we are handling more data's, even after cleaning and changing it. We have enough data set than we need it. Data reduction makes the analysis more easier and most accurate.
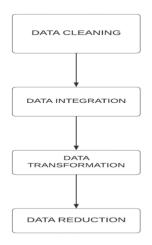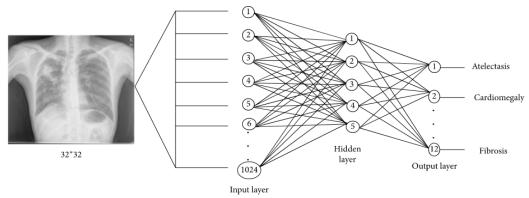
```
┌─────────────────────┐
│   DATA CLEANING     │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│  DATA INTEGRATION   │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│       DATA          │
│  TRANSFORMATION     │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│   DATA REDUCTION    │
└─────────────────────┘
```

Figure 4: Data Processing Methods

3.3.   Algorithm and Model

A classification algorithm is a quantitative process of mapping input data to a certain category using a classifier. Classifiers come in a variety of forms. One of them is Convolutional Neural Network. A convolution is a quantitative process that transforms one function into another and calculates the cumulative of their integer combination. It is intimately linked to the Laplace and Fourier transforms. Cross-volution's work in a similar fashion to convolutional layers. The first layer of a CNN is crucial since it connects the input image to the first layer's receptive fields. CNNs are the most widely used deep learning algorithm, and they are made up of brains with adaptable prejudices and parameters. Several inputs are received by each node. The sum of the inputs is then calculated. The sum is then fed into a

convolution operation, which generates an output. CNN differs from other neural networks because it includes several convolutional layers. When training, CNNs usually have two elements: feature extraction and classification. Convolution is applied to the input using a kernel during the feature extraction stage. Following that, a feature map is created. During the classification stage, the CNN calculates the likelihood that the image parts to a given class or label.

Figure 5: CNN model

The image has been converted to grayscale. After that, noise removal and contrast enhancement are completed to generate enhanced photos. CNN divides it into two categories: no findings and other labelled diseased lungs, and so it identifies lung diseases. The X-rays' small characteristics serve as a template for feeding the classifier. The part of the sickness that has been recognised is depicted in the diagram.
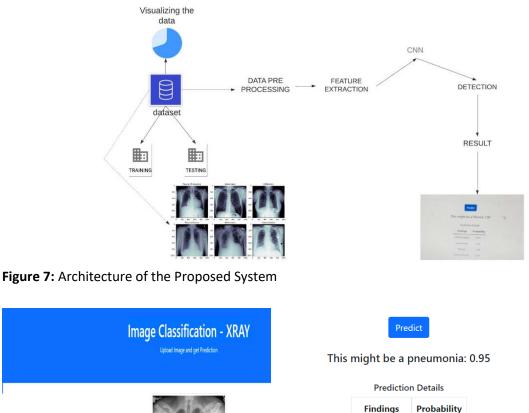

Figure 6: Disease identified in images

The dataset is first sufficiently separated into the train and test groups. For the purpose of visualizing how data is classified the python library matplotlib is used to visualize data. Pre-processing of data is a technique in data mining for converting unprocessed data into a consumable and practical format. There may be various insignificant and missing sections in the data. Information cleansing has been completed in order to deal with this section. Using the feature extraction methods, we can create new aspects that are a quadratic mixture of current features. The training method entails retrieving features from an image, which is repeated over numerous epochs. At least 10 photos must be processed for each epoch. As a result, the system may intelligently forecast a disease based on the labels. The training algorithm is CNN [Convolutional Neural Network] and the language used for creating the model is Python. This is a binary classification concept that requires extracting structural and physiological information from photos and

30

masks. The characteristics are linear and quantitative, although they can be divided into groups. We created a user interface using vanilla Javascript for easy access after training and testing the data.

4. Results and Comparison

Result of this model is to classify the given input as nofinding or the prediction of specific lung disease like cardiomegaly ,emphysema, fibrosis, pneumothorax, infiltration, nodule, effusion, pneumonia, pleural thinking ,atelectasis . The output is the prediction of the different lung diseases.The accuracy of the CNN model is 97 Percentage. We can also Predict the lung disease using the real time data so that it will be very useful for the civilians to know the lung diseases based on the X-Ray images itself.

Figure 7: Architecture of the Proposed System

5. Conclusion

In this paper, we present a method for detecting lung disease from lung X-ray pictures. We developed a lung disease categorization system based on deep learning algorithms and evaluated it on modest lung image datasets. We want to show that using a deep learning algorithm will help us acquire more precise

31

results. As a result, we've obtained a great level of accuracy. With the right feature selection strategy and unified methodology, this can be predicted.

6. References

[1] 16.Kingsley Kuan, Mathieu Ravaut, Gaurav Manek, Huiling Chen, Jie Lin,Babar Nazir, Cen Chen, Tse Chiang Howe, Zeng Zeng, Vijay Chandrasekhar. Deep Learning for Lung Cancer Detection: Tackling the Kaggle Data Science Bowl 2017 Challenge.

[2] Campos HS, Lemos ACM. Asthma and COPD according to the pulmonologist. Brazilian Journal of Pulmonology. 2009; 35(4):301-9.

[3] Roth HR, Lu L, Seff A, Cherry KM, Hoffman J, Wang S, Liu J, Turkbey E, Summers RM. A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations. Lecture Notes in Computer Science. 2014; 8673: 520–527.

[4] Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nat Med. 2018; 24:1559–1567. doi:10.1038/s41591-018-0177-5

[5] P. Pattrapisetwong and W. Chiracharit, "Automatic lung segmentation in chest radiographs using shadow filter and multilevel thresholding," in Proceedings of 2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Manchester, UK, 2016.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proceedings of Advances in Neural Information Processing Systems, pp. 1097–1105, Lake Tahoe, Nevada, USA, 2012.

[7] D. Manos, J. M. Seely, J. Taylor, J. Borgaonkar,H. C. Roberts, and J. R.Mayo, "The lung reporting and data system (LU-RADS): a proposal for computed tomography screening,"Canadian Association of Radiologists Journal, vol. 65,pp. 121–134, 2014.

[8] Hussein, S., Cao, K., Song, Q., Bagci, U.: Risk Stratification of Lung Nodules Using 3D CNN-Based Multi-task Learning. In: International Conference on Information Processing in Medical Imaging.

[9] Arnaud A. A. Setio, Francesco Ciompi, Geert Litjens, Paul Gerke, Colin Jacobs, Sarah J. van Riel, Mathi"Pulmonary nodule detection in CT images:false positive reduction using multi-view convolutional networks"

[10] Wang, L.; Lin, Z.Q.; Wong, A. COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images.

[11] Melendez, J.; Ginneken, B.V.; Maduskar, P.; Philipsen, R.H.H.M.; Reither, K.; Breuninger, M.; Adetifa, I.M.O.; Maane, R.; Ayles, H.; Sánchez, C.I. A Novel Multiple-Instance Learning-Based Approach to Computer-Aided Detection of Tuberculosis on Chest X-Rays. IEEE Trans. Med.

[12] Angeline R, Mrithika M, Raman A, Warrier P. Pneumonia detection and classification using chest X-ray images with convolutional neural network. In: Smys S, Iliyasu AM, Bestak R, Shi F, editors. New trends in computational vision and bio-inspired computing. ICCVBIC. Cham: Springer; 2020

[13] Ge Z, Mahapatra D, Chang X, Chen Z, Chi L, Lu H. Improving multi-label chest X-ray disease diagnosis by exploiting disease and health labels dependencies. Multimed Tools Appl. 2019

[14] Justin Johnson Andrej Karpathy Li Fei-Fei,DenseCap:Fully Convolutional Localization Networks for Dense Captioning, IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2016

[15] Andrew Ward, Nicholas Bambos. Quantum Annealing Assisted Deep Learning for Lung Cancer Detection. http://cs231n.stanford.edu/reports/2017/pdfs/534.pdf

[16] Alcantud, J.C.R., Varela, G., Santos-Buitrago, B., Santos-Garcia, G., Jimenez, M.F.: Analysis of survival for lung cancer resections cases with fuzzy and soft set theory in surgical decision making. PLoS ONE 14(6), e0218283 (2019)