

COMPARITIVE ANALYSIS OF LUNG DISEASE DETECTION USING DEEP LEARNING MODELS

G.V.GAYATHRI ^{#1}, B. SIVA JYOTHI ^{#2}

^{#1} Assistant Professor, ^{#2} Assistant Professor,
Department of Computer Science & Engineering,
Anil Neerukonda Institute of Technology & Sciences,
Sanghivalasa, Visakhapatnam, AP 531162.

ABSTRACT

Now a days for identifying or predict any diseases on human beings, we should have proper diagnosis for predicting the disease which is present in that human body. In general for prediction of diseases we try to use either X-Ray, CT or MRI scan techniques for taking decision on that appropriate disease. In general medical person need complete knowledge on that appropriate domain to find out the abnormality which is present in human beings. As we all know that India tops the world for having more deaths due to lung diseases. After the second highest cause of deaths in India due to heart disease, this lung disease is one which is increasing its rank more and more. In order to reduce that problem early diagnosis and treatment of lung diseases is critical to prevent complications including death. Normally for finding the abnormality present in lung, chest X-ray is playing very important role to detect the complete information about the lungs. In this current article we try to present an effective way for expert diagnosis of lung diseases using deep learning models. It focuses on creating a system for assistance of Radiologists in detection of lung diseases. This will especially benefit rural areas where radiologists aren't easily available. We use two models like Vgg16 and Vgg19 for predicting the lung disease from chest x ray images and then tell which model gives high accuracy and performance. We conclude by discussing research obstacles, emerging trends, and possible future directions for improving some more advancement.

Keywords:

Radiologists, Lung Diseases, Deep Learning Models, Early Diagnosis, X-Ray.

1. INTRODUCTION

In recent days, the introduction of IT and e-health care system in the medical field try to provide medical experts to give proper treatment for the patients who are in emergency. One of the most critical disease which is ranked second in India after the heart disease is lung diseases, also known as respiratory diseases [1]. Some of the diseases which come under respiratory disease are pneumonia, tuberculosis and

currently Coronavirus Disease 2019 (COVID-19). As per the IRS (International Respiratory Societies [2]), report more than three hundred million people are continuously suffering from asthma disease and more than 2 million people die due to this lung diseases. From the recent analysis, we know the COVID-19 pandemic infected millions of people and healthcare systems and also there was great loss for the humans. In general these lung diseases are major cause of death and create disaster for the world. Normally early detection of lung disease plays an major role in the chance of disease recovery and there are very few recovery rates if they are early detected and treated. In the primitive days the lung diseases are detected via blood test, skin test and some X-ray and CT scan. The report need to be examined by the radiology department and the concern person who has enough knowledge will try to tell the report from the test sample, which is becoming a very complex task if the radiologist is not available all the time. Recently deep learning has gained a lot of users attention towards medical domain for disease prediction and finding abnormality. Hence we try to use this deep learning technique on lung disease prediction and try to classify the abnormality which is present in the lungs using chest X-ray examination[3]-[8].

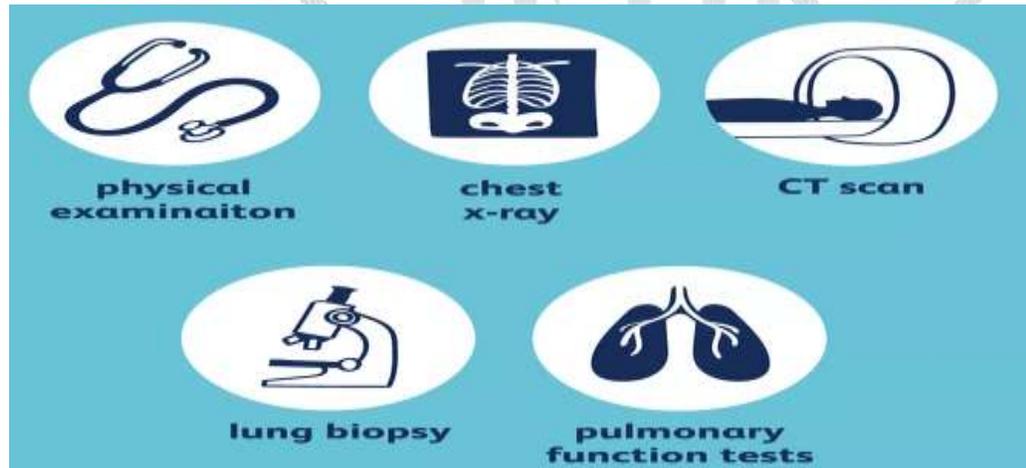


Figure 1. Represent the Several Methods for Detecting Lung Diseases

From the above figure 1, we can clearly identify several types of examinations are done for identifying the abnormality which is present in human lungs. In general we try to apply deeplearning in the field of medical domain to identify the pattern which is present in the chest X-ray and then try to derive the possible learned features from that image[9]. As we all know that deep learning is becoming state of the art by increasing its performance in huge number of medical applications which can assist the medical department persons or clinicians to detect and classify some minute medical abnormalities very effectively and efficiently[10]. There was a lot of research work undergone for the lung diseases detection and to the best of our knowledge we can see one survey paper which is published based on some previous published papers references on this topic [11]. If we look in this paper we can see all the history related to deep learning and how deep learning is integrated in the applications of pulmonary imaging. Major applications of deep learning techniques on several lung diseases, namely pulmonary nodule diseases, pulmonary embolism, pneumonia, and interstitial lung disease, are also described.

2. LITERATURE SURVEY

Literature survey is that the most vital step in the software development process. Before developing the new application or model, it's necessary to work out the time factor, economy, and company strength. Once all these factors are confirmed and got approval then we can start building the application. The literature survey is one that mainly deals with all the previous work which is done by several users and what are the advantages and limitations of those previous models. This literature survey is mainly used for identifying the list of resources to construct this proposed application.

MOTIVATION

1) A Survey of Deep Learning for Lung Disease Detection on Medical Images: State-of-the-Art, Taxonomy, Issues and Future Directions

AUTHORS: Stefanus Tao Hwa Kieu and Abdullah Bade

In this paper, the authors mainly concentrated on the lung disease and survey about deep learning for lung disease detection based on X-ray images. The authors mainly studies one survey paper published in the last five years regarding deep learning directed at lung diseases detection. In that paper they saw a taxonomy and analysis of the trend of recent work, but they saw some missing information present in that contents. So in order to improve the information related to this taxonomy they studies nearly 98 articles published from 2016 to 2020. From all these articles they gathered seven main attributes which are required for the lung disease detection such as: image types, features, data augmentation, types of deep learning algorithms, transfer learning, the ensemble of classifiers and types of lung diseases.

2) Deep Convolutional Neural Networks for Chest Diseases Detection

AUTHORS: Rahib H. Abiyev

In this paper, the authors mainly discussed about deep Convolutional Neural Networks for Chest Diseases Detection. In general the authors concentrated on the chest diseases as one of the complex health problem in day to day life. This lung disease includes chronic obstructive pulmonary disease, pneumonia, asthma, tuberculosis, and lung diseases. The authors said that timely diagnosis of chest diseases is very important for finding the abnormality which is present and cure of that lung disease. In this article the authors described that use of CNN model is very efficient for diagnosis of chest diseases and this can increase the chance of accuracy.

3) A Review of Deep-Learning-Based Medical Image Segmentation Methods.

AUTHORS: Xiangbin Liu and Liping Song

In this paper, the authors mainly discussed the review of several deep learning based Image segmentation methods. This paper focuses on the research of medical image segmentation based on deep learning. First, the basic ideas and characteristics of medical image segmentation based on deep learning are introduced. By explaining its research status and summarizing the three main methods of medical image segmentation and their own limitations, the future development direction is expanded[6]-[7].

3. EXISTING SYSTEM AND ITS LIMITATIONS

In the existing system, there was no concept like lung cancer prediction using CNN models. All the prediction is done using manual approach or by using primitive Machine Learning models. In the ML we can able to classify whether lung cancer is present or not, but those models cannot classify the records with accuracy and parameters.

LIMITATIONS OF THE EXISTING SYSTEM

1. All the existing schemes are limited to the few classes classification only.
2. All the existing systems are failed to classify the chest x-ray images and then try to find out the cancer symptoms.
3. All the existing ML approaches try to classify the patients information from the raw dataset
4. There is no accurate model to classify the real time chest x-ray for detecting and prediction of accuracy of that image.

4. PROPOSED SYSTEM AND ITS ADVANTAGES

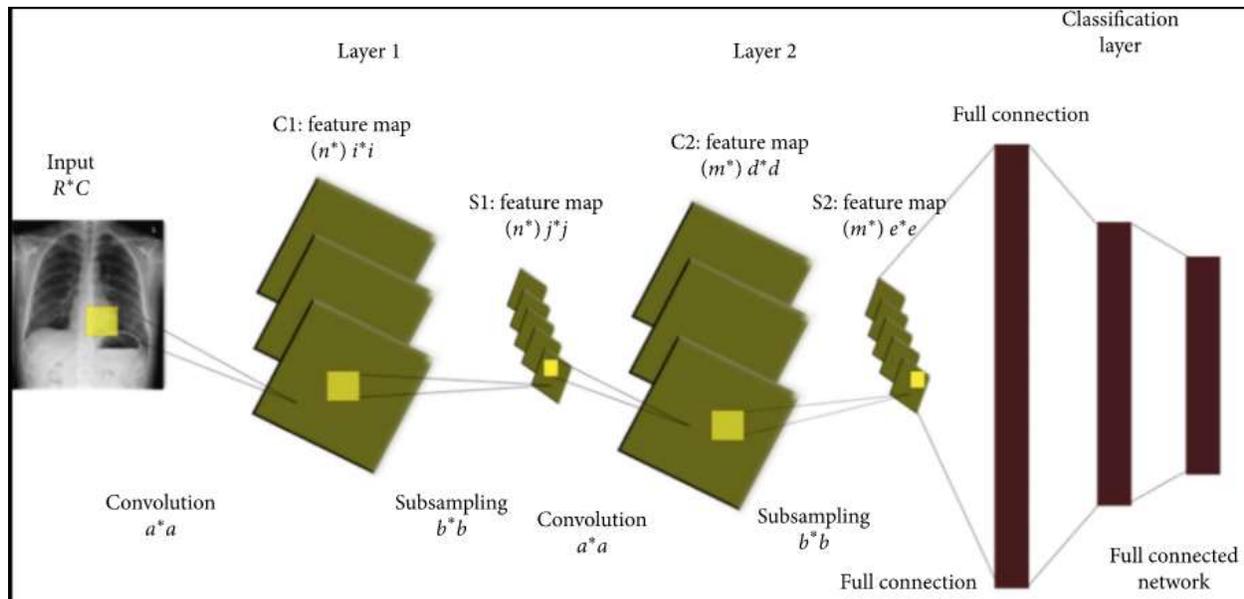
In this proposed work we try to design an application which can be used for prediction of lung cancer from real world chest x-ray images. For training the system we try to collect the sample chest X-ray images which contain cancer symptoms from KAGGLE website and then train the system. Once the system is trained now we can check the model performance by giving dynamic images and check the performance of each and every individual Model. Using Deep Learning to predict lung diseases from Chest X-rays can be a lifesaving factor for an individual suffering from the disease. This is possible as the results can be predicted with a high percentage of accuracy instantly. This paper presents an effective way for expert diagnosis of lung diseases using Deep Learning. It focuses on creating a system for assistance of Radiologists in detection of lung diseases. This will especially benefit rural areas where radiologists aren't easily available. We use two models like Vgg16 and Vgg19 for predicting the lung cancer from chest x ray images and then tell which model gives high accuracy and performance.

ADVANTAGES OF THE PROPOSED SYSTEM

1. The proposed scheme is very accurate in classification of chest x-ray images
2. The proposed system gives accurate recommendation for the doctors.
3. The proposed system is capable of classification of chest x-ray and find out the accuracy of that image.

5. PROPOSED CNN MODEL FOR CHEST DISEASE DETECTION

In this section we try to discuss about proposed CNN model which is used to detect chest disease detection using pre-trained CNN models such as VGG16 and VGG19.



The Application is mainly divided into 4 modules. They are as follows:

1. Convolution Layer
2. Rectified Linear Unit (RELU) Layer
3. Pooling Layer
4. Fully Connected layer

A) CONVOLUTION LAYER

A convolution is defined as an operation on two functions. In image analysis, one function consists of input values (e.g. pixel values) at a position in the image, and the second function is a filter (or kernel) each can be represented as array of numbers. Computing the dot product between the two functions gives an output. The filter is then shifted to the next position in the image as defined by the stride length. The computation is repeated until the entire image is covered, producing a feature (or activation) map. This is a map of where the filter is strongly activated and 'sees' a feature such as a straight line, a dot, or a curved edge. If a photograph of a face was fed into a CNN[12], initially low-level features such as lines and edges are discovered by the filters. These build up to progressively higher features in subsequent layers, such as a nose, eye or ear, as the feature maps become inputs for the next layer in the CNN architecture[8]-[10].

B) RECTIFIED LINEAR UNIT (RELU) LAYER

The RELU layer is an activation function that sets negative input values to zero. This simplifies and accelerates calculations and training, and helps to avoid the vanishing gradient problem. Mathematically it is defined as:

$$f(x)=\max(0,x).$$

Where x is the input to the neuron. Other activation functions include the sigmoid, tanh, leaky RELUs, Randomized RELUs and parametric RELUs.

C) POOLING LAYER

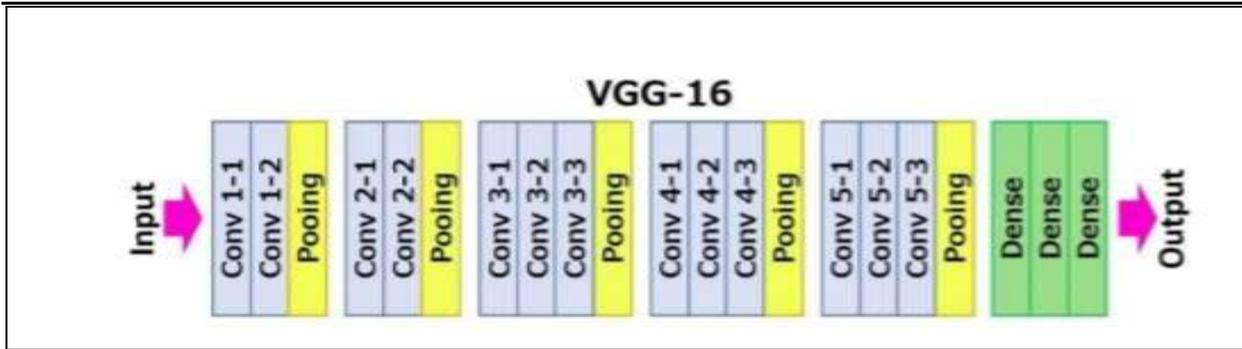
The Pooling layer is inserted between the Convolution and RELU layers to reduce the number of parameters to be calculated, as well as the size of the image (width and height, but not depth). Max-pooling is most commonly used; other pooling layers include Average pooling and L2-normalization pooling. Max-pooling simply takes the largest input value within a filter and discards the other values; effectively it summarizes the strongest activations over a neighborhood. The rationale is that the relative location of a strongly activated feature to another is more important than its exact location.

D) FULLY CONNECTED LAYER

The final layer in a CNN is the Fully Connected Layer, meaning that every neuron in the preceding layer is connected to every neuron in the Fully Connected Layer. Like the convolution, RELU and pooling layers, there can be 1 or more fully connected layers depending on the level of feature abstraction desired. This layer takes the output from the preceding layer (Convolutional, RELU or Pooling) as its input, and computes a probability score for classification into the different available classes. In essence, this layer looks at the combination of the most strongly activated features that would indicate the image belongs to a particular class. For example, on histology glass slides, cancer cells have a high DNA to cytoplasm ratio compared to normal cells. If features of DNA were strongly detected from the preceding layer, the CNN would be more likely to predict the presence of cancer cells.

VGG-16/VGG-19 ARCHITECTURE

VGG19 is a variant of VGG model which in short consists of 19 layers (16 convolution layers, 3 fully connected layer, 5 MaxPool layers and 1 SoftMax layer). In general **VGG16 and VGG 19 is quite similar** except the layers difference, actually VGG 19 gives more accuracy due to increase of some hidden layers compared with VGG 16. In this proposed application we applied VGG 19 on given dataset, so that we got accuracy of 21 percent on 5 epochs.



6. EXPERIMENTAL RESULTS

Implementation is a stage where the theoretical design is converted into a programmatic manner. In this proposed application we try to use PYTHON as a programming language in which Google Collaboratory or Jupiter Notebook as a working platform to process the current application.

STEP 1: IMPORTING ALL NECESSARY LIBRARIES

```
Please download the dataset from the below url
https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

# import the libraries as shown below
from keras.layers import Input, Lambda, Dense, Flatten
from keras.models import Model
#from keras.applications.resnet50 import ResNet50
from keras.applications.vgg16 import vgg16
from keras.applications.vgg16 import preprocess_input
from keras.preprocessing import image
from keras.preprocessing.image import ImageDataGenerator
from keras.models import Sequential
import numpy as np
from glob import glob
import matplotlib.pyplot as plt
```

STEP 2: LOAD TRAIN AND TEST DATA INTO SEPARATE VARIABLES

```
[ ] # re-size all the images to this
IMAGE_SIZE = [224, 224]

train_path = 'Datasets/train'
valid_path = 'Datasets/test'

# Import the Vgg 16 library as shown below and add preprocessing layer to the front of VGG
# Here we will be using imagenet weights

vgg = VGG16(input_shape=IMAGE_SIZE + [3], weights='imagenet', include_top=False)
```

STEP 3: DATA PRE-PROCESSING

```
[ ] from keras.preprocessing.image import ImageDataGenerator

train_datagen = ImageDataGenerator(rescale = 1./255,
                                   shear_range = 0.2,
                                   zoom_range = 0.2,
                                   horizontal_flip = True)

test_datagen = ImageDataGenerator(rescale = 1./255)
```

STEP 4: Building CNN Model

```
[ ] # create a model object
model = Model(inputs=vgg.input, outputs=prediction)

# view the structure of the model
model.summary()
```

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 224, 224, 3)	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160

STEP 5 : Start Training CNN with Parameters and fit the model.

```
# fit the model
# Run the cell. It will take some time to execute
r = model.fit_generator(
    training_set,
    validation_data=test_set,
    epochs=5,
    steps_per_epoch=len(training_set),
    validation_steps=len(test_set)
)
```

WARNING:tensorflow:From C:\Users\krish.naik\AppData\Local\Continuum\anaconda3\envs\myenv\lib\site-packages\tensorflow\python\ops\math_ops.py:3866: to_int32 (from tensorflow.python.ops.math_ops) is deprecated. Instructions for updating:
Use tf.cast instead.

Epoch 1/5
53/163 [=====] - ETA: 23:22 - loss: 0.4488 - acc: 0.8510

After training the model we got classification accuracy of 85 % For VGG16 Model

STEP 6 : TEST ON VGG-19 MODEL

```

from keras.applications.vgg19 import VGG19
from keras.applications.inception_v3 import InceptionV3

```

After training the model we got classification accuracy of 95 % For VGG19 Model

STEP 7 : PREDICTIONS**PERFORMANCE ANALYSIS OF CNN MODELS FOR LUNG DISEASE DETECTION**

Accuracy	Deep Learning Method	
	VGG-16	VGG-19
	85 %	95 %

From the above table we can clearly identify the accuracy of proposed CNN model is almost 95 percent when compared with other existing vgg-16 MODEL

7. CONCLUSION

In this current work we for the first time designed and implemented an application using deep learning CNN model in the medical field for Chest or Lung disease detection from chest X-ray images. We try to design an application which can able to identify the abnormality present in the human chest or lungs from the affected part of image and then find out the abnormality. At present, it is very interesting to design the deep intricate neural network (CNN) is the latest image recognition solution. Here we try to gather several infected lung images as well as normal lung images and then try to train the system with all these images. Once the model is trained then we try to give a sample chest X-ray image as input and check whether that image is having abnormality or not. To solve the above problem, we developed a deep learning model using a CNN algorithm to detect the abnormality present in human chest. By conducting various experiments on our proposed model, we achieved a classification accuracy of 95.0% when applied to the test dataset. If this application is designed perfectly we can able to use this model on covid patients so that we can able to check the current lung status of that covid affected patients and this can greatly reduce lot of physical examination in this pandemic situation.

8. REFERENCES

1. Bousquet, J. Global Surveillance, Prevention and Control of Chronic Respiratory Diseases; World Health Organization: Geneva, Switzerland, 2007; pp. 12–36.
2. Forum of International Respiratory Societies. The Global Impact of Respiratory Disease, 2nd ed.; European Respiratory Society, Sheffield, UK, 2017; pp. 5–42.

3. World Health Organization. Coronavirus Disease 2019 (COVID-19) Situation Report; Technical Report March; World Health Organization: Geneva, Switzerland, 2020.
4. Rahaman, M.M.; Li, C.; Yao, Y.; Kulwa, F.; Rahman, M.A.; Wang, Q.; Qi, S.; Kong, F.; Zhu, X.; Zhao, X. Identification of COVID-19 samples from chest X-Ray images using deep learning: A comparison of transfer learning approaches. *J. X-Ray Sci. Technol.* 2020, 28, 821–839.
5. Yahiaoui, A.; Er, O.; Yumusak, N. A new method of automatic recognition for tuberculosis disease diagnosis using support vector machines. *Biomed. Res.* 2017, 28, 4208–4212.
6. Hu, Z.; Tang, J.; Wang, Z.; Zhang, K.; Zhang, L.; Sun, Q. Deep learning for image-based cancer detection and diagnosis-A survey. *Pattern Recognit.* 2018, 83, 134–149.
7. American Thoracic Society. Diagnostic Standards and Classification of Tuberculosis in Adults and Children. *Am. J. Respir. Crit. Care Med.* 2000, 161, 1376–1395.
8. Setio, A.A.A.; Traverso, A.; de Bel, T.; Berens, M.S.; van den Bogaard, C.; Cerello, P.; Chen, H.; Dou, Q.; Fantacci, M.E.; Geurts, B.; et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. *Med. Image Anal.* 2017, 42, 1–13.
9. Shen, D.; Wu, G.; Suk, H.I. Deep Learning in Medical Image Analysis. *Annu. Rev. Biomed. Eng.* 2017, 19, 221–248.
10. Wu, C.; Luo, C.; Xiong, N.; Zhang, W.; Kim, T.H. A Greedy Deep Learning Method for Medical Disease Analysis. *IEEE Access* 2018, 6, 20021–20030.
11. Ma, J.; Song, Y.; Tian, X.; Hua, Y.; Zhang, R.; Wu, J. Survey on deep learning for pulmonary medical imaging. *Front. Med.* 2019, 14, 450–469.
12. Rajaraman, S.; Candemir, S.; Xue, Z.; Alderson, P.O.; Kohli, M.; Abuya, J.; Thoma, G.R.; Antani, S.; Member, S. A novel stacked generalization of models for improved TB detection in chest radiographs. In Proceedings of the 2018 40th Annual International Conference the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 17–21 July 2018; pp. 718–721.