OPEN ACCESS
International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958 (Online), Volume-9 Issue-5, June 2020

Detection of Lung Cancer Disease using Machine Learning

Amit Kore, Saurabh Kandekar, Anurag Yadav, Shubham Ohol

Abstract: Cancer causes cell to split uncontrollably. Lung Cancer results in rapid cell growth and division of such infected cell, such growth of cells called tumor. Lung is the first organ where lung tumor begins and can spread to lymph nodes and so on. Early identification of lung cancer would facilitate in sparing a large no. of lives. If we compare death rates in any cancer then lung cancer has highest mortally rate. This article presents an automated web-based system for disease detection in lung using X-Ray images. To identify disease in lung in X-ray images, as it provides detailed picture and gives clear idea of lung in the body. For this project dataset of chest x-ray was taken from Kaggle. Using Mobile Net model we predicted the lung disease. Using this approach, we can early detect the disease present in lung which causes lung cancer

Keywords: X-Ray, lung cancer, Mobile Net model, Images.

I. INTRODUCTION

Lung cancer is the cause of majority of cancer related deaths in both men and women. This is because there is no significant signs or symptoms related to lung cancer during its initial onset. This causes the detection of lung cancer during its advanced phase. Even after diagnosis, the survival rate is very less, thereby continuously increasing the death rate year by year. So, it is very important to detect or diagnose the lung cancer in early stages. Survival rate can go up if we diagnose the lung cancer disease in early stages. Early lung cancer detection has been accomplished with the technological advancements with reduced cost[3]. Another cause of late detection is human error factor because of which radiologist might misdiagnose lung cancer to some other problem. As the treatment of cancer is effective only in early stages. Among other types of cancer disease Lung cancer has the second most death rate. So even after it gets detected the chances of survival is very less Early detection of cancer might increase the chances of survival to 80%. There are various methods to detect lung cancer. These are PET scan, MRI, CT scan and chest X-Rays.[1] To achieve method which are cost effective and more reliable, X-ray scan images are the best method. Automated system is the need in today's world to help radiologist in diagnosing more accurately. We have used MobileNet CNN model to predict the disease in in the lungs over x-ray images[4]. 'JPEG' images are taken from the dataset to process the image. Our design is found to be 94% accurate.

Revised Manuscript Received on June 15, 2020.
* Correspondence Author
Amit Kore*, Assistant Professor, Department of Information Technology AISSMS IOIT SP Pune University Sangamvadi, Pune , India. E-mail:amit.kore88@gmail.com
Saurabh Kandekar, Department of Information Technology AISSMS IOIT SP Pune University Sangamvadi, Pune – 411001, India. E-mail: Saurabh.kandekar@live.in
Anurag Yadav, Department of Information Technology AISSMS IOIT SP Pune University Sangamvadi, Pune – 411001, India. E-mail: yadav12anurag@gmail.com
Shubham Ohol, Department of Information Technology AISSMS IOIT SP Pune University Sangamvadi, Pune – 411001, India. E-mail: shubhamohol1997@gmail.com

II. METHODOLOGY

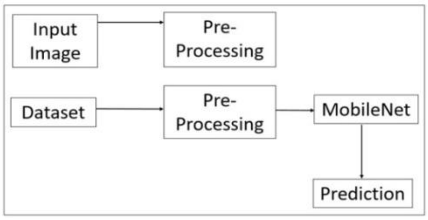


Fig. Block diagram of system.

A. MobileNet Model

We have used Mobile Net in our model. Mobile Nets are CNN based models which works very well on low latency and low power. Detection, Prediction, Classification, embeddings, are also build using Mobile Net model. similar to other large scale or heavy model used. Mobile Nets can be used efficiently using TensorFlow [2]. Easy light weight Deep Neural network is developed using depth wise separable convolutions which is based on simple framework of Mobile Net. Which basically means on each channel it performs a single convolution rather than combining all channels.

The job of the convolution layer was split into two subtasks: 1) depth wise convolution 2) pointwise convolution layer.

Depthwise Convolution - A depthwise convolution on each input channel maps a single convolution separately. Therefore, its number of output channel is the same of the number of input channel. Its computational cost is $Df^2 * M * Dk^2$.

Pointwise Convolution - The last operation is a pointwise convolution. It is a convolution with a kernel size of 1x1 that simply combines the features that are created by the depthwise convolution. Its computational cost is $M * N * Df^2$.

The input shape of the image taken was (128,128,3) for the same output shape is (None , 4 , 4 , 1024.) Input shape arguments are 'batch size', 'Hight', 'Width', 'Depth'.

B. Loading dataset –

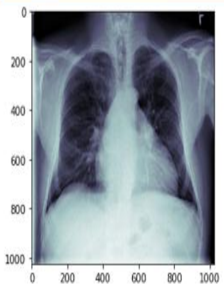
For this project dataset was taken from the well-known source of data "Kaggle". The name of dataset was 'Chest X-ray'. This Dataset has 30,800 special patients with disease labels, also 1,12,120 X-ray images. We can easily load the dataset. The largest openly available dataset of chest X-ray was 'Openi' image with 4143 images available.

C. Data analysis, data processing –

Data processing such as standardized age in digital form per year, age noise filtering, one hot attribute such as gender, snapshot of the image as well as the specific type of illness that the patient suffers. Analysis of data such as age, gender, and photo-taking will affect the likelihood that a patient will develop a specific disease. Image processing such as resizing images to the same size, for black and white or color images for parallel processing and comparison. Loading of sample Image.

```
import matplotlib.image as mpimg

img = mpimg.imread('/home/ubuntu/Downloads/lung/data/images_001/images/00000001_000.png')
imgplot = plt.imshow(img, cmap='bone')
plt.show()
```



As 1,12,120 images are available in the dataset out of which 2250 are for training and random no of images are for testing purpose. The shape of each image is 128x128.

D. Building the model –

Here we have used Sequential model. This model is build using sequential It helps us to create the model layer by layer. Sigmoid is the activation function we have used for the layer. Our first layer takes input shape(128x128), We have used 'Flatten layer'. Flatten works as a bridge between the dense layer and convolution layer. As a output layer we have used 'Dense' layer. In many neural networks 'Dense' layer is used as a standard layer.

E. Compiling the Model –

Next step is to compile the model. Following three parameters were used while Compiling the model. 'optimizer', 'loss' and 'metrics', learning rate was controlled by the 'optimizer'. In our model 'adam' was used. In many cases 'Adam' is generally used. Throughout the training The "adam" optimizer adjusts the learning rate.
We have used 'binary_crossentropy' for our loss function. In multi-label problems, where at the same time, an example can belong to multiple classes. The model decides whether example belongs which class. To make things even easier to interpret, to get the Validation accuracy rate on training model we have used binary accuracy metric.

F. Training Model –

Next step is to train our model. Fit () function model was used to train the model as follows - train data - train_X, target data train_y. Test set data was used to validate the data. As per dataset, which we was divided into X_test and y_test.
The model will go through the given data many no. of times which is called as 'epochs'. Model will be improved if we run more no. of epochs. up to a certain point.

```
Train on 15749 samples, validate on 6751 samples
Epoch 1/8
15749/15749 [==============================] - 27s 2ms/step - loss: 0.1728 - binary_accuracy: 0.9487 - mae: 0.0898 - val_loss: 0.1826 - val
_binary_accuracy: 0.9475 - val_mae: 0.0864
Epoch 2/8
15749/15749 [==============================] - 27s 2ms/step - loss: 0.1701 - binary_accuracy: 0.9486 - mae: 0.0890 - val_loss: 0.1884 - val
_binary_accuracy: 0.9475 - val_mae: 0.0781
Epoch 3/8
15749/15749 [==============================] - 27s 2ms/step - loss: 0.1686 - binary_accuracy: 0.9488 - mae: 0.0883 - val_loss: 0.1845 - val
_binary_accuracy: 0.9475 - val_mae: 0.0966
Epoch 4/8
15749/15749 [==============================] - 27s 2ms/step - loss: 0.1672 - binary_accuracy: 0.9489 - mae: 0.0880 - val_loss: 0.2018 - val
_binary_accuracy: 0.9472 - val_mae: 0.0751
Epoch 5/8
15749/15749 [==============================] - 27s 2ms/step - loss: 0.1642 - binary_accuracy: 0.9489 - mae: 0.0872 - val_loss: 0.1848 - val
_binary_accuracy: 0.9473 - val_mae: 0.0849
Epoch 6/8
15749/15749 [==============================] - 28s 2ms/step - loss: 0.1622 - binary_accuracy: 0.9489 - mae: 0.0863 - val_loss: 0.1867 - val
_binary_accuracy: 0.9475 - val_mae: 0.0967
Epoch 7/8
15749/15749 [==============================] - 27s 2ms/step - loss: 0.1609 - binary_accuracy: 0.9489 - mae: 0.0856 - val_loss: 0.1953 - val
_binary_accuracy: 0.9473 - val_mae: 0.0811
Epoch 8/8
15749/15749 [==============================] - 27s 2ms/step - loss: 0.1573 - binary_accuracy: 0.9493 - mae: 0.0848 - val_loss: 0.1949 - val
_binary_accuracy: 0.9406 - val_mae: 0.1037
```

We have gotten to 94.00% accuracy on our validation set after 8 rounds.(Epochs)

G. Using our model for predictions –

we use the predict function to predict the result.

III. RESULT

The designed system was implemented as web-based system software. The Mobile Net based on Convolutional Neural Network (CNN) has been implemented. To understand the different lung disease the model has been trained on training data. The final app was tested & deployed over Heroku server with a web interface to upload a new image for lung cancer disease detection. Dataset for this project was taken from Kaggle (NIH X-Ray dataset). Our system shows the disease detected on the web-based system by our MobileNet CNN based system automatically on X-ray. The result was generated in graphical format. We have achieved 94% accuracy.

IV. CONCLUSION

To detect the disease present in the Lung which causes the cancer, in the lungs, we used Mobile Net model bases on the CNN. To train the proposed and designed system input taken as X-ray images of lung with different shape and size. The designed system achieves 94% accuracy. It also detects presence and absence of diseases in the lungs. In the upcoming days the system will be trained with more accurate data.

International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958 (Online), Volume-9 Issue-5, June 2020

Training the model using X-ray is prone to many problems such as less availability of data, inaccurate imaging etc. This can be reduced by the use of GAN networks. The final aim is to develop a model to detect the disease present in the lung with low cost using machine learning.

ACKNOWLEDGEMENT

We would like to express our thanks of gratitude to our project guide, Asst.Prof. Amit Kore sir, for his guidance and valuable feedback in completing our project. We also would also like to extend our gratitude to Dr.P. B. Mane sir, Principal, AISSMS'S Institute of Information Technology and Dr. Meenakshi Thalor mam, HOD, Information Technology for providing us with all the facility that was required during final year project. Further on, we want to thank all the faculty member of IT department for their valuable suggestions. We also extend our heartful thanks to our friends who helped us during this project.

REFERENCES

1. Nidhi S. Nadkarni, Prof. Sangam Borkar. "Detection of Lung Cancer in CT Images using Image Processing.", Goa College of Engineering.
2. S. Sasikala, M. Bharathi, B. R. Sowmiya. "Lung Cancer Detection and Classification Using Deep CNN". (IJITEE)
3. S.Kalaivani, Pramit Chatterjee, Shikhar Juyal, Rishi Gupta. "Lung Cancer Detection Using Digital Image Processing and Artificial Neural Networks". International Conference on Electronics, Communication and Aerospace Technology ICECA 2017.
4. Vinod Kumar, Ashu Gupta, Rattan Rana, Kanwal Garg. "Lung Cancer Detection from X-Ray Image Using Statistical Features*". International Journal of Computing Academic Research (IJCAR)
5. Christine Tataru, Darvin Yi, Archana Shenoyas, Anthony Ma." Deep Learning for abnormality detection in Chest X-Ray images". Stanford University: Computer Science
6. Md. Badrul Alam Miah, Mohammad Abu Yousuf. "Detection of Lung Cancer from CT Image Using Image Processing and Neural Network".
7. https://www.medicalnewstoday.com/articles/323701.php
8. https://en.wikipedia.org/wiki/Lung_cancer
9. Albert Chon, Peter Lu, NiranjanBalachandar "Deep Convolutional Neural Networks for Lung Cancer Detection".
10. M.S. Al-Tarawneh, "Lung cancer detection using image processing techniques," Leonardo Electronic Journal of Practices and Technologies, vol. 20, pp. 147– 58, May 2012
11. Shubhangi Khoragade, Aditya Tiwari, " Automatic Detection of Major Lung diseases using Chest Radiographs and classification by Artificial Neural Network", 1st IEEE International Conference on Power Electronics, Intelligent control and energy systems, IEEE, 2016.
12. Anita Chaudhary, Sonit Sukhraj Singh, "Lung Cancer Detection on CT Images using Image Processing", International conference computing

    sciences, IEEE, 2012.

AUTHORS PROFILE

Asst. Prof. Amit Kore - Working with AISSMS IOIT as an assistant professor. Pursuing PhD from KIIT, Bhubaneswar. Domain of expertise are Computer network, security, WSN.

Saurabh Kandekar - Pursuing Bachelor of Engineering in Information Technology at AISSMS Institute of Information Technology, SP Pune University. So far paper published in Fake News Detection using Ensemble Learning.

Anurag Yadav - An IT student at AISSMS Institute of Information Technology, Pune. Learner at heart. Working on challenging projects to change the world.

Shubham Ohol- An IT student at AISSMS institute of technology, Pune.