**PAPER** • **OPEN ACCESS**

# Deep Learning Based Lung Cancer Detection and Classification

To cite this article: N Kalaivani *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **994** 012026

# Deep Learning Based Lung Cancer Detection and Classification

**N Kalaivani[1*], N Manimaran[2], Dr. S Sophia[3]&D Devi D[4]**
[1]Assistant Professor,Sri Krishna College of Engineering and Technology,
[2]Assistant Professor,SACS MAVMM Engineering College,
[3]Professor,Sri Krishna College of Engineering and Technology,
[4]Assistant Professor,Sri Krishna College of Engineering and Technology.

Corresponding author mail id: kalaivanin@skcet.ac.in

**Abstract:**Lung diseases are indeed the lung-affecting diseases which impair the respiratory mechanism. Lung cancer has been one of the leading causes of mortality in humans worldwide. Early detection can enhance survival chances amid humans. If the condition is diagnosed in time, the average survival rates for people with lung cancer rise from 14 to 49 percent. While computed tomography (CT) is far more effective than X-ray, a thorough diagnosis includes multiple imaging approaches to support each other. A deep neural network for detecting lung cancer from CT images is developed and evaluated. For the classification of the lung image as normal or malignant, a densely connected convolution neural network (DenseNet) and adaptive boosting algorithm wasused.A dataset of 201 lung images is used in which 85% of the images are used for training and 15% of the images are used for testing and classification. Experimental results showed that the proposed method achieved an accuracy of 90.85%.
Keywords: DenseNet, Image Processing, Deep Learning, Convolution Neural Networks (CNN).

## 1. Introduction

Lung cancers have been identified as one of the world's most serious causes of death [1]. It is among the most malignant tumors that can affect human wellbeing. Its death rate scores among all tumor deaths, and is also the top killer towards male and female cancer death [2-3]. There have been nearly 1.8 million fresh cases of lung cancer annually (13 percent among all cancers), 1.6 million deaths worldwide (19.4 percent among all cancers). Lung cancer is a proliferation of expanding and developing irregular cells into a tumour. Of the other forms of cancer, the death rate of lung cancer is the greatest. Cigarette smoke induces an approximate 85 percent of cases of lung cancer in males and 75 percent in females. Lung cancer is amongst the most terrible illnesses in the developing countries, with a death rate of 19.4 percent. Lung cancer is among the most dangerous cancer worldwide, with lowest success rate following diagnosis, with a steady rise in casualty count per year [4-6]. Advantages of Fuzzy logic in the earlier predictions will lead to result oriented analysis [5]. Survival of lung cancer as a result of diagnosis is directly related to its progress. Yet individuals have a greater success rate it will be found in the early stages of life. Cancer cells are distributed in blood from the lungs, the lymph fluid that covers the lung tissues. The lymph passes into lymph vessels that discharge through lymph nodes in the lungs and chest region. Examination and treatment of lung disease has become one of the biggest obstacles that humanity faces in recent years. Early tumor diagnosis will reliably promote its survival of vast numbers of life around the world. This paper introduces a method that uses a convolutional neural Network (CNN) to identify the lung tumors as malignant/benign. The results

achieved by CNN are 81 percent, which are more effective than the results achieved by the traditional neural network.

## 2. Existing System

Sharma and Jindal have suggested a method for identifying lung cancer tumor by using fuzzy interference system and the active contour model. This device makes use of gray level transformation to improve the image contrast. Binarization of images is achieved prior to segmentation, and the resulting image is divided by using active contour model. Moreover, the system efficiency is 94.12 percent. Bhatnagar et al., devised a technique utilizing watershed segmentation. It utilizes Gabor filter in preprocessing to improve the image quality. They compared the effectiveness of the algorithm with neural fuzzy and the region growing method. The accuracy was observed as 89.1 percent which is significantly greater than the method of segmentation using neural fuzzy and region growing method. This framework being the current best solution, but it has certain drawbacks. Just a few characteristics of cancer nodules have been identified. No preprocessing features such as noise reduction and image smoothing has been applied which will potentially help improve the effective identification of nodules. There has been no proper classification method as benign or malignant. This approach varies from the current best solution, and proposed new method. Median filter and Gaussian filter is introduced in preprocessing stage rather than Gabor Filter. The best system ends after cancer nodule detection, extraction of feature and accuracy estimation. Yet, it has not enforced its classification as normal or irregular.

## 3.Proposed Model

Convolutional neural networks are designed for minimizing the number of parameters and adjusting the architecture of the network for image classification. Convolutional neural networks are made up of a series of layers organized according to their features and functionality. A ConvNet 's architecture is close to that of the human brain connectivity pattern of neurons inspired by the Visual Cortex structure. Data augmentation is the process by which data quantity and complexity increase. We will collect fresh data rather than converting the data already available. Data augmentation is an important phase in deep learning, since we require vast quantities of data in deep learning and in certain instances it's not really possible to capture thousands or even millions of images, so this data augmentation comes to the scene. It allows us to maximize the dataset size and add uncertainty within the datasets. In addition to this, webpage is created to store the database of the patients, here the patients can login to their page at any time for the future references. Hospital locality server details can also be inferred easily.Fig.1 depicts the proposed model.

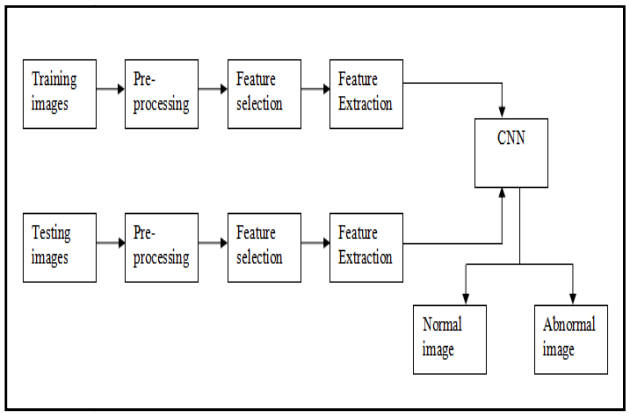


**Fig.1** Proposed Model

## 4. Operation of the Proposed Model

The working is divided into two parts, i.e., training and testing parts. A sample of 201 images is used where 171 images are used for training and the remaining 30 images are used for testing. In testing part, images are first pre-processed, where resizing and blur removal of images are done by using

histogram equalization due to which non-linear images are stretched and pixel values are redistributed. Fig.2shows the articulation of data processing.

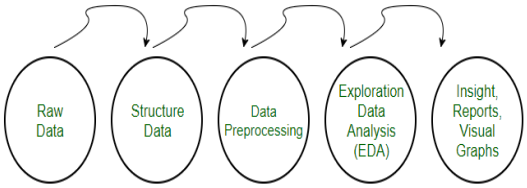


**Fig.2** Data Processing

The results obtained from training and testing part are fed into in CNN layers where the images are classified and the output is obtained. The algorithm used for classification is ADABOOST (Adaptive Boosting), where accuracy calculation is of the images is done based on the sample weights of images. Fig.3shows the conventional CNN architecture.

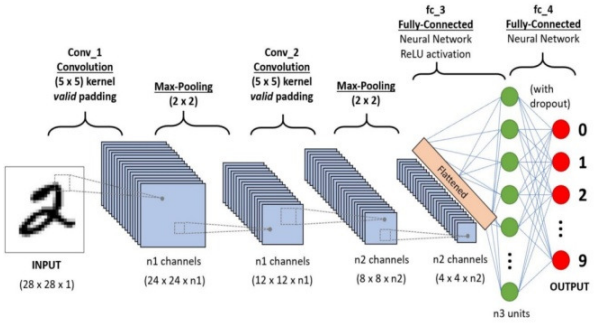


**Fig.3** CNN Architecture

The convolutional layer is the first and foremost layer and is the essentiallayer of a CNN. The layer's parameters contain a set of filters that are referred as kernels. Then the output obtained from ReLu layer is passed on to the next Pooling layer.There are two common pooling approaches, they are average(optimized) pooling and max pooling that condense the average presence of a feature.Fully connected layer is the most essential layer component of Convolutional Neural Networks (CNNs). Which has been successful in recognized for classifying images for computer vision. AdaBoost algorithm is the best usage of algorithm in deep learning. XAMPP is an open-source and cross-platform web server stack package. XAMPP was mainly developed for Apache HTTP Server, scripts written in the PHP etc., XAMPP is used as web server in most cases, and thus it is designated as the local test server to a live server. Thus the web page consists of home page which included the hospital login and user login. Patients can login to the portal using their patient id at any time to have their results.

## 5. Results and Discussions

Total number of images taken is about 201, among which 171 images are been trained and the remaining 30 images are tested accordingly. The trained images are then moved on to the classification part where each image is classified as either normal or malignant. Output of each image is been displayed with its accuracy as either normal or malignant. The formula to calculate accuracy is as follows:

$Accuracy = \frac{TP}{TP+FP}$, *here TP – True Positive and FP – False Positive*

Output of each image has its own true positive(TP),true negative(TN), false positive(FP) and false negative(FN).

| S. No. | Parameters | Accuracy |
|---|---|---|
| 01 | Trained images | 171 |
| 02 | Tested images | 30 |
| 03 | True Positive | 5 |
| 04 | True Negative | 20 |
| 05 | False Positive | 4 |
| 06 | False Negative | 1 |
| 07 | Accuracy | 90.85% |

Thus, by this approach the output is obtained using deep learning, which includes detection and classification of the images. Finally, the overall accuracy obtained is about 90.85%. This is obtained at an accurate ratio. The contents are then stored to the web server which contains hospital and admin users.

| Dataset Image Number | Accuracy of each image |
|---|---|
| 15 | 99.8% |
| 30 | 96.41% |
| 22 | 78.92% |
| 01 | 99.89% |
| 10 | 98.17% |
| 24 | 99.90% |
| 19 | 59.56% |
| 21 | 99.76% |
| 16 | 98.84% |
| 03 | 74.93% |
| 02 | 93.23% |
| **Overall accuracy** | **90.85%** |

Accuracy of 11 images are calculated for predicting the overall accuracy, the average of all individual accuracy's of each image is been taken. Thus, the overall accuracy obtained is the average of accuracy of individual images is obtained as 90.85%.

## 6. Conclusion

The greatest benefit of deep learning above other algorithms on machine learning is its ability in executing feature engineering on its own. This examines the data to search for associated features and incorporates them to provide for quicker learning.It takes advantage of spatial coherence in the input. The training and testing of images are done where images are pre-processed and feature selection and feature extraction of images are done. Once training and testing part is done successfully, the CNN algorithm classifies the input lung image either as normal or abnormal and the output will be displayed. Hence, a Deep CNN network is used for the classification of lung images for the detection of cancer.

## References

[1] RLSiegel KDMiller AJemal Cancer statistics 2017*Cancer Journal for Clinicians* 67(1): 7–30 2017

[2] RLSiegel KDMiller SAFedewa Colorectal cancer statistics *Cancer Journal for Clinicians* 68(1): 7130 2018

[3] JLortet-Tieulent ISoerjomataram JFerlay MRutherford EWeiderpass FBray ''International trends in lung cancer incidence by histological subtype: Adenocarcinoma stabilizing in men but still increasing in women Lung Cancer 84(1): 13-22 2014

[4] DSharma GJindal ‘‘Computer aided diagnosis system for detection of lung cancer in CT scan images *International Journal of Computer and Electrical Engineering* 3(5): 714-718 2011
[5] Janarthanan R Balamurali R Annapoorani A and Vimala V 2020 Prediction of rainfall using fuzzy logic Materials Today: Proceedings
[6] DBhatnagar AKTiwari VVijayarajan AKrishnamoorthy ‘‘Classification of normal and abnormal images of lung cancer *IOP Conference Series: Materials Science and Engineering* Vol 263 2017