# Prediction and Classification Lung Cancer using Machine Learning

**Shruthi G R [1], Ambika K B [2]**

[1] M-Tech Student, CSE, Maharaja Institute of Technology, Belavadi, Mandya, India, 571477
[2]Assitant Professor, Department of CSE, Maharaja Institute of Technology, Belavadi, Mandya, India, 571477

**Abstract:**

In wealthy nations, lung cancer is one of the most lethal diseases, and early detection is challenging. One of the most difficult problems people have faced recently is lung cancer diagnosis and treatment. Every day, early tumor diagnosis will continue to save many lives all around the world. This research presents a strategy that integrates a Convolutional Neural Network (CNN) with the AlexNet Network Model to categorize lung cancers as benign or malignant. One of the transfer learning models is AlexNet CNN. Here, in this project, the aim will be focussed onto list, discuss, and analyse several ML algorithm to classify and detect lung cancer and its stages. The proposed CNN achieves a high degree of accuracy, which is more effective than the accuracy attained by traditional neural network systems.

**Keywords:** Alex Net , Convolution Neural Network, Machine learning, , Transfer learning, Tumor.

## Introduction

In 2018 it was estimated that approximately 9.6 million deaths were claimed by lung cancer. Lung cancer tops the list if a person talks about the types and their shares. Estimated cases of lung cancer are around 2.09 million with 1.76 million deaths which account for around 84% deaths [1]. Due to this reason lung cancer has been entitled as one of the most fatal diseases. Cancer is when cells in the body begin to grow out of controlAs per the statistics of American Cancer Society, lung cancer is the major cause of the cancer related death in USA. The estimates of new cases and deaths in 2012 are 2, 26,160 and 1, 60,340 respectively. In 2013 the estimated new cases and deaths are 2, 28,190 and 1, 59,480 [2-3] respectively. In India approximately 63,000 lung cancer cases are reported every year. When a person has lung cancer, then they have abnormal cells that cluster together to form a tumor (nodule). Not all tumors are cancerous. These non-cancerous tumors are called benign nodules. The other cancerous nodules that grow without order, control and obliterate the healthy lung tissues around them are called malignant nodules [4-6].

Tumor is made by multiplication of abnormal cells in lung cancer. Cancer cells tend to spread really fast due to blood streams anslymphfluid that is present in lung tissue. In general, due to normal lymph flow, cancer cells frequently migrate to the middle of the chest [7-10]. As cancer cells migrate to other tissues, metastasis occurs. CT image technique is the most common out of the mentioned methods due to its ability to give a view excluding overlapping structures. Interpreting and recognizing cancer is complicated for doctors. CT photographs are accurate for the diagnosis of lung cancer. To identify lung cancer, image processing, and deep learning methods will be used. Accuracy can be improved using these approaches. Tumor detection and determination of its form, size, and location is a tough task [11-13]. Timely detection helps in saving a lot of time. And this time can be used in providing early treatment to the patient. In this project, pre-processing (removing noise if any), post processing (segmentation) and classification techniques will be used to classify tumors into one of the two groups i.e. Malignant and Benign. Benign refers to a non-cancerous tumor and it doesn’t spread to other parts. Exploring different methods to diagnose lung cancer will be a prime aim in this paper. Computed tomography can be used to capture images of lungs across various dimensions so that a 3D image of the chest can be formed. This 3D image can be used to detect tumors present [14]. Normally a doctor or any field expert uses a CT image to detect cancer. Due to the large number of CT images, it is difficult for a doctor or radiologist to detect cancer quickly and accurately [15]. But with the advancement in technology, Computer-Aided Diagnosis (CAD) can be utilized to complete this duty efficiently and in considerably less time. A technique for quick detection of lung cancer using CT scans was proposed in this work. We discovered that the random forest classifier and regress or gave more accurate results than other algorithms when we examined different machine learning techniques from the survey. In addition, Convolution Neural Networks with class weights provide reliable results for categorizing CT scan pictures by overcoming unbalanced data issues. As a result, the proposed model overcomes the disadvantages and provides an application in which the user may obtain information ranging from a basic level, such as symptom prediction. Then it allows lung cancer stage classification using CT scans and presents insurance costs. The user may obtain all of the information they want from a single application in less time.

## Methodology

A typical lung cancer detection scheme is shown in Figure.1, which comprises of the following sequences of processing steps. Image Acquisition, Image Enhancement, and Lung Parenchyma Segmentation, Candidate nodule Detection, False Positive Reduction and Nodule Classification.

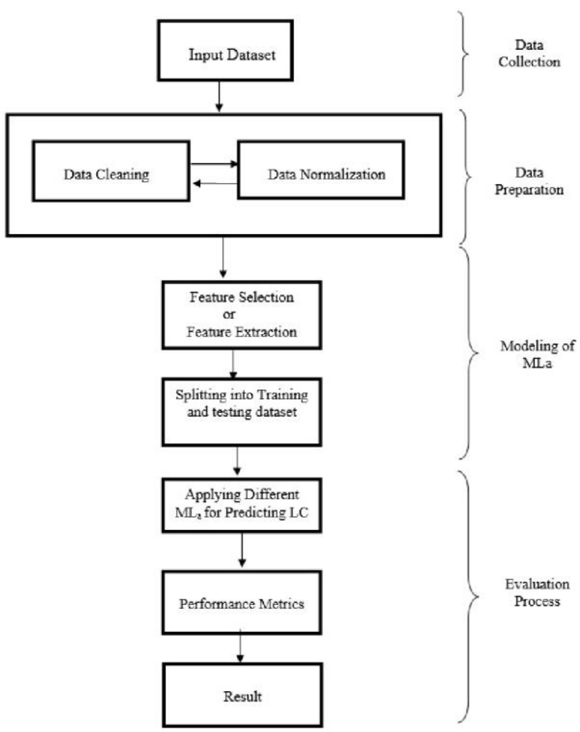


**Figure 1: Prediction and Classification Lung Cancer**

## Data Pre-Processing

The general process of building a machine learning model follows a specific process which includes image pre-processing. Since the images in dataset may contain images of different sizes, different extension and might contain noisy and blurry data. The images should be pre-processed before training by machine learning. Gaussian and Gabor filtering, Adaptive Gaussian Filtering, Wiener filtering and CLAHE are the various pro-processing techniques widely used for analysis as in Figure 2

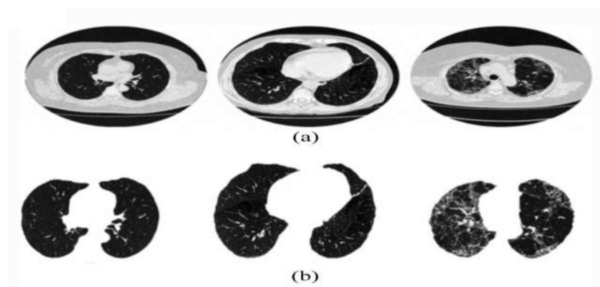


**Figure 2: Data Pre-Processing**

## Architecture

The first layer of Alex Net is as in figure 3. The first layer of Alex Net utilizes a convolution window with a specific form. Given that the objects in Image Net data tend to occupy more pixels and contain more visual information, a larger convolution window is required to effectively capture these objects.

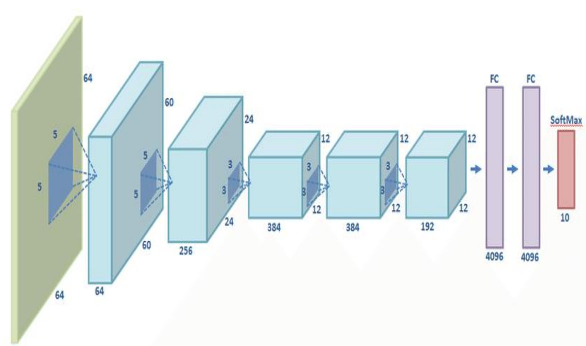


**Figure 3: Alex net Architecture**

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 08 | August - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Convolution Layers in a alex network are designed to compute the output of nodes that are connected to local regions of the input matrix. These layers use a set of weights, also known as filters, to perform dot products with the values associated with a local region of the input.

## Results and Discussion

The main intension of the present work is to determine whether a person's tumor present in the lung is malignant or benign. Once the cancer has been detected, further investigation continues to find the specific class of the malignance. The LIDC-IDRI dataset consists of 1018 cases gathered from a collaboration of seven academic centers and eight medical imaging companies. The dataset consists of 1018 CT scans from 1010 patients, with a total of 244,527 images. The Data set from LIDC-IDRI have a resolution of 512 × 512.

## Prediction and Classification of Lung Cancer

Our proposed model is able to classify the stages of lung cancer into Adenocarcinoma, Large cell carcinoma, Squamous cell carcinoma and Normal class. Figure 4 depicts performance the proposed model through mAP curve towards identification of a particular stage of the malignancy. The mAP values for the proposed method are presented n the graph for a specific epoch and iteration.

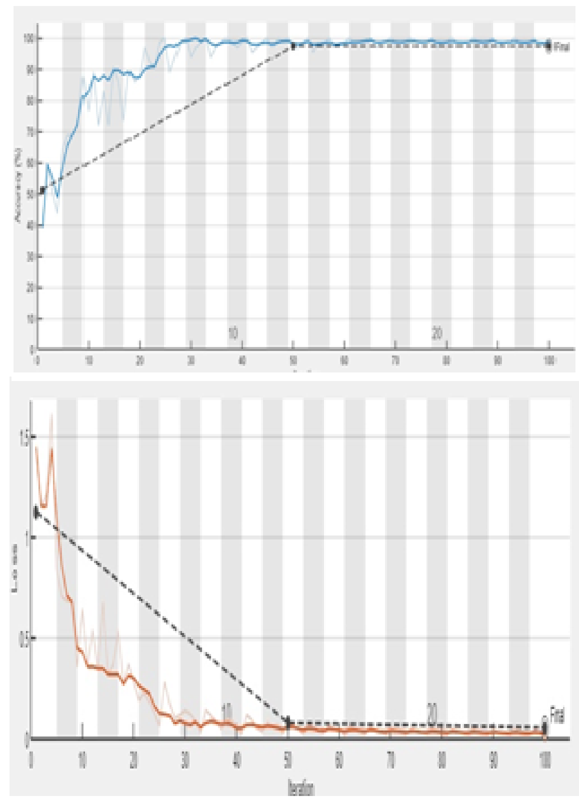

Figure 4: mAP curve (Accuracy and Loss)

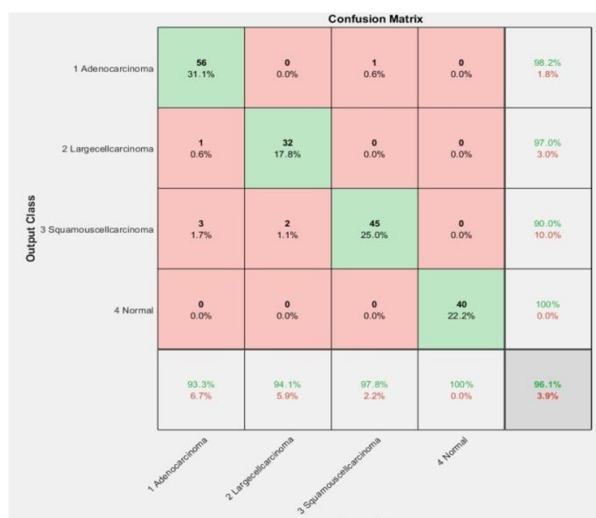


Figure 5: Confusion Matrix

The Alex Net model has a ROC-AUC value more than 0.99, indicating that it can distinguish all the classes with greater accuracy and lower false-positive rates than other models. The Alex Net model's overall performance is 96.7% for cancer detection and 96.1%for cancer categorization. According to the developers, the suggested diagnostic system would provide doctors an accurate and rapid diagnosis.

**Conclusion**

Using an Alexnet convolution neural network-based system, the malignancy tissues visible in the input lung CT picture were observed. Using an image of a lung with variously sized and shaped malignant tissues, the system was taught. Lung cancer classification would be implemented based on the following processes: multilevel thresholding for both lung extraction and cancer portion identification in lung extraction; morphological analysis and thresholding for lung cancer classification; and Alexnet CNN for lung cancer classification. In categorizing the presence and absence of malignant cells, the suggested technique achieves a classification accuracy of 96.7% for cancer detection and 96.1%for cancer categorization. Further, The Alex Net model has a ROC-AUC values are more than 0.99 for most of the classes, indicating that it can distinguish all the classes accurately.

REFERENCES:

[1] World Health Organisation‘s Official website https://www.who.int/news-room/fact-sheets/detail/cancer#:~:text=The%20most%20common%20causes%20of,Lung%20(1.76%20 million%20deaths)

[2]. Kanitkar SS, ThombareND and Lokhande SS 2015 Detection of lung cancer using marker-controlled watershed transform *Int. Conf. on Pervasive Computing* (ICPC) pp. 1-6.

[3]. Vas M and Dessai A 2017 Lung cancer detection system using lung CT image processing *Int. Conf. on Computing, Communication, Control and Automation* (ICCUBEA) pp. 1-5.

[4]. Punithavathy K, Ramya MM and Poobal S 2015 Analysis of statistical texture features for automatic lung cancer detection in PET/CT images *Int. Conf. on Robotics, Automation, Control and Embedded Systems* (RACE) pp. 1-5.

[5]. Sharma D and Jindal G 2011 Identifying lung cancer using image processing techniques *Int. Conf. on Computational Techniques and Artificial Intelligence* (ICCTAI) vol. 17 pp. 872-880.

[6]. Asuntha A, and Srinivasan A 2020 Deep learning for lung Cancer detection and classification. Multimedia Tools and Applications pp 1-32.

[7]. Teramoto A, Tsukamoto T, Kiriyama Y and Fujita H 2017 Automated classification of lung cancer types from cytological images using deep

convolutional neural networks *BioMed research International*.

[8]. Ozdemir O, Russell RL and Berlin AA 2019 A 3D Probabilistic Deep Learning System for Detection and Diagnosis of Lung Cancer Using Low-Dose CT Scans *IEEE Transactions on Medical Imaging* 1419-29.

[9]. Huang PW, Lin PL, Lee CH and Kuo CH 2013 A classification system of lung nodules in CT images based on fractional brownian motion model *Int. Conf. on system science and engineering* (ICSSE) pp. 37-40.

[10]. Šarić M, Russo M, Stella M and Sikora M 2019 CNN-based method for lung cancer detection in whole slide histopathology images *Int. Conf. on Smart and Sustainable Technologies* (SpliTech) pp. 1-4.

[11]. Alam J, Alam S and Hossan A 2018 Multi-stage lung cancer detection and prediction using multi-class svm classifier *Int. Conf. on Computer, Communication, Chemical, Material and Electronic Engineering* (IC4ME2) pp. 1-4.

[12]. Aggarwal T, Furqan A and Kalra K 2015 Feature extraction and LDA based classification of lung nodules in chest CT scan images *Int. Conf. on Advances in Computing, Communications and Informatics* (ICACCI) pp. 1189-1193.

[13]. Farahani FV, Ahmadi A and Zarandi MF 2015 Lung nodule diagnosis from CT images based on ensemble learning *IEEE Conf. on Computational Intelligence in Bioinformatics and Computational Biology* (CIBCB) pp. 1-7.

[14]. Rahman MS, Shill PC and Homayra Z 2019 A New Method for Lung Nodule Detection Using Deep Neural Networks for CT Images *Int. Conf. on Electrical, Computer and Communication Engineering* (ECCE) pp. 1-6.

[15]. Fakoor R, Ladhak F, Nazi A and Huber M 2013 Using deep learning to enhance cancer diagnosis and classification. *In Proceedings of the Int. Conf. on machine learning* vol. 28 (New York, USA).