



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

**Volume 9, Issue 9, September 2020**

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 7.512**

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)

# Detecting Lung Cancer from Live Dataset Using CNN

Sanyogita Bobade<sup>1</sup>, Pratik Deshpande<sup>1</sup>, Abhinav Deshpande<sup>1</sup>, Aditi Deshmukh<sup>1</sup>, Shalini Wankhade<sup>2</sup>

U.G. Student, Department of Computer Engineering, Sinhgad Academy of Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India<sup>1</sup>

M.Tech, Assistant Professor, Department of Computer Engineering, SAE Engineering College, Pune, Maharashtra, India<sup>2</sup>

**ABSTRACT:** Lung cancer is one of the major causes of high mortality rate, compared with the other cancer types. Detecting lung cancer at an early stage is still a challenge even after the advancement in technology. Early detection of lung cancer would lead to a significant drop in the mortality rate. Different methods have been developed for automatic detection of lung nodules from the images obtained from various medical imaging techniques. It has become possible to diagnose the probability of presence or absence of lung cancer without human interference, due to the advancement in image analysis using Deep Learning techniques. This paper presents a CNN (Convolutional Neural Network) system used for analysing the patient's CT (Computed Tomography) images and classify them according to their stages (Stage I or Stage II). The accuracy acquired by this system is 85%, which is progressively productive.

**KEYWORDS:** CNN, DICOM Image, Live Dataset, Lung Cancer.

## I. INTRODUCTION

Cancer ranks the second in leading causes of death globally, after cardiovascular diseases, and is responsible for approximately 9.6 million deaths in 2018. All over the world, almost 1 in 6 deaths is due to cancer [1]. Major cancers can be treated in the early stages of detection. Cancer cells grow faster than normal cells in the body and they form clusters ultimately leading to the development of nodules or tumor [2]. Lung cancer basically begins in the lungs and may either spread to the lymph nodes or to the other organs in the body. It is caused due to the major reasons including smoking, exposure to toxins and sometimes even due to genetic inheritance [3]. About 2.09 million cancer cases are of lung cancer which thus, makes it the most common cause of cancer deaths [1]. Early recognition of lung tumor is possible by utilizing many imaging strategies like Computed Tomography, Sputum Cytology, Chest X-beam, and Magnetic Resonance Imaging. Locating the tumor implies grouping it in two classes: (i) non-carcinogenic tumor (benign) and (ii) carcinogenic tumor (malignant) [4].

CT screening has been proven to be effective for diagnosing lung cancer at its early stage, present in the form of pulmonary nodules, thus decreasing the mortality rate [5]. Manual examination and determination methodologies can be improved with the use of different image processing procedures. Various advancements in the image processing methods for the early detection of the disease location are being worked on. However, the early detection of malignant growth is not extraordinarily improved. With further progress in the AI procedures, early detection of the malignant growth is endeavoured by few scientists. The neural system, thus, plays a vital role in the identification and classification of the malignant cells among the other typical tissues, which thus provides powerful and assistive AI-based disease detection. The disease treatment will be successful only when the tumor cells are precisely differentiated from the typical cells. AI-based disease analysis is used for classification of the tumor cells and preparation of the neural system [6]. This paper presents a CNN based system to classify the lung tumor as harmful or harmless.

## II. METHODOLOGY

The proposed system identifies the presence of lung cancer on the basis of the CT scans (DICOM images) using Convolutional Neural Network. In the first stage, the lung images (in pixel array format) are extracted from the DICOM image. The images are further processed by performing image pre-processing before providing them to the CNN. The principle aim of this examination is to identify whether the tumor present in the lung of the patient is

malignant (stage I or stage II) or benign. Fig. 1 shows the outline of the proposed system. As shown in the fig. 1, the proposed system readily recognizes the extent to which the tumor can be classified as harmful.

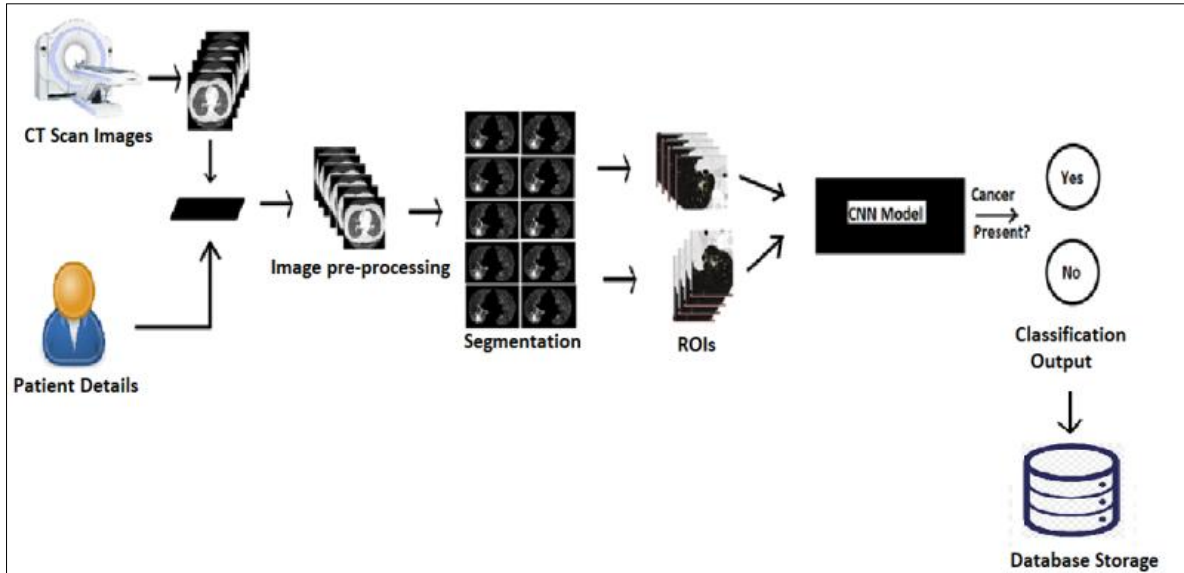


Fig. 1. Proposed System

### III. PREPROCESSING

The image to be provided to the neural network for classification, should be processed. In the preprocessing stage, initially the image is converted into a gray scale image. Different preprocessing and division strategies of the lung nodules are discussed in “Detection and Segmentation of Lymph Nodes for Lung Cancer Diagnosis” [7]. The image, initially of dimensions 512 X 512, is resized to 100 X 100. Further, the image is cleared of the blobs connected to the border. Subsequently, the pixel details which are altogether different from their neighbour pixels will be disposed off. Then the binary image is extracted by performing erosion operation. Erosion operation shrinks the brighter regions and enlarges the darker regions. Finally, the binary mask is superimposed onto the input image. This pre-processed image is further fed to the CNN which includes a series of hidden layers for complex data processing.

### IV. LIVE DATASET

The proposed system uses a live dataset acquired from a multispecialty hospital, which contains CT images of the patients. The system is provided with these images which predicts the presence and absence of cancer in them and classifies the cancer type into stage I or stage II, if present.

### V. CONVOLUTIONAL NEURAL NETWORKS (CNNs)

Mathematically, convolution refers to a function which is derived from two functions (by integration) which expresses how the shape of one function is modified by the other [8]. The Convolutional Neural Network consists of a number of hidden layers including Convolution layer, Pooling layer and Fully Connected layer. A shallow neural network consists of a limited number of hidden layers, while a deep neural network includes a series of hidden layers which allows complex data processing. Therefore, CNN is used as a deep learning model which can extract or build better features than a shallow model. In 2012, the neural networks were upgraded to superior quality, as Alex Krizhevsky implemented them to win that year’s ImageNet rivalry (essentially, the yearly Olympics of PC vision), dropping the unwanted record of NPV (negative predictive value) from 26% to 15%, which was considered as a huge improvement at that time [9]. The architecture of CNN for the proposed system is given in the Fig. 2.

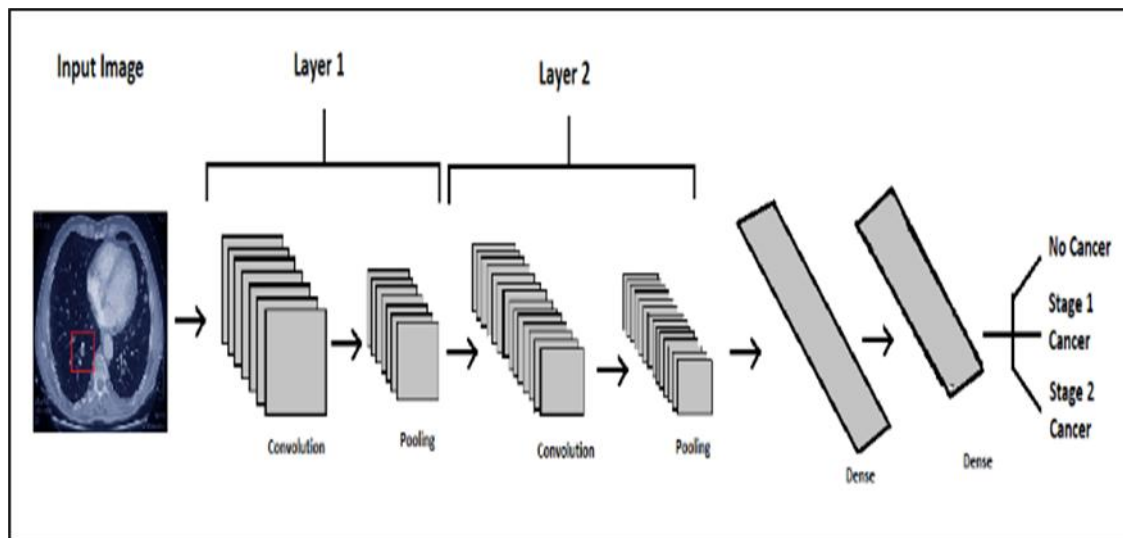


Fig. 2. Architecture of CNN

### VI. CONVOLUTION (CONV2D) LAYER

Convolution, being the first layer of CNN, is an application of the filter to the input image matrix, resulting in activation (feature) maps. The feature map indicates the location and strength of the feature; in this case, nodules in the input image. Since large size filters can precisely detect only the large sized objects, smaller sized filters (3X3) are chosen to detect small sized pulmonary nodules. It basically resembles a lot of learnable channels. 16 channels have been set for the first conv2D layer and 32 channels for the last. Each channel changes a part of the image (characterized by its size) utilizing the bit channel. The channel grid is applied to the entire image. CNN can identify the differences wherever these images have changed. Every Convolution layer is associated with an activation function. Here, ReLU (Rectified Linear Unit) is used as an activation function for obtaining the output as the input directly, if it is positive, otherwise the output will be zero. The dimensions of the images eventually decrease after passing through this layer, whereas there is an increase in the number of channels.

### VII. POOLING (MAXPOOL2D) LAYER

This layer performs down inspection of channels. It takes a group of 2X2 blocks and picks the maximal worth in each block, thus performing max pooling. These are used to reduce the computational expense and assist in controlling over fitting. The pooling size (i.e. the zone size pooled each time) needs to be selected carefully. It reduces the size of the feature map without losing significant information from the image. Eventually after passing through the pooling layer, dimensions of the image are further reduced along the two axes, thus, reducing the processing time. Joining convolution and pooling layers, CNN can learn the patterns of the image precisely.

### VIII. COMPLETELY (FULLY) CONNECTED LAYER

This layer is simply a Feed Forward Neural Network forming the last layer of the CNN. It takes the outputs from the previous layers, flattens them, and converts them into a single continuous vector that can be an input for the next layer. The last layer from the Fully Connected layers, yields the N-dimensional vector where, N refers to the number of classes that the program needs to output. Each value in the vector represents the probability that a certain feature belongs to a particular label. The objective of fully connected layer is simply to take the results of the convolution and pooling processes and use them to classify the image into a label. Here, the image is classified into no cancer, stage I cancer or stage II cancer. The fully connected part of the CNN goes through the backpropagation process to determine the most accurate weights in the training phase. Each neuron receives weights which prioritize the most appropriate label.

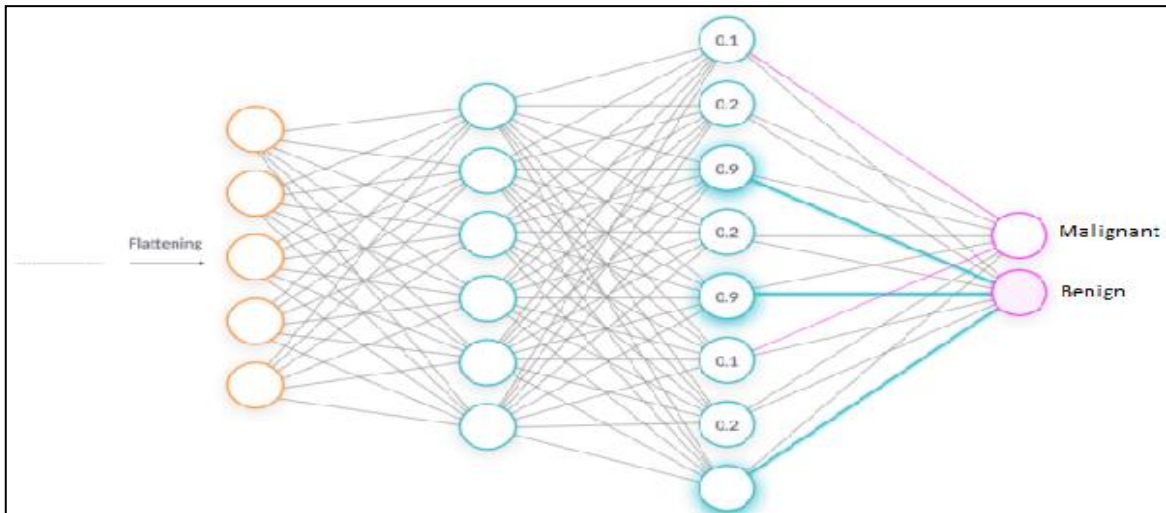


Fig. 3. Fully connected layers

Essentially, a FC layer determines the probability by which the highlights connect to a specific class. It also determines specific weights such that when there is processing between the weights and the output of past layer, correct probabilities for various classes are eventually obtained. Softmax function is used for finding the probability to identify the class that the image belongs to. The softmax function essentially converts the output of the final layer into a probability distribution. It is given by the following formula:

$$Softmax(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)}$$

### IX. IMAGE CLASSIFICATION

Depending on the probability obtained for the classes, the class with the highest probability is chosen as the final output of the system.

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 96, 96, 16)	416
max_pooling2d_1 (MaxPooling2)	(None, 48, 48, 16)	0
dropout_1 (Dropout)	(None, 48, 48, 16)	0
flatten_1 (Flatten)	(None, 36864)	0
dense_1 (Dense)	(None, 3)	110595
Total params: 111,011		
Trainable params: 111,011		
Non-trainable params: 0		

Fig. 4. Proposed model Summary

## X. EXPERIMENTAL RESULTS

There are 2 users performing the operations in the system:

- A) **Admin:** The admin, solely, has the authority to view the results of any patient which are diagnosed by the system earlier. The admin log-in has its own unique credentials to log into the system.
- B) **Doctor:** The doctor can obtain the result by uploading their CT image directory on the system. The doctor cannot access or view the details of any patient as the system is secured with unique credentials for every patient. The doctor can save the data into the system for future reference of the result.

The UI comprises of four pages:

- i. Home Page
- ii. Patient Registration
- iii. Doctor Login
- iv. Admin Login

### i. Home Page

Home Page of the system is the first page displayed when the program starts executing. There are three buttons on the Home Page which are:

- a. Registration: This button will redirect the user to the Registration Page
- b. Login: This button will redirect the user to the Login Page
- c. Exit: This button will terminate the execution of the program

### ii. Doctor Login

This page asks the user to create their account by setting their credentials to log into the system. There are two fields on this page asking to set User ID and Password for the user to log into the system. The page also contains a button which saves the credentials into the database of the system.

### iii. Admin Page

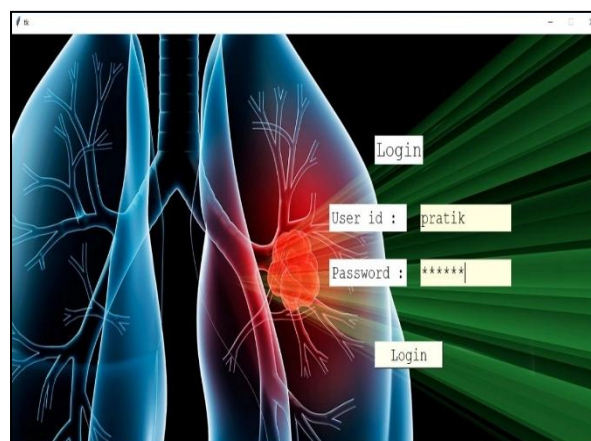
The admin page displays a list of all the patients whose testing has been carried out earlier for lung cancer and by selecting any particular patient, the lung cancer detection results of that patient are displayed.

### iv. Patient Registration

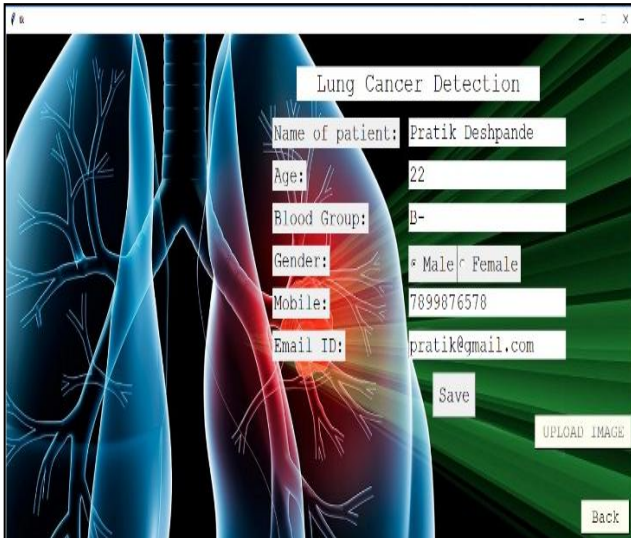
This page asks the user to enter the basic information of the patient to be diagnosed for lung cancer. It contains various fields which include name of the patient, age, gender, blood group, contact number, email id. If all the fields satisfy the validating criteria, the “Upload” button is enabled for uploading the CT images of the patient for the process of detecting lung cancer. Once all the processing is executed, the result is displayed on the very same window. On clicking on the “Back” button, the user gets logged out and is redirected to the home page.



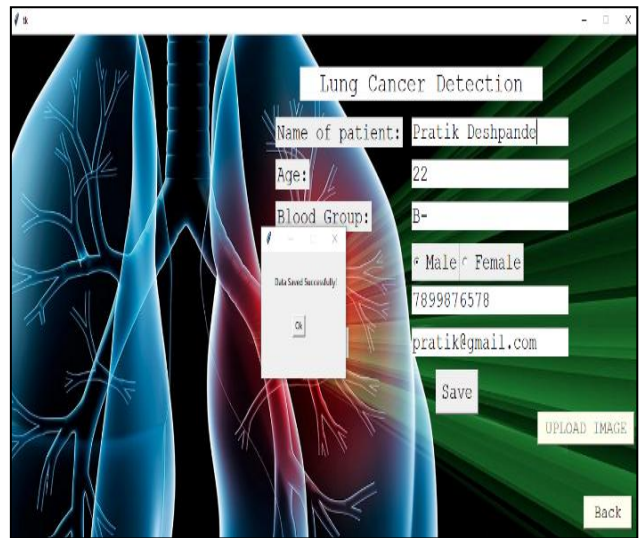
(a)



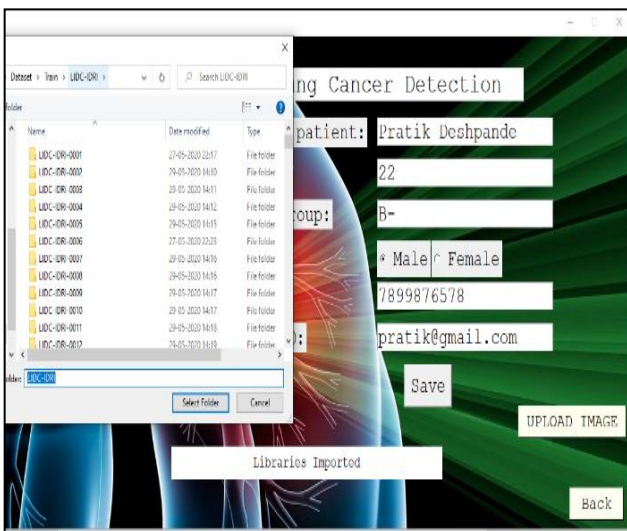
(b)



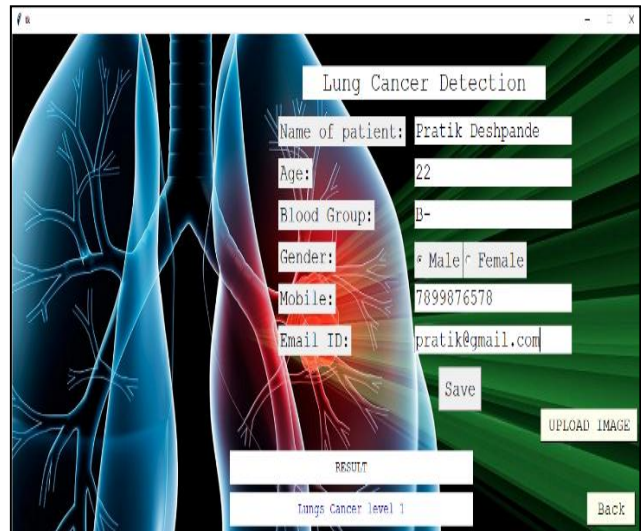
(c)



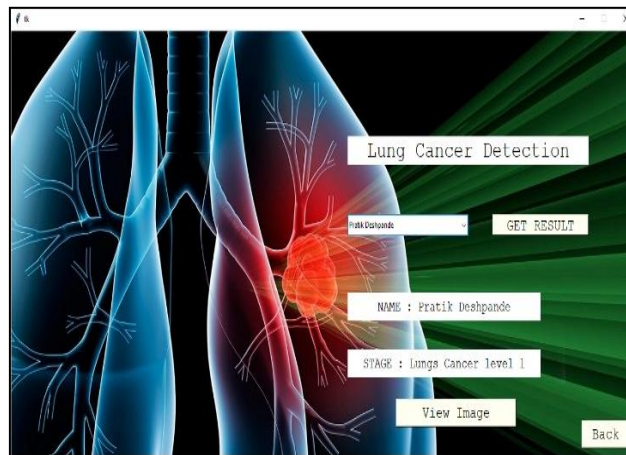
(d)



(e)



(f)



(g)

Fig. 5. Experimental Results (a) Home Page (b) Login Page (c) Registration Page (d) Registration Successful (e) Select Patient's CT images (f) Patient Result (g) View User Result

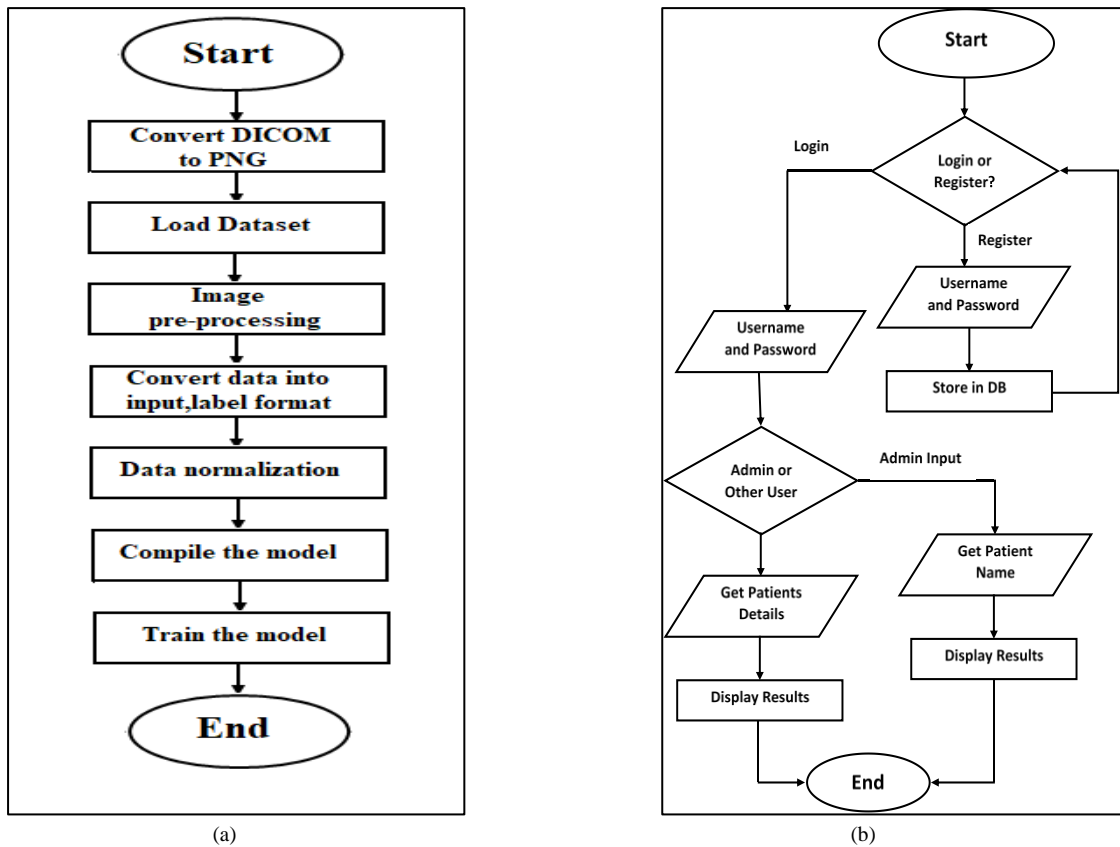


Fig. 6. System Flowcharts (a) Training Flowchart (b) Testing Flowchart

### XI. CONCLUSION

The system is built for detecting lung cancer and reducing the workload on radiologists, with an accuracy of 85%. The system works on CT scan based LIDC-IDRI dataset. It classifies lung cancer into three classes, particularly, stage I, stage II and no lung cancer. The overall accuracy of the system can be improved in future using 3D CNN.

### REFERENCES

- [1] World Health Organization, Available at: <https://www.who.int/news-room/fact-sheets/detail/cancer>.
- [2] Shalini Wankhade, Vigneshwari.S, “Prediction and Classification of Lung Nodule using Artificial Neural Network (ANN)”, Scopus, ICMEET 2019.
- [3] Sanyogita Bobade, Aditi Deshmukh, Abhinav Deshpande, Pratik Deshpande, “Detecting Lung Cancer from Pet Images”, International Journal of Innovative Research in Science, Engineering and Technology, vol 8, issue 11, Nov 2019.
- [4] RuchitaTekade, K. Rajeswari, et. al., “Lung Cancer Detection and Classification Using Deep Learning”, DOI: 10.1109/ICCUBEA.2018.8697352, 25 April 2019.
- [5] Patrice Monkam, Shouliang Qi, He Ma, Weiming Gao, Yudong Yao, (Fellow, IEEE), AND Wei Qian, “Detection and Classification of Pulmonary Nodules Using Convolutional Neural Networks: A Survey”, Digital Object Identifier 10.1109/ACCESS.2019.2920980.
- [6] Wenya Linda, Ahmed Hosny, Matthew B. et. al., “Artificial intelligence in cancer imaging: Clinical challenges and applications”, 05 February 2019, <https://doi.org/10.3322/caac.21552>.
- [7] A Manikandarajan, S Sasikala, “Detection and Segmentation of Lymph Nodes for Lung Cancer Diagnosis”, National Conference on System Design and Information Processing, 2013.
- [8] 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.
- [9] Convolutional Neural Network, Available at: <https://hackr.io/blog/convolutional-neural-network>





**INNO SPACE**  
SJIF Scientific Journal Impact Factor

**Impact Factor:  
7.512**

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details